Is the Quantity-Quality Trade-off a Trade-off for All, None, or Some?

Daniel L. Millimet
Le Wang

Is the Quantity-Quality Trade-off a Trade-off for All, None, or Some?

Daniel L. Millimet
*Southern Methodist University
and IZA*

Le Wang
University of New Hampshire

ABSTRACT

Is the Quantity-Quality Trade-off a Trade-off for All, None, or Some?[*]

Although the theoretical trade-off between the quantity and quality of children is well-established, empirical evidence supporting such a causal relationship − particularly on child health − is limited. We use two measures of child health to asses the quantity-quality trade-off across the entire distribution. Using data from the Indonesia Family Life Survey and controlling for the potential endogeneity of child quantity, we find evidence of a causal trade-off only for some and only in the short-run.

JEL Classification: C14, D10, I12, O12

Keywords: intrahousehold allocation, health, human capital, fertility,
              quantile treatment effects, stochastic dominance

Corresponding author:

Daniel L. Millimet
Department of Economics
Box 0496
Southern Methodist University
Dallas, TX 75275-0496
USA
E-mail: millimet@mail.smu.edu

---

1 Introduction

According to the classic theory on intra-household resource allocation, households make interdependent choices regarding the number of children and investments in child-specific human capital. The theory predicts a negative relationship between child outcomes (quality) and the number of children in the household (quantity) (Becker and Tomes 1976; Becker and Lewis 1973). Here, we test the empirical validity of this so-called *quantity-quality trade-off* using data on household size and child health in Indonesia. We do so accounting for the potential endogeneity of fertility decisions and allowing for heterogeneous effects across the distribution of child health.

The results are striking, yielding three main conclusions. First, distributions of weight-for-age are statistically different when we account for the endogeneity of the quantity of children. Second, despite the significant difference in the distributions across households with more than two or only two children, we fail to find statistically meaningful evidence of the quantity-quality trade-off over the whole distribution or on average as Two-Stage Least Squares (TSLS) estimates of the mean trade-off are statistically insignificant. In particular, using an identification strategy based on gender composition of the first two children, the answer to the question posed in the title is *some.* Finally, while modest evidence of the trade-off is found when using a more short-run measure of health (based on weight), there is no statistically meaningful evidence of a trade-off on a long-run measure (based on height) once the endogeneity of fertility decisions is addressed.

In the prior literature, the quantity-quality trade-off is typically modeled as arising from parental preferences for equal levels of quality across children combined with a binding budget constraint (Rosenzweig and Wolpin 1980). Empirical tests of the trade-off center on estimating demand equations for child-specific outcomes, where the number of children is one potential determinant of demand. Such studies typically find a negative relationship between the number of children and human capital investments (e.g., Rosenzweig and Zhang 2006; Conley and Glauber 2005; Glick et al. 2005; Lee 2004), although a few find no effect (e.g., Black et al. 2005) or even a positive effect (e.g., Qian 2008).

Empirical tests of the trade-off have focused, however, mainly on schooling. While clearly impor-

tant, health constitutes another salient component of child quality. Researchers and policymakers are cognizant of the impact of child health on adult health and other economic outcomes. For example, Thomas et al. (1990, 1991) note the relationship between child anthropometric measures and the probability of survival and skill development, and Thomas and Frankenberg (2002) state that adult stature is largely determined during the fetal and early childhood periods. Moreover, adult health is associated with and labor market outcomes at both the microeconomic and macroeconomic levels. Many studies have found that there is a positive impact of height on earnings (Strauss and Thomas (1998) provide an excellent review), Fogel (1994) documents the parallel historical increases in height and economic growth, and Weil (2005) suggests that variation in health explains approximately 20% of the cross-country variation in (log) income per worker, roughly the same fraction as explained by variation in education.[1]

Given the importance of children's health, a small literature has developed investigating its determinants; Strauss and Thomas (1995) survey the literature. In many of these studies, household size enters the analysis as a control, although the estimated relationship is not of primary interest and the issue of causation is often ignored. Two recent exceptions are Glick et al. (2007) and Angrist et al. (2006). Glick et al. (2007) utilize data on twins to isolate the casual effect of fertility on child health and school enrollment using Romanian data, finding sizeable negative effects that increase in magnitude after accounting for the endogeneity of the number of children. Angrist et al. (2006) use Israeli data on twins and the gender composition of children to estimate the causal impact of fertility on a variety of children's outcomes as adults (e.g., completed education, labor market outcomes, and own marital and fertility patterns), finding little impact.

In this paper, we advance this literature in two ways. First, we assess the empirical validity of the quantity-quality trade-off using two measures of child health: height-for-age and weight-for-age. Both are frequently used measures, where the former (latter) is a reflection of relatively long-term (short-term) health status.[2] Second, we assess this trade-off within a *distributional* framework,

---

[1]López-Casasnovas et al. (2005) offer a detailed theoretical and empirical account of the linkages between health and economic development.

[2]Cogill (2003, p. 11) states that height-for-age "identifies past undernutrition or chronic malnutrition" and "cannot measure short term changes in malnutrition." He notes that weight-for-age "reflects both past (chronic) and/or present (acute) undernutrition." See also Thomas et al. (1991, 1996).

via the estimation of quantile treatment effects (QTE). Moreover, we provide a welfare-consistent method of summarizing the QTEs based on the notion of stochastic dominance (SD).[3] This approach uncovers any heterogeneity in the magnitude and existence of the trade-off across the distribution of child quality, enabling one to answer: Is the quantity-quality trade-off a trade-off for all, none, or some? The answer to this question may not only shed light on the moderately inconsistent empirical findings detailed at the outset, but also is vital for sound policymaking. For instance, if policymakers are interested in improving the health of the least healthy children, but the trade-off is more pronounced in the upper tail of the distribution, then inferring the impact of fertility-reducing programs (e.g., investments in family planning clinics) based on the mean trade-off may vastly overstate the effects of such programs.

To perform the analysis, we first present a simple theoretical model based on Becker and Tomes (1976) showing why the trade-off may not be homogeneous. Then, we utilize data from the 2000 wave of the Indonesian Family Life Survey (IFLS) on roughly 3,000 children ten years of age and younger to assess the trade-off. Specifically, we assess the treatment effect of residing in a household with more than two children (relative to only two children) on child health, controlling for potentially confounding observable *and* unobservable characteristics by implementing an instrumental variable (IV) method put forth in Abadie (2002). The method relies on a binary instrument, and we use the gender composition of the first two children, as utilized in Butcher and Case (1994), Angrist and Evans (1998), Cruces and Galiani (2004), Conley and Glauber (2005), Angrist et al. (2006), and Henderson et al. (2006).

Assessing the distributional consequences of the quantity-quality trade-off in Indonesia represents more than academic curiosity. After the Asian financial crisis of the late 1990s, children's health and poverty have become even more of a policy concern in Indonesia. At the 27th Special Session of the United Nations General Assembly on Children in 2002, Dr. Achmad Sujudi, Indonesian Minister for Health, stated: "[U]nless vigorous measures are urgently taken, the threatened menace of a 'lost generation' of malnourished, unhealthy and poorly educated children could

---

[3]Although there exist alternative frameworks for comparing distributions (or portions of distributions), the information content provided by QTE and SD analysis has led to an increasing number of applications (see, e.g., Bitler et al. 2006; Abadie 2002; Maasoumi and Heshmati 2000).

become a stark reality," where vigorous measures include "investing more of its resources for the poor, particularly for children and women, and support for the 20/20 initiative in which the government should allocate 20% of its national budget to social development programs."[4] Such extreme measures are necessary since even as the Indonesian economy recovers at the macroeconomic level, the microeconomic consequences are not easily reversed given past inadequacies in the provision of health care (FitzGerald 2001). Rukumnuaykit (2003) estimates that infant mortality increased roughly 1.4 percentage points after the financial crisis. Recent figures provided by UNICEF (2005) indicate that 28% of children under age five are moderately or severely underweight; the under age five (one) mortality rate is 38 (30) per 1,000 live births, nearly five times higher than in the United States and placing it in the bottom half of countries in the world.

Indonesia also has a relatively long history of encouraging smaller families, particularly families with two children. The National Family Planning Coordinating Board, BKKBN (*Badan Koordinasi Keluarga Berencana Nasional*), was established as part of a national population campaign in 1970 in order to reverse the trend of rapid population growth and to promote the welfare of women and children. The campaign conveyed two consistent messages: (i) "A Small Family Is a Happy, Healthy, and Prosperous Family" and (ii) *Dua Anak Cukup*, meaning "Two Children Are Enough" (Weidemann 1999). Weidemann (1999, p. 9-10) summarized the extent of the campaign, which continues at present:

> "This theme was repeated everywhere, in all kinds of forums. The back of a five-rupiah coin, displayed a two-child family with the message "Family Planning: the Way to Prosperity." Everywhere, posters with this message greeted people from billboards and storefront windows. Family planning subjects are written into soap opera and film scripts, as well as radio and television programs. Even Indonesia's ancient puppet theatre programs, the wayang kulit and wayang golek, feature family planning. These nation-wide information campaigns are still going on around Indonesia."

The result of the campaign has been a reduction in the average number of children per woman from 5.57 in 1970, to 4.73 in 1980, to 3.50 in 1990, and further to 2.60 in 2000.[5]

The political desire to improve child health combined with the message that the 'ideal' family

---

[4]Available at `http://www.un.org/ga/children/indonesiaE.htm`.

[5]Data are from the United Nations Common Database (`http://globalis.gvu.unu.edu/`).

contains two children motivates our focus on the treatment effect of more than two versus only two children. Moreover, the large number of severely unhealthy children suggests that one should examine the determinants of child health at all parts of the distribution. The remainder of the paper is organized as follows: section 2 provides a simple theoretical framework to motivate the distributional analysis; section 3 details the econometric approach and data; section 4 discusses some preliminary analysis focused mainly on the validity of the instrument; section 5 discusses the distributional results; and, section 6 concludes.

2 Theoretical Model

To motivate the distributional analysis and clarify the identification strategy, we present a simple extension to the model in Becker and Tomes (1976). To begin, assume households maximize utility given by $U(n, q, r, c)$, where n is the number of children, q is the quality per child, r is the sex ratio of children ($r \in [0, 1]$), and c is consumption. Child quality depends on market purchased health inputs and a household-level health endowment; the production function is $q = q(w, \theta)$, where w is a vector of market purchased health inputs and θ is the household health endowment.[6] We assume positive marginal products for each input, $q_w > 0$ and $q_\theta > 0$; we make no assumptions about the cross-derivative $q_{w\theta}$. The household budget constraint is given by

$$p_c c + p_n n + p_w w n - \delta(r - 0.5)^2 = I \tag{1}$$

where p_c is the price of c, p_n is the fixed cost per child independent of the level of child quality, p_w is a vector of input prices, δ is a parameter reflecting cost-savings due to having more children of one gender, and I is household income.

The household maximizes U given the production function for child quality and the budget

---

[6]Designation of the health endowment at the household level implies that the endowment of each child is equal. As Becker and Tomes (1976) discuss, heterogeneous endowments do not alter the primary implications of the model.

constraint. The equilibrium conditions are:

$$
\begin{aligned}
\frac{\partial U}{\partial c} &= \lambda p_c = \lambda \pi_c \\
\frac{\partial U}{\partial q} &= \lambda \frac{p_w}{\partial q/\partial w} n = \lambda \pi_q \\
\frac{\partial U}{\partial n} &= \lambda(p_w w + p_n) = \lambda \pi_n
\end{aligned}
\tag{2}
$$

where λ is the marginal utility of income and π_c, π_q, and π_n are the shadow prices of consumption, child quality, and child quantity, respectively. As is well known, the equilibrium conditions imply that the shadow price of child quality, π_q, is positively related to the number of children, n. Thus, an 'exogenous' increase in fertility increases the shadow price of child quality, which reduces the demand for quality per child, q, which reduces the shadow price of child quantity, π_n, further increasing n, and so on. This reaction yields the familiar quantity-quality trade-off.

Further examination, however, reveals that the magnitude of the trade-off – although not the trade-off itself – depends on the health endowment, θ, as well as the form of the health production function. Specifically, the impact of an exogenous increase in fertility depends on the resultant change in the shadow price of child quality. The magnitude of the change in π_q depends on the prices of market purchased health inputs, p_w, which are assumed fixed, and the marginal productivity of market purchased health inputs, q_w, which in turn may depend on the household health endowment, θ. Consequently, the sign of the cross-derivative $q_{w\theta}$ has important implications.

Consider the three possible cases. First, if $q_{w\theta} > 0$ (as in a Cobb-Douglas production function), then the change in π_q from an exogenous increase in n is deceasing in θ. Thus, the quantity-quality trade-off will be larger in magnitude in the lower tail of the distribution of θ. Second, if $q_{w\theta} < 0$, then the opposite occurs and the trade-off will be larger in magnitude in the upper tail. Finally, if $q_{w\theta} = 0$ (e.g., if the production function is additively separable), then the change in π_q from an exogenous increase in n is independent of θ. Thus, the magnitude of the trade-off will be independent of θ. Assessing heterogeneity in the trade-off is one of the goals of the empirical analysis. However, since θ is unobserved, we test for such heterogeneity not across the distribution of θ, but rather across the distribution of q itself (net of other observable inputs, w). An additional goal – assuming the

quantity-quality trade-off is found to exist over at least a portion of the distribution – is to assess the robustness of distributional comparisons over a large class of social welfare functions.

3 Estimation

3.1 Regression Approach

To initially examine the data, we use Ordinary Least Squares (OLS) and TSLS to estimate the model

$$q_i = X_i\beta + D_i\tau + \varepsilon_i \tag{3}$$

where q_i is the health of individual i, X is a vector of individual, family, and community attributes (which includes health inputs, w), D is an indicator variable equal to one if there are more than two children in the household (zero otherwise), and ε is a mean zero, possibly heteroskedastic, normally distributed error term. We also estimate (3) by TSLS, instrumenting for D using an exclusion restriction based on the gender composition of children (discussed below).

3.2 Distributional Approach

3.2.1 Quantile Treatment Effects

To allow for heterogeneous effects of household size, we estimate *quantile treatment effects* (QTE).[7] To begin, let Q_0 and Q_1 denote two health variables to be compared. For instance, Q_0 (Q_1) may represent a measure of weight of children residing in households with only two children (more than two children). $\{q_{0i}\}_{i=1}^{N_0}$ is a vector of N_0 observations of Q_0 (denoted by $D_i = 0$); $\{q_{1i}\}_{i=1}^{N_1}$ is an analogous vector of realizations of Q_1 (denoted by $D_i = 1$). Let $F_0(q) \equiv \Pr[Q_0 < q]$ represent the cumulative density function (CDF) of Q_0; define $F_1(q)$ similarly for Q_1. The pth quantile of F_0 is given by the smallest value q_0^p such that $F_0(q_0^p) = p$; q_1^p is defined similarly for F_1. Under this

---

[7]As an alternative, one could utilize IV techniques recently developed for quantile regression models (e.g., Chernozhukov and Hansen 2005). We pursue the current approach as it lends itself more naturally to stochastic dominance testing (discussed below), which is informative for comparing entire distributions when treatment effects vary across the distribution.

notation, the QTE for quantile p is given by $\Delta_p = q_1^p - q_0^p$, which is simply the horizontal difference between the CDFs at probability p.[8] Estimates, $\widehat{\Delta}_p$, are obtained using the sample analogues of $q_j^p \equiv \inf_q\{\Pr[Q_j \leq q] \geq p\}$, $j = 0, 1$ and $p = 0.01, ..., 0.99$. In the results below, we plot $\widehat{\Delta}_p$, as well as 90% confidence intervals based on a simple bootstrap technique, similar to Bitler et al. (2006).

3.2.2 Test of Equality

In addition to examining the QTEs at each quantile, we test the joint null $H_o : \Delta_p = 0\ \forall p \in (0, 1)$, or equivalently $H_o : F_0 = F_1$, utilizing a two-sample Kolmogorov-Smirnov (KS) statistic and a bootstrap procedure (Abadie 2002). See Appendix A for detail.

3.2.3 Selection Issues

Selection on Observables Thus far, the distributional analysis has only considered *unconditional* distributions. However, dependence between the treatment (number of children) and other determinants of the outcome (child health) most certainly precludes one from inferring causation. To alleviate the bias attributable to *selection on observables*, we utilize several sets of observable determinants of child health and analyze the health distributions *adjusting for covariates*. Here, we utilize the inverse propensity score weighting procedure as applied in Bitler et al. (2006) (see also Firpo (2007) and Appendix A for detail).

Selection on Unobservables Adjusting the distributions of child health for observable covariates is not sufficient to identify the causal effect of household size if there is *selection on unobservables*. As a result, we implement the IV procedure developed in Abadie (2002) to compare the distributions of potential health outcomes for (a subpopulation of) children. According to Imbens and Rubin (1997), given a binary instrument, the potential distributions of the outcome are identified for the subpopulation (referred to as *compliers*) whose treatment assignment (in this case, number of children) is determined by the instrument.

---

[8]It is important to note that the QTEs do not correspond to quantiles of the distribution of the treatment effect unless the assumption of *rank preservation* holds (Heckman et al. 1997; Firpo 2007). Absent this assumption, whereby the ranking of children would remain unchanged under of the two household types, the QTE simply reflects differences in the quantiles of the two marginal distributions.

To proceed while avoiding new notation, re-define Q_0 and Q_1 as two *potential* health variables for the untreated (children in households with two children) and treated (children in households with more than two children). Let q_{0i} and q_{1i} represent the corresponding values for observation i, $i = 1, ..., N_0 + N_1$, from the respective distribution, and let Z_i be a binary instrument. Denote $D_i(0)$ the value of D_i if $Z_i = 0$; similarly for $D_i(1)$. Given this setup, for any child i, the pair of treatment indicators $\{D_i(0), D_i(1)\}$ and the pair of potential health outcomes $\{q_{0i}, q_{1i}\}$ are not both observed since only one state of the world – $Z_i = 0$ or $Z_i = 1$ – is realized. Instead, the realized treatment assignment $D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)$ and the realized potential outcome $q_i = q_{1i}D_i + q_{0i}(1 - D_i)$ are observed.

Let $F_0^c(q)$ and $F_1^c(q)$ represent the CDFs of potential health outcomes for compliers in the control and treatment groups, respectively, which are defined as follows:

$$
\begin{aligned}
F_0^c(q) &= \mathbb{E}\left[I\{q_{0i} \leq q\} | D_i(1) = 1, D_i(0) = 0\right] \\
F_1^c(q) &= \mathbb{E}\left[I\{q_{1i} \leq q\} | D_i(1) = 1, D_i(0) = 0\right]
\end{aligned}
\tag{4}
$$

If Z_i satisfies the following three assumptions:

(i) Independence: $\{q_{0i}, q_{1i}, D_i(0), D_i(1)\} \perp Z_i$

(ii) Correlation: $\Pr[Z_i = 1] \in (0, 1)$ and $\Pr[D_i(0) = 1] < \Pr[D_i(1) = 1]$

(iii) Monotonicity: $\Pr[D_i(0) \leq D_i(1)] = 1$,

then QTEs based on the distributions $F_0^c(q)$ and $F_1^c(q)$ identify the causal effect of household size for the subpopulation of compliers despite unobservables being correlated with both the quantity and quality of children (Imbens and Angrist 1994; Angrist et al. 1996). Moreover, as shown in Abadie (2002), QTEs based on the distributions $F_0^c(q)$ and $F_1^c(q)$ are proportional to QTEs based on the distributions $G_0(q)$ and $G_1(q)$, where G_0 (G_1) represents the distribution of health outcomes for children with $Z_i = 0$ $(Z_i = 1)$. Thus, QTE estimates obtained using the empirical CDFs of G_0 and G_1 identify both the *sign* and *statistical significance* of the QTEs.

Finally, we also estimate the QTEs combining Abadie's (2002) approach with the previous method of adjusting for observables. Conditioning on X, in combination with the IV strategy, increases the likelihood that the instrument satisfies the necessary criteria. In addition, even in randomized experiments, controlling for potentially confounding variables may be advisable to capture residual covariance, as well as the fact that randomization only balances confounders in expectation (Imai and van Dyk 2004).

In the analysis, the instrument is an indicator of whether the first two children are of the same gender. As shown in Rosenzweig and Wolpin (2000), for gender composition to be a valid exclusion restriction certain assumptions are required (see also Rosenzweig and Zhang (2006)). Specifically, we require that (i) the sex ratio of children, r, enters the household utility function, (ii) child-rearing costs do not vary with r (i.e., $\delta = 0$ in (1)) and/or consumption, c, and per child quality, q are strongly separable in the utility function, and (iii) per child quality, q, and r are strongly separable in the child quality production function. We assess the validity of these requirements below through examination of the first-stage regressions and various sensitivity analyses.[9]

3.2.4 Stochastic Dominance

While examination of the QTEs is of great interest, in the event that the QTE estimates vary in sign or statistical significance over the distribution, tests for SD enable welfare comparisons of distributions. To begin, assuming general von Neumann-Morgenstern conditions, let $\mathcal{U}_1$ denote the class of (increasing) social welfare functions u such that welfare is increasing in health (i.e. $u' \geq 0$), and $\mathcal{U}_2$ the sub-class of functions in $\mathcal{U}_1$ such that $u'' \leq 0$ (i.e. concavity). Concavity represents an aversion to inequality in the health of children. Note that u refers to the welfare function of a policymaker, not the household.

Under this notation, Q_0 First Order Stochastically Dominates Q_1 (denoted Q_0 FSD Q_1) *iff* $\mathbb{E}[u(Q_0)] \geq \mathbb{E}[u(Q_1)]$ for all $u \in \mathcal{U}_1$, with strict inequality for some u, where $\mathbb{E}[\cdot]$ is the expected

---

[9]The monotonicity assumption is inherently untestable. It rules out the presence of *defiers*: households who have only two children because the first two children were of the same gender, or households who have a third child because the first two children were of opposite gender.

value operator. Equivalently,

$$F_0(q) \leq F_1(q) \qquad \forall q \in \mathcal{S}, \text{ with strict inequality for some } q \qquad (5)$$

where $\mathcal{S}$ denotes the union of the supports of Q_0 and Q_1. Condition (5) may be alternatively stated as

$$\Delta_p \leq 0 \qquad \forall p \in (0,1), \text{ with strict inequality for some } p. \qquad (6)$$

If Q_0 FSD Q_1, then the expected social welfare from Q_0 is at least as great as from Q_1 for all increasing welfare functions, with strict inequality for some function(s) in the class. The distribution of Q_0 Second Order Stochastically Dominates Q_1 (denoted as Q_0 SSD Q_1) *iff* $\mathbb{E}[u(Q_0)] \geq \mathbb{E}[u(Q_1)]$ for all $u \in \mathcal{U}_2$, with strict inequality for some u. Equivalently,

$$\int_{-\infty}^{q} F_0(v)dv \leq \int_{-\infty}^{q} F_1(v)dv \qquad \forall q \in \mathcal{S}, \text{ with strict inequality for some } q, \text{ or} \qquad (7)$$

$$\int_{0}^{p} \Delta_v dv \leq 0 \qquad \forall p \in (0,1), \text{ with strict inequality for some } p. \qquad (8)$$

If Q_0 SSD Q_1, then the expected social welfare from Q_0 is at least as great as that from Q_1 for all social welfare functions in the class $\mathcal{U}_2$, with strict inequality holding for some function(s) in the class. FSD implies SSD and higher orders. To test for FSD and SSD, we use generalizations of the Kolmogorov-Smirnov test criteria and the simple bootstrap proposed in Maasoumi and Heshmati (2000) (see Appendix A for detail).

3.3 Data

The data are obtained from the IFLS. The IFLS contains a longitudinal sample of households representing about 83% of the Indonesian population living in 13 of the 26 provinces in 1993; see Strauss et al. (2004a, 2004b) for a complete description. We utilize the 2000 wave to form a sample of roughly 3,000 children aged ten and under, with at least one identifiable birth parent in the survey, and who come from a household with at least two children.

We use two measures of child health – height and weight – standardized to the reference popu-

lation for the child's age and sex utilizing the 1990 British Growth Reference data.[10] The quantity of children is defined as the number of children 'belonging' to a given set of parents. Assuming that households take into account their spouses' fertility history, this definition *includes* any children from previous marriages. For example, if one (or both) parents were previously married and entered the current marriage with children, then these children are counted in the number of siblings. Our definition, however, does *exclude* children 'belonging' to other couples who reside in the same home. Finally, our definition *includes* children 'belonging' to the couple, but not currently residing in the household if the child resided in the household during the prior survey waves (1993 or 1997).

Once the number and identity of siblings are established, we create a dummy variable, *MoreThan2*, equal to one if the household has more than two children and zero otherwise; households with less than two children are omitted. As the instrumental variable, we define *SameSex2* equal to one if the first two children are of same gender, and zero otherwise. Because the first two children are the subjects of the quasi-experimental design, we restrict the sample to these children (Angrist et al. 2006). For comparison, Appendix B reports results using all children under ten years of age.

When adjusting for covariates, we utilize three sets of covariates, X. Control Set A includes the child's gender, age in months, and birth order. Control Set B includes the variables in A plus:

Parental: dummy variables for mother's and father's education, mother's and father's height, mother's and father's weight, mother's and father's age, dummy variables for mother's and father's work status, dummy variables for mother's and father's religion, dummy indicating whether or not parents' height are missing, dummy indicating whether or not father's age is missing;

Household: dummy variable for ownership of farm, dummy variable indicating if the household head is female, dummy variables for decision-making powers concerning children's health, dummy variables for type of dwelling, house size, number of rooms in the house, dummy variables for type of floor materials in the house, dummy variables for type of wall materials in the house, dummy variables for water source for the house, dummy variable indicating

---

[10]Note, there are a few missing values for the health outcomes. Thus, the samples utilized to analyze the two measures are not identical.

if water is boiled prior to consumption, dummy variables for type of sanitary conditions of household, region dummy, province dummies, and a dummy variable indicating whether or not house size is missing.[11]

Control Set C includes the variables in B plus community-level variables: number of and minimum distance to various health facilities in the community. These variables are missing for a significant portion of the sample; we replace missing data with the sample mean and include dummy variables denoting missing data.

Table 1 provides summary statistics. On average, children in a household with two children are healthier in terms of their height-for-age and weight-for-age z-scores, consonant with the quantity-quality trade-off.

4 Preliminaries

Prior to presenting the distributional results, Table 2 presents the regression results utilizing OLS (Panel A) and TSLS (Panel B). This is useful not only for comparison to the distributional results, but also relative to the literature. The OLS results indicate that moving from a two child household to one with more than two children is associated with a small, statistically significant reduction in both height-for-age and weight-for-age: roughly 0.11 to 0.21 standard deviations. Treating fertility as endogenous, however, we fail to find any statistically significant impact of the number of children regardless of conditioning set utilized, consonant with Angrist et al. (2006). While the estimates are imprecise, we note that the point estimates are *positive* using Control Sets B and C for the long-run measure, height-for-age.

For the TSLS estimates to be meaningful, our instrument, *SameSex2*, has to be correlated with the endogenous variable, but independent of health outcomes conditional on the endogenous variable. We discuss the evidence concerning each of these requirements in turn.

To assess the former requirement, we provide results from several specification tests, as well as the results from the first-stage regressions, in Table 3. In terms of the first-stage results, as

---

[11]Missing values are replaced with sample means.

expected, our *SameSex2* has a positive and statistically significant effect on *Morethan2*, regardless of sample or conditioning set. Specifically, having the first two children of same gender increases the probability of having more than two children by roughly five percent.

Although encouraging, statistical significance alone is not sufficient to rule out a weak instrument problem. Thus, Table 3 presents results from two additional tests assessing the relevance of our instrument: the Kleibergen-Paap rk Wald Statistic and the first-stage F-statistic for the significance of the instrument. Our instrument fares very well in terms of these tests. Finally, despite these comforting findings, we also utilize several weak-instrument robust inference approaches to assessing the statistical significance of the endogenous variable in the second-stage: Anderson-Rubin F-test Statistic and Stock-Wright S-statistic. The results, in Panel C of Table 2, indicate that our inference is robust to the use of these alternatives. Moreover, results from the reduced form regressions of child health on *SameSex2* are presented in Panel D of Table 2. These regressions also fail to yield any statistically significant estimates.

The second requirement for a valid instrument is independence; the instrument must be independent of potential health outcomes (conditional on X). Such dependence could arise from either a direct impact of the instrument on health outcomes (i.e., the instrument belongs in the second-stage), or an indirect effect arising from correlation between the instrument and unobservable health determinants (i.e., the instrument itself is endogenous). The existing literature suggests several reasons for concern. First, as discussed above, Rosenzweig and Wolpin (2000) point out that there may exist economies of scale in households with children of the same sex. This implies a (positive) direct effect of our instrument on child health due to an income effect. We are not able to test for a direct link between gender composition and the intra-household allocation of resources or availability of household public goods due to data limitations (although a direct link does not preclude the viability of the identification strategy, it only implies that additional restrictions on household preferences are needed). Second, Baez (2008) suggests that interactions with opposite sex siblings may benefit individuals due to complementarities arising in the production of health. This implies a negative direct effect of our instrument of child health.

As an informal test of a direct effect of the instrument, we regress each health measure on *Same-*

Sex2 and *Morethan2*, along with the other covariates discussed above. Conditional on *Morethan2*, the coefficient on *SameSex2* is statistically insignificant in both cases (height-for-age: $\widehat{\beta}_{samesex2} = 0.02$, $s.e. = 0.049$; weight-for-age: $\widehat{\beta}_{samesex2} = -0.014$, $s.e = 0.052$). Moreover, the negative point estimate for weight-for-age is not consistent with the mechanism suggested in Rosenzweig and Wolpin (2000).

A third possibility that would invalidate independence of the instrument is endogenous sex selection. While child gender is generally exogenous, this need not be the case if abortion or infant mortality rates differ by gender. Such a selection process leads to an imbalance in the sex ratio. However, in 1992, the population sex ratio was unity (in contrast, it was 1.06 men per woman in mainland China, 1.07 in India, 1.08 in Bangladesh, and 1.02 in Thailand).[12] This is consistent with the fact that abortion is illegal in Indonesia except under certain extreme circumstances. In addition, the under age five infant mortality rate in 2000 was *higher* for boys than girls (47.3 versus 34.3).[13]

Absent gender-biased abortion or infant mortality rates, the gender of children past the first child could nonetheless be endogenous if there exists male preference, as suggested in Jensen (2002).[14] However, as suggested by the figures referenced above, there is little evidence to suggest gender bias in Indonesia. In addition, Kevane and Levine (2001) find no evidence of 'missing girls' in Indonesia, and document a narrowing, if not a complete disappearance, in the gender gap in educational attainment. Similarly, Levine and Ames (2003) find that even during the Asian financial crisis in the late 1990s, girls did not suffer disproportionately relative to boys, and may have actually benefited, in terms of a wide range of measures including school enrollment, immunizations, and mortality. Finally, Wongboonsin and Ruffolo (1995) and Soeradji and Hatmadji (1994), among others, find that son preference appears nonexistent in Indonesia.

Despite this prior evidence, we nonetheless assess the existence of son preference in our data. First, a simple tabulation of our data reveals that less than 52% of first born children are male, as are first and second born children and all children under ten years of age in our sample. Second, we

---

[12]See `http://www.census.gov/apsd/www/statbrief/sb93_18.pdf`.

[13]See `http://www.census.gov/ipc/www/idb/tables.html`.

[14]We thank the editor for bringing this to our attention.

utilize regression analysis to assess determinants of second and higher order births. As suggested in Lee (2008), if son preferences exist, households with daughters at earlier parity should have more children. Specifically, we utilize two fertility measures: *Morethan1* equal to one if a household has more than one child (zero otherwise), and *Morethan2* (defined above). We also use two measures of sex composition: *Daughter1* equal to one if the first child is a girl (zero otherwise), and *Daughter2* equal to one if the first two children are girls (zero otherwise). These regressions can be thought of as alternative first-stage results of a TSLS procedure or as parity progression models as in the demography literature.[15] The results – obtained by conditioning on Control Set C – are displayed in Table 4. Despite the relative precision of the estimates (the standard errors are similar to those in Table 3), the alternative measures of sex composition do not have a statistically meaningful relationship with family size. In addition, the point estimates in Panels A and B are *negative*, indicating if anything a lower probability of larger families when the first child is a daughter.

The discussions above indicate that our instrument is at least *plausibly exogenous*. Nonetheless, we undertake a final test based on a method proposed in Conley et al. (2008). The method is couched in a modified version of the model given in (3). The modification entails permitting the instrument to have a direct impact on child health. Formally, the model is now given by

$$q_i = D_i\tau + Z_i\gamma + \varepsilon_i \tag{9}$$

$$D_i = Z_i\pi + u_i \tag{10}$$

where we omit the other covariates, X, purely for notational convenience.

A valid instrument requires $\gamma = 0$. Conley et al. (2008) seek to construct a valid confidence interval for τ even when this requirement does not hold. Their approach – referred to as the Union of Confidence Intervals (UCI) with γ support assumption approach – assumes that $\gamma \in \Gamma$, where Γ is the bounded support of γ. Given a specific value of γ from the support, say γ_0, one can subtract

---

[15]Except that in the demography literature, the conditional mean is usually modeled using a logit model instead of a linear probability model.

$Z_i\gamma_0$ from equation (9), yielding

$$\widetilde{q}_i = q_i - Z_i\gamma_0 = D_i\tau + Z_i(\gamma - \gamma_0) + \varepsilon_i. \tag{11}$$

If $\gamma = \gamma_0$, then one can consistently estimate τ via TSLS using Z as an instrument and construct a symmetric $(1-\alpha)\%$ confidence interval, $CI_N(1-\alpha, \gamma_0)$, based on the asymptotic variance of the TSLS estimator. However, because the true value is unknown, one can estimate τ for all values within the support Γ via TSLS regressions of $\widetilde{q}_i$ on D_i and construct the union of the resulting confidence intervals. As long as $\gamma \in \Gamma$, the union will contain the true parameter value of τ as $\Pr[\tau \in \underset{\gamma_0 \in \Gamma}{\cup} CI_N(1-\alpha, \gamma_0)] \geq (1-\alpha)$ asymptotically.[16]

To implement the UCI approach, we utilize an interval for Γ consistent with our discussions above. Specifically, we assume the positive income effects and negative complementarity effects of *SameSex2* are roughly equal, yielding a value of γ close to zero. As such, we use a symmetric support centered at zero: $\Gamma = \{-\delta, \delta\}$ for different values of δ.

The results are shown in Figure 1 and obtained using Control Set C. The left (right) column displays the results for height-for-age (weight-for-age). The figures reveal only that this approach adds extra uncertainty, thereby increasing the width of the confidence intervals. Thus, allowing for $\gamma \neq 0$ only strengthens our inability reject the null hypothesis that τ is zero, consonant with our original TSLS results.

While all these discussions do not provide a definite answer to the question of whether our instrument is valid, they do increase our confidence in the identification strategy. We now turn to the distributional results.

---

[16]Since $\Pr[\tau \in CI_N(1-\alpha, \gamma_0)] \to (1-\alpha)$ when $\gamma = \gamma_0$, it follows that $\Pr[\tau \in \underset{\gamma_0 \in \Gamma}{\cup} CI_N(1-\alpha, \gamma_0)] \geq (1-\alpha)$. In practice, we approximate the interval by taking the union of the confidence intervals for grid points in the support Γ.

5 Results

5.1 Unconditional Distributions

The QTEs of the unconditional distributions are plotted in Figure 2. Treating *MoreThan2* as exogenous (denoted as 'No Instrument' in Figure 2), the estimated QTEs are negative at every quantile for both height-for-age and weight-for-age, consonant with the quantity-quality trade-off applying to all, instead of just on average. Moreover, the estimates are statistically different from zero in both cases over the majority of the distribution.[17] Finally, in both cases, the QTE estimates are relatively constant across the distribution. In light of the theoretical model, this is consonant with the production function being additively separable for both anthropometric measures.

The corresponding tests for equality and SD are provided in Table 5. Several findings are noteworthy. First, we easily reject the null of equal distributions for both health measures (height-for-age: p = 0.000; weight-for-age: p = 0.014). Second, despite the fact that the QTEs are negative at all quantiles for both health measures, we fail to observe either a first- or second-order SD ranking as the distributions cross below the first percentile.[18] Moreover, the simple bootstrap confirms the inability to rank the distributions in even the second-order sense (height-for-age: $\Pr(s \leq 0) = 0.134$; height-for-age: $\Pr(s \leq 0) = 0.448$). The inability to find a statistically significant first- or second-order ranking utilizing the simple bootstrap is attributable to the relatively frequent occurrence of positive QTEs at low quantiles, as indicated in Figure 2.

Although interesting, these results treat *MoreThan2* as exogenous (and fail to adjust for any covariates). Thus, we now turn to the IV results (still not controlling for any covariates). Examination of the plots (Figure 2, second row) yields two findings. First, for height-for-age, the QTEs are almost always negative below roughly the 90th quantile but positive above it, although the 90% confidence intervals for the QTEs nearly always contain zero. Second, for weight-for-age, the QTEs are negative at nearly all quantiles, and the 90% confidence intervals exclude zero around the median,

---

[17]All inference is based on 500 bootstrap repetitions.

[18]To be more specific, because the d and s statistics used to assess the existence of SD rankings are based on 500 points along the support, rather than just the 99 quantiles displayed in the plots, the SD tests capture crossings not shown in the figures.

supporting the quantity-quality trade-off. In terms of the statistical tests (Table 5), we now reject equality of the distributions only for weight-for-age (height-for-age: p = 0.220; weight-for-age: p = 0.026). In addition, we again fail to obtain a statistically meaningful SSD ranking in either case.

In sum, the unconditional results yield modest statistically meaningful evidence of a quantity-quality trade-off over at least some portions of the distribution using the more short-run measure of health (weight-for-age) whether one treats the number of children as exogenous or endogenous. However, we only find evidence of an impact of household size on the long-run measure (height-for-age) when we ignore endoeneity. The fact that the evidence of a trade-off is stronger when household size is treating as exogenous is consistent with negative selection into larger households. To further assess the sensitivity of these results, we turn to the results adjusting for covariates.

5.2   Adjusting for Covariates

The QTEs based on the distributions obtained using inverse propensity score weighting are displayed in Figures 3 (height-for-age) and 4 (weight-for-age). In each figure, the top row utilizes Control Set A, whereas the second (third) row uses Control Set B (C). In addition, the first column considers the treatment, *MoreThan2*, as exogenous, whereas the second column instruments for the treatment using *SameSex2*.

Treating the number of children as exogenous, we find that the point estimates of the QTEs are negative at every quantile for height-for-age (Figure 3), and negative over the majority of the distribution for weight-for-age (Figure 4). However, unlike the unconditional analysis, the estimates are rarely statistically significant, particularly when using Control Set B or C. Moreover, the lack of statistical significance is as much attributable to a fall (in absolute value) in the point estimates as due to a widening of the confidence intervals. Thus, household size appears strongly correlated with observable attributes associated with worse child health.

The corresponding test statistics are displayed in Table 6. In terms of the tests for equality, we reject the null of equality at conventional levels in all six cases. Thus, despite the fact that many of the QTEs are individually not statistically significant in Figures 3 and 4, we still easily reject the null of no effect of the number of the children under the selection on observables assumption.

19

In terms of making robust welfare statements, we fail to observe any SD ranking in the first- or second-order sense for height-for-age, and this is confirmed by the simple bootstrap. While we observe an SSD ranking in two cases using weight-for-age, none are statistically meaningful.

Turning to the IV results, there are several findings of note. First, although the point estimates of the QTEs are negative at nearly every quantile below roughly the 90th quantile in figures, the estimates are rarely statistically significant. The exception is for quantiles between then 20th and 50th quantiles when using weight-for-age. Thus, we again find modest statistically meaningful evidence of a quantity-quality trade-off for some when using a more short-run measure of health. Second, although the estimates are not statistically significant elsewhere, due to the wider confidence intervals further above the median, the QTEs do appear relatively uniform across the majority of the distributions. In light of theory, this uniformity suggests that the health production function is additively separable. Finally, in most cases, the confidence intervals for the IV QTEs are tighter compared with the estimates under exogeneity. Moreover, once we instrument for the treatment, the results are invariant to the control set utilized.

In terms of the test statistics (Table 6), two findings stand out. First, as in the unconditional case, we reject the equality of the distributions at conventional levels in all cases using weight-for-age and when using Control Set C using height-for-age. Second, while find no evidence of a statistically meaningful SSD ranking . As a result, when adjusting for covariates and allowing for the endogeneity of the number of children, we obtain little support for the trade-off over the *whole* distribution. Although, as in the unconditional analysis, evidence favoring the existence of the trade-off even in the case of weight-for age arises when one moves beyond a narrow focus on average effects (mostly concentrated in between 20th and 50th quantiles).

In sum, upon adjusting for covariates, we reach three conclusions. First, the distribution of weight-for-age continues to be statistically different. Such differences imply the existance of a trade-off at least over portions of the distribution when treating household size as endogenous. This contrasts starkly with the TSLS analysis which indicated no statistically meaningful trade-off on average. Second, failure to find any FSD or SSD welfare ranking implies that the trade-off is not robust across the whole distribution; the trade-off exists at most for some. Finally, the relative

uniformity of the point estimates of the quantile treatment effects when treating household size as endogenous is consonant with the health production function being additively separable in inputs and endowments.

5.3   Sensitivity to Sample Selection

As mentioned earlier, we restrict our sample to first and second born children since these are the children subjected to the quasi-experimental design. Nonetheless, we also performed our analysis on the full sample of all children, although the results based on first and second born children only constitute our preferred estimates. We continue to restrict the sample to children residing in households with at least one sibling. The results are relegated to Appendix B.

In the interest of brevity, we simply highlight the main differences. First, the negative (local) average treatment effect estimated by TSLS is now statstically significant when using weight-for-age and Control Set A; all other TSLS estimates remain statistically insignificant (see Table B2). Second, in Tables B4 (Unconditional Results) and B5 (Adjusted for Covariates), we now reject equality of the distributions in all cases when using height-for-age. Finally, in all cases in Tables B4 and B5, the distribution of weight-for-age in households with only two children is observed to second order dominate the corresponding distribution from larger households. Moreover, the rankings of the unconditional distributions (Table B4) are statistically meaningful.

In sum, while the majority of results – particularly those concerning the validity of our instrument – are unchaged when utilizing the full sample of children, we do find moderately stronger results pertaining to a quantity-quality trade-off for children between the 20th and 50th quantiles when using the more short-run measure of health (weight-for-age).

6   Conclusion

Although the theoretical trade-off between the quantity and quality of children is well-established, empirical evidence supporting such a causal relationship is limited. Moreover, most existing empirical studies focus on education as a measure of child quality and are limited to linear regression

analysis. While such results are easily interpretable, at best they provide evidence of the average effect of the number of children on child-specific investments. In contrast, this study analyzes the impact of household size on the entire distribution of two measures of health using data from Indonesia, while accounting for the potential endogeneity of the quantity of children.

The analysis yields three main conclusions. First, distributions of weight-for-age are statistically different when we account for the endogeneity of the quantity of children. Second, despite the significant difference in the distributions across households with more than two or only two children, we fail to find statistically meaningful evidence of the quantity-quality trade-off over the whole distribution. In particular, using an identification strategy based on gender composition of the first two children, there is evidence that the quantity-quality trade-off applies to only *some*. Finally, while modest evidence of the trade-off is found when using a more short-run measure of health (based on weight), there is no statistically meaningful evidence of a trade-off on a long-run measure (based on height) once the endogeneity of fertility decisions is addressed.

While these findings are striking, future research is necessary to answer questions generated by this analysis. First, how robust are the results to alternative instruments that identify the trade-off from other subpopulations of compliers (e.g., twins, such as in Rosenzweig and Zhang (2006))? Second, utilizing other instruments for identification, are there larger gains from reducing the number of children from, say, six children in a family to five, or from two children to only one? Finally, are robust rankings possible if one examines bivariate distributions, such as health and education, or health and family income? Despite these open questions, answering these and other similar questions within a distributional framework is necessary for a deeper understanding of the nature of intrahousehold allocation as well as sound policymaking.

References

[1] Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284-292.

[2] Angrist, J.D. and W. Evans (1998), "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 88, 450-477.

[3] Angrist, J.D., V. Lavy, and A. Schlosser (2006), "New Evidence on the Causal Link Between the Quantity and Quality of Children," IZA Discussion Paper No. 2075.

[4] Becker, G.S. and H.G. Lewis (1973), "On the Interaction Between the Quantity and Quality of Children," *Journal of Political Economy*, 82, S279-88.

[5] Becker, G.S. and N. Tomes (1976), "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84, S143-62.

[6] Bitler, M.P., J.B. Gelbach, and H.W. Hoynes (2006), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988-1012.

[7] Black, S.E., P.J. Devereux, and K.G. Salvanes (2005), "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, 120, 669-700.

[8] Butcher, K.F. and A. Case (1994), "The Effect of Sibling Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, 109, 531-63.

[9] Chernozhukov, V.and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

[10] Cogill, B. (2003), *Anthropometric Indicators Measurement Guide*, Food and Nutrition Technical Assistance Project, Academy for Educational Development, Washington, D.C.

[11] Conley, D. and R. Glauber (2005), "Parental Education Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Changes in Fertility," NBER Working Paper 11302.

[12] Conley, T.G., C. Hansen, and P.E. Rossi (2008), "Plausibly Exogenous," unpublished manuscript, University of Chicago Business School.

[13] Cruces, G. and S. Galiani (2007), "Fertility and Female Labor Supply in Latin America: New Causal Evidence," *Labour Economics*, 14, 565-573.

[14] Firpo, S. (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259-276.

[15] FitzGerald, V. (2001), "Financial Globalisation and Child Wellbeing," in G.A. Cronia (ed.) *Harnessing Globalisation for Children: A Report to UNICEF*, http://www.unicef-icdc.org/research/ESP/globalization/.

[16] Fogel, R. (1994), "Economic Growth, Population Theory and Physiology: the Bearing of Long-Term Processes on the Making of Economic Policy," *American Economic Review*, 84, 369-395.

[17] Glick, P., A. Marini, and D.E. Sahn (2007), "Estimating the Consequences of Changes in Fertility on Child Health and Education in Romania: an Exercise Using Twins Data," *Oxford Bulletin of Economics & Statistics*, 69, 667-691.

[18] Heckman, J., J. Smith, with the assistance of N. Clements (1997) "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

[19] Henderson, D.J., D.L. Millimet, C.F. Parmeter, and L. Wang (2008), "Fertility and the Health of Children: A Nonparametric Investigation," *Advances in Econometrics*, 21, 167-195.

[20] Imai, K. and D.A. van Dyk (2004), "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association*, 99, 854-866.

[21] Imbens, G.W and D.B. Rubin (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economic Studies*, 64, 555-574.

[22] Jensen, R. T. (2002), "Fertility Preferences and Female Disadvantage: Equal Treatment, Unequal Outcomes?" unpublished manuscript, John F. Kennedy School of Government, Harvard University.

[23] Kevane, M. and D. I. Levine (2001), "The Changing Status of Daughters in Indonesia," University of California, Berkeley Institute of Industrial Relations Working Paper No. 74.

[24] Lee, J. (2008), "Sibling Size and Investment in Children's Education: An Asian Instrument," *Journal of Population Economics*, 21, 855-875.

[25] Levine, D.I. and M. Ames (2003), "Gender Bias and The Indonesian Financial Crisis: Were Girls Hit Hardest?" University of California, Berkeley Center for International and Development Economics Research (CIDER) Working Paper No. C03-130.

[26] Linton, O., E. Maasoumi, and Y.J. Whang (2005), "Consistent Testing for Stochastic Dominance: A Subsampling Approach," *Review of Economic Studies*, 72, 735-765.

[27] López-Casasnovas, G. , B. Rivera, and L. Currais (2005), *Health and Economic Growth: Findings and Policy Implications*, Cambridge, MA: MIT Press.

[28] Maasoumi, E. and A. Heshmati (2000), "Stochastic Dominance Amongst Swedish Income Distributions," *Econometric Reviews*, 19, 287-320.

[29] Rosenzweig, M.R. and K.I. Wolpin (1980), "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48, 227-240.

[30] Rosenzweig, M.R. and K.I. Wolpin (2000), "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38, 827-874.

[31] Rosenzweig, M.R. and J. Zhang (2006), "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One Child' Policy," *Review of Economic Studies*, forthcoming.

[32] Rukumnuaykit, P. (2003), "Crises and Child Health Outcomes: The Impacts of Economic and Drought/Smoke Crises on Infant Mortality and Birthweight in Indonesia," unpublished manuscript, Michigan State University.

[33] Qian, N. (2008), "Quantity-Quality and the One Child Policy: The Only-Child Disadvantage on School Enrollment in Rural China," unpublished manuscript, Department of Economics, Brown University.

[34] Soeradji, B. and S.H. Hatmadji (1994), "Issues on Sex Preference: The Case of Indonesia," presented at the International Symposium on Issues Related to Sex Preference for Children in the Rapidly Changing Demographic Dynamics in Asia, Seoul, Republic of Korea, 21-24 November 1994.

[35] Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawatt, and F. Witoelar (2004a), "The Third Wave of the Indonesia Family Life Survey: Overview and Field Report," Rand Labor and Population Working Paper series.

[36] Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawatt, and F. Witoelar (2004b), "User's Guide for the Indonesia Family Life Survey, Wave 3," Rand Labor and Population Working Paper series.

[37] Strauss, J. and D. Thomas (1995), Human Resources: Empirical Modeling of Household and Family Decisions, Chapter 34. in J. Behrman and T.N. Srinivasan (eds.) *Handbook of Development Economics*, Vol. 3B, Elsevier Science B.V.

[38] Strauss, J. and D. Thomas (1998), "Health, Nutrition and Economic Development," *Journal of Economic Literature*, 36, 766-817.

[39] Thomas, D. and E. Frankenberg (2002), "Health, Nutrition and Prosperity: a Microeconomic Perspective," *Bulletin of the World Health Organization*, 80, 106-113.

[40] Thomas, D., J. Strauss, M. Henriques (1990), "Child Survival, Height for Age and Household Characteristics in Brazil," *Journal of Development Economics*, 33, 197-234.

[41] Thomas, D., J. Strauss, and M. Henriques (1991), "How Does Mother's Education Affect Child Height?," *Journal of Human Resources,* 26, 183-211.

[42] UNICEF (2005), *The State of the World's Children 2006: Excluded and Invisible*, New York, NY: UNICEF.

[43] Weidemann, A. (1999), "The Indonesian Family Planning Programme and the International Conference on Population and Development: A Comparative Follow-up Field Study," PROP Report No. 26, Lund University

[44] Weil, D. (2005), "Accounting for the Effect of Health on Economic Growth," NBER Working Paper No. 11455.

[45] Wongboonsin, K. and V.P. Ruffolo (1995), "Sex Preference for Children in Thailand and Some Other South-East Asian Countries," *Asia-Pacific Population Journal*, 10, 43-62.

Table 1: SUMMARY STATISTICS

Variable	More than Two Children			Two Children		
	Mean	SD	Obs	Mean	SD	Obs
Height-for-age (z-score)	-1.843	1.119	810	-1.484	1.335	2103
Weight-for-age (z-score)	-1.822	1.198	796	-1.545	1.344	2115
First two children are same sex (1 = yes)	0.514	0.5	1014	0.48	0.5	2594
First two children are daughter2 (1 = yes)	0.266	0.442	1014	0.229	0.421	2594
Age in months	93.677	27.061	966	62.929	37.128	2522
Gender (1=male)	0.499	0.5	1014	0.515	0.5	2594
First Child's Gender (1 = male)	0.492	0.5	1014	0.494	0.5	2594
Second Child's Gender (1 = male)	0.489	0.5	1014	0.527	0.499	2594
Father's Education						
Elementary School and below	0.427	0.495	1014	0.412	0.492	2594
Junior High School	0.17	0.375	1014	0.177	0.381	2594
Senior High School	0.259	0.439	1014	0.287	0.452	2594
University	0.144	0.351	1014	0.125	0.33	2594
Mother's Education						
Elementary School and below	0.527	0.5	1008	0.447	0.497	2574
Junior High School	0.157	0.364	1008	0.196	0.397	2574
Senior High School	0.229	0.421	1008	0.278	0.448	2574
University	0.087	0.282	1008	0.08	0.271	2574
Father's Height	161.997	5.972	749	162.428	6.47	2011
Mother's Height	150.968	5.36	849	151.01	5.363	2180
Father's Weight	57.161	9.83	745	57.398	9.801	2005
Mother's Weight	51.568	9.491	846	51.745	8.988	2172
Father's Age	36.705	5.543	1014	34.878	6.48	2588
Mother's Age	31.996	4.495	1008	30.085	5.318	2568
Father's Religion (1 = Islam)	0.853	0.354	1014	0.891	0.312	2592
Mother's Religion (1 = Islam)	0.854	0.353	1008	0.893	0.309	2572
Father's Work Status (1 = Wrok)	0.988	0.108	1014	0.984	0.125	2594
Mother's Work Status (1 = Wrok)	0.481	0.5	1008	0.486	0.5	2572
Region (1 = Urban)	0.485	0.5	1014	0.507	0.5	2594
Dweilling Type						
Single Unit	0.777	0.416	1014	0.836	0.37	2594
Duplex	0.087	0.282	1014	0.079	0.269	2594
Multiple Unit	0.04	0.197	1014	0.035	0.185	2594
House on Stilts	0.096	0.294	1014	0.05	0.217	2594
Floor Type						
Ceramic/Marble/Granite/Stone	0.162	0.369	1012	0.188	0.391	2594
Tiles/Terrazzo	0.186	0.389	1012	0.223	0.416	2594

Table 1 − (cont.) Summary Statistics

Variable	More than Two Children			Two Children		
	Mean	SD	Obs	Mean	SD	Obs
Cement/Bricks	0.373	0.484	1012	0.354	0.478	2594
Lumber/Board/Bamboo	0.182	0.386	1012	0.122	0.328	2594
Dirt	0.098	0.297	1012	0.113	0.316	2594
Wall Type						
Masonry (cement/brick)	0.609	0.488	1014	0.665	0.472	2594
Lumber/Board	0.271	0.445	1014	0.247	0.432	2594
Bamboo/Woven/Mat	0.119	0.324	1014	0.088	0.283	2594
Water Type						
Pipe water	0.284	0.451	1006	0.266	0.442	2569
Well/Pump	0.262	0.44	1006	0.313	0.464	2569
Well water	0.308	0.462	1006	0.28	0.449	2569
Spring/Rain Water	0.056	0.229	1006	0.069	0.253	2569
River/Creek Water	0.051	0.219	1006	0.028	0.165	2569
Other	0.039	0.193	1006	0.044	0.204	2569
Decision on Children's Health						
Jointly made by husband and wife	0.025	0.155	889	0.032	0.175	2243
Only made by husband	0.126	0.332	889	0.117	0.322	2243
Only made by wife	0.094	0.293	889	0.088	0.283	2243
Otherwise	0.755	0.43	889	0.763	0.425	2243
House surrounded by						
human/animal waste (1 = yes)	0.085	0.279	1014	0.068	0.252	2594
House surrounded by						
piles of trash (1 = yes)	0.116	0.321	1014	0.119	0.324	2594
House surrounded by						
stagnant water (1 = yes)	0.106	0.307	1014	0.097	0.296	2594
Stable under/next to						
house (1 = yes)	0.217	0.412	1014	0.197	0.398	2594
Sufficient ventilation (1 = yes)	0.769	0.422	1012	0.801	0.399	2594
Yard is cleaned up (1 = yes)	0.681	0.466	1014	0.718	0.45	2594
House has a moderately						
sized yard (1 = yes)	0.61	0.488	1014	0.616	0.486	2594
House has kitchen						
outside (1 = yes)	0.271	0.445	1012	0.266	0.442	2594
No. of Rooms	5.386	2.639	1014	5.463	2.612	2594
House Size	75.536	67.265	1014	82.532	182.343	2594
Boil Water (1 = yes)	0.921	0.27	983	0.95	0.218	2499
Own Farm (1 = yes)	0.331	0.471	1006	0.299	0.458	2569

Table 1 − (cont.) Summary Statistics

Variable	More than Two Children			Two Children		
	Mean	SD	Obs	Mean	SD	Obs
Female Household Head (1 = yes)	0.033	0.178	1014	0.028	0.165	2594
Number of Health Facilities						
Hospitals	2.655	1.432	1014	2.797	1.59	2594
Integrated Health Post	8.888	5.588	1014	9.502	5.54	2594
Private Practice	27.86	9.191	1014	28.751	9.719	2594
Commnuity Health Centers	7.855	2.89	1014	8.209	3.543	2594
Traditional Practices	5.514	3.384	1014	5.381	3.054	2594
Minimum distance of Health Facilities						
Hospitals	17.012	22.689	1014	15.882	21.751	2594
Integrated Health Post	5.643	62.677	1014	3.986	48.03	2594
Private Practice	0.416	0.512	1014	0.438	0.602	2594
Commnuity Health Centers	1.137	1.385	1014	1.045	1.05	2594
Traditional Practices	0.775	1.821	1014	0.796	2.103	2594

Table 2: Regression Results

Control Set	Height-for-age			Weight-for-age		
	Model 1 (1) A	Model 2 (2) B	Model 3 (3) C	Model 1 (4) A	Model 2 (5) B	Model 3 (6) C
			Panel A: OLS Results			
Morethan 2	-0.207***	-0.123**	-0.114*	-0.190***	-0.117*	-0.110*
	[0.058]	[0.059]	[0.059]	[0.064]	[0.064]	[0.065]
			Panel B: IV Results			
Morethan 2	-0.294	0.170	0.312	-1.426	-0.474	-0.378
	[0.916]	[1.143]	[1.033]	[1.029]	[1.131]	[1.012]
			Panel C: Weak-IV Robust Inference			
Anderson-Rubin F	0.102	0.022	0.090	2.192	0.173	0.136
p-value	0.749	0.883	0.764	0.139	0.677	0.713
Stock-Wright S stat	0.102	0.022	0.093	2.192	0.178	0.140
p-value	0.749	0.882	0.760	0.139	0.673	0.708
			Panel D: Reduced Form			
Samesex2	-0.016	0.007	0.015	-0.078	-0.022	-0.019
	[0.050]	[0.049]	[0.049]	[0.053]	[0.052]	[0.052]
Number of Obs	2863	2399	2399	2861	2399	2399

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[2] Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

[3] Panel D reports reduced form estimates where morethan2 is excluded from these regressions.

Table 3: The Effects of IV on More Than 2 (First Stage)

	Height-for-age			Weight-for-age		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	(1)	(2)	(3)	(4)	(5)	(6)
Control Set	A	B	C	A	B	C
Same Sex 2	0.055***	0.043**	0.048***	0.055***	0.045***	0.051***
	[0.017]	[0.018]	[0.018]	[0.017]	[0.018]	[0.018]
F-test	10.872	5.928	7.318	10.880	6.727	8.310
p-value	0.001	0.015	0.007	0.001	0.010	0.004
Kleibergen-Paap rk Wald Stat	10.891	6.080	7.551	10.899	6.899	8.574
p-value	0.001	0.014	0.006	0.001	0.009	0.003
Number of Obs	2863	2399	2399	2861	2399	2399

1 Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
2 Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

Table 4: The Effects of Alternative IVs on Fertility (First Stage)

		Height-for-age		Weight-for-age	
		Model (1)	No. of obs (2)	Model (3)	No. of obs (4)
Panel A	Endog Var: Morethan 1 IV: Daughter 1	-0.014 [0.010]	3762	-0.014 [0.010]	3785
Panel B	Endog Var: Morethan 2 IV: Daughter 1	-0.027 [0.019]	2399	-0.028 [0.019]	2399
Panel C	Endog Var: Morethan 2 IV: Daughter 2	0.03 [0.025]	2399	0.032 [0.025]	2399

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[2] Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

Table 5: Distributional Tests of Unconditional Children's Health.

Distributions		Observed Ranking	Test of Equality	First Order Dominance				Second Order Dominance			
X	Y			$d_{1,MAX}$	$d_{2,MAX}$	d	$Prd^* \leq 0$ (Simple Boot)	$s_{1,MAX}$	$s_{2,MAX}$	s	$Prs^* \leq 0$ (Simple Boot)
No IV											
A. Height-for-Age											
Two Children	More than Two	None	p=0.000	0.038	3.048	0.038	0.056	0.137	222.195	0.137	0.134
B. Weight-for-Age											
Two Children	More than Two	None	p=0.014	0.018	2.216	0.018	0.286	0.018	200.475	0.018	0.448
With IV											
A. Height-for-Age											
Two Children	More than Two	None	p=0.220	1.018	1.353	1.018	0.002	1.360	59.607	1.360	0.030
B. Weight-for-Age											
Two Children	More than Two	None	p=0.026	0.131	1.891	0.131	0.002	0.177	107.394	0.177	0.184
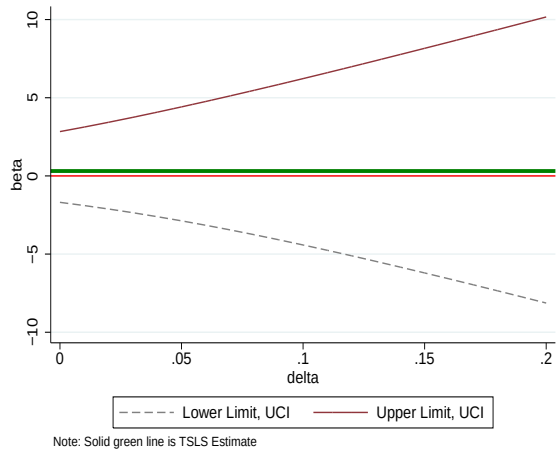
[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
[2] P-values obtained using 500 bootstrap repetitions.

Table 6: Distributional Tests of Children's Health Adjusted for Covariates
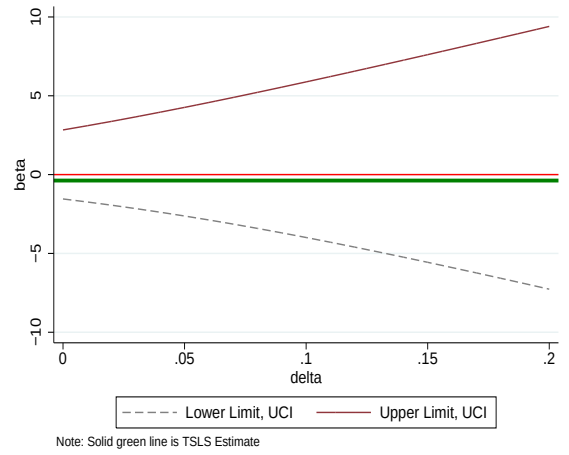
Distributions		Control Set	Observed Ranking	Test of Equality	First Order Dominance				Second Order Dominance			
X	Y				$d_{1,MAX}$	$d_{2,MAX}$	d	$Prd^* \leq 0$ (Simple Boot)	$s_{1,MAX}$	$s_{2,MAX}$	s	$Prd^* \leq 0$ (Simple Boot)
No IV												
A. Height-for-Age												
Two Children	More than Two	A	None	p=0.000	0.048	2.719	0.048	0.022	0.194	165.807	0.194	0.092
Two Children	More than Two	B	None	p=0.000	0.054	3.729	0.054	0.010	0.274	257.775	0.274	0.074
Two Children	More than Two	C	None	p=0.000	0.052	3.076	0.052	0.012	0.257	201.350	0.257	0.072
B. Weight-for-Age												
Two Children	More than Two	A	None	p=0.004	0.020	2.335	0.020	0.252	0.020	195.729	0.020	0.448
Two Children	More than Two	B	X SSD Y	p=0.012	1.219	1.948	1.219	0.076	0.000	112.883	0.000	0.564
Two Children	More than Two	C	X SSD Y	p=0.008	1.013	1.980	1.013	0.062	0.000	116.933	0.000	0.594
With IV												
A. Height-for-Age												
Two Children	More than Two	A	None	p=0.148	0.901	1.470	0.901	0.000	1.473	68.956	1.473	0.024
Two Children	More than Two	B	None	p=0.124	0.951	1.418	0.951	0.000	2.411	49.879	2.411	0.026
Two Children	More than Two	C	None	p=0.068	0.971	1.539	0.971	0.000	2.339	49.618	2.339	0.028
B. Weight-for-Age												
Two Children	More than Two	A	None	p=0.012	0.114	2.028	0.114	0.004	0.170	120.190	0.170	0.204
Two Children	More than Two	B	None	p=0.046	0.383	1.716	0.383	0.000	1.369	64.454	1.369	0.084
Two Children	More than Two	C	None	p=0.052	0.496	1.754	0.496	0.000	1.374	63.689	1.374	0.102

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[2] P-values obtained using 500 bootstrap repetitions.
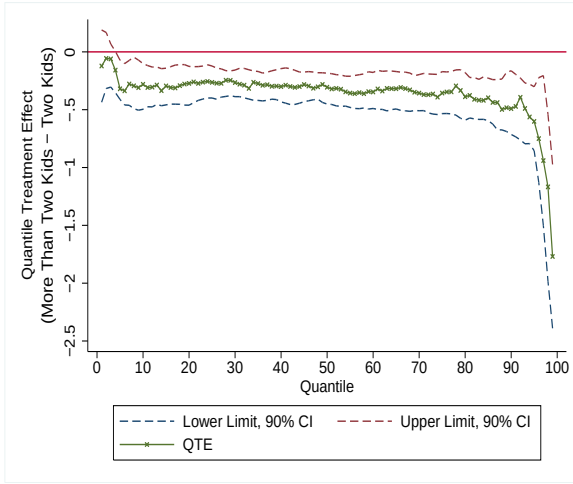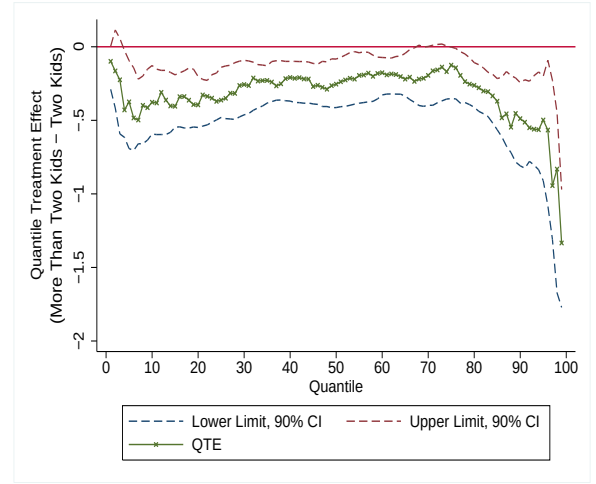
(a) Height-for-age

(b) Weight-for-age

Figure 1: Sensitivity Analysis of Instrument Variable
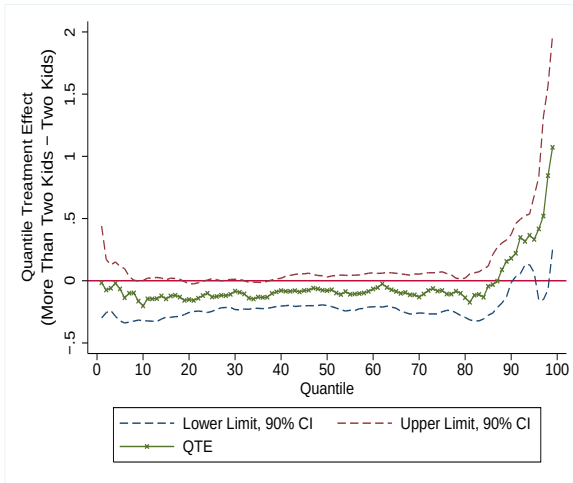
Note: Control set C is utilized.
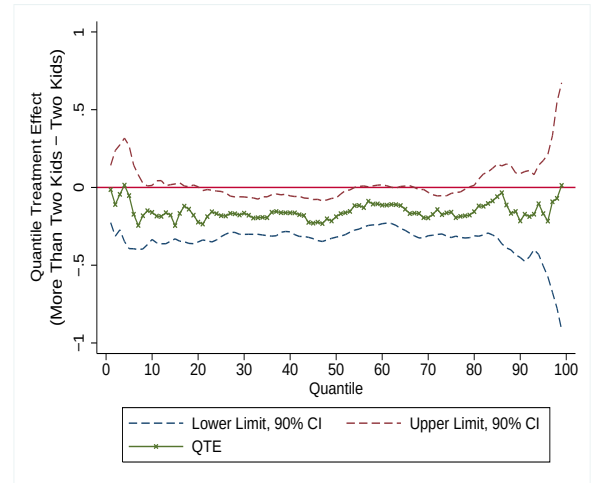
(a) Height-for-age   (b) Weight-for-age
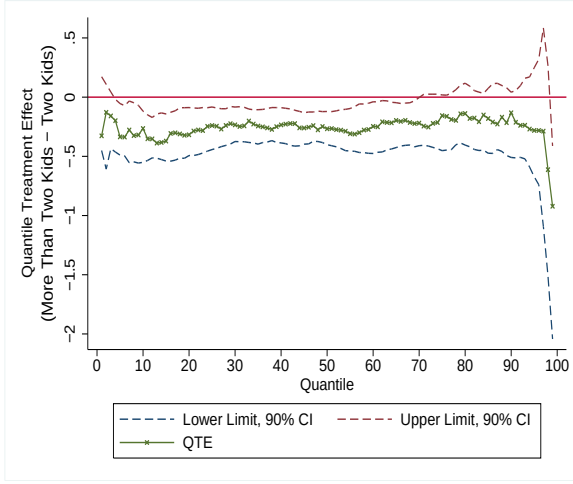
No Instrument
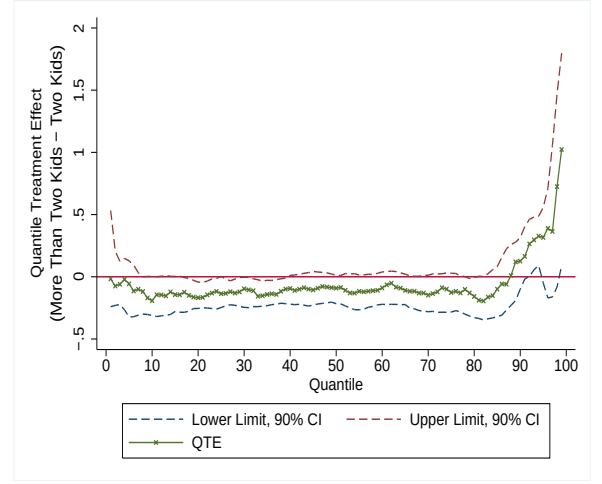
(c) Height-for-age   (d) Weight-for-age

Instrument: SameSex2

Figure 2: Unconditional Quantile Treatment Effects
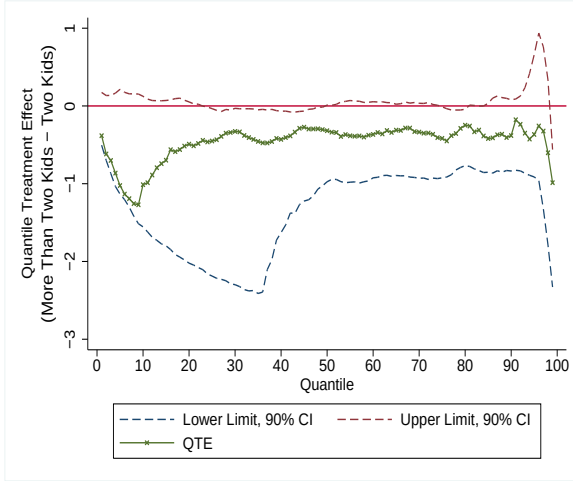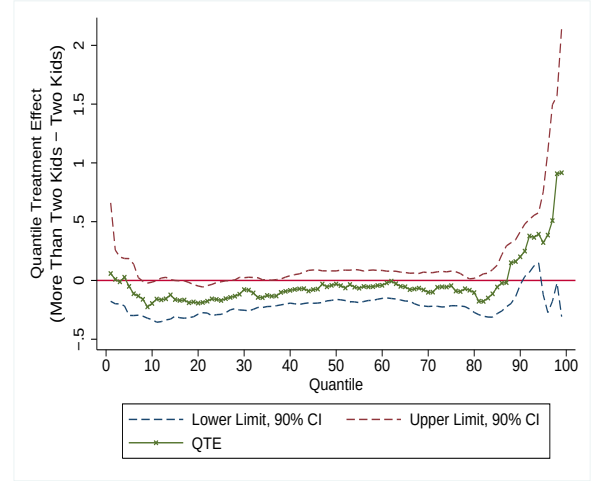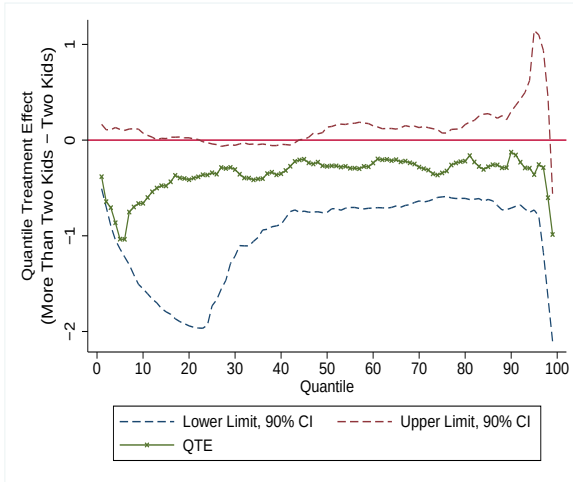
(a) No Instrument

(b) Instrument

Control Set A
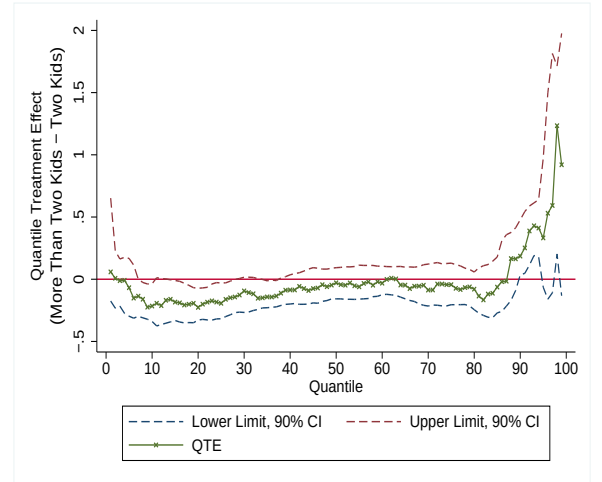
(c) No Instrument

(d) Instrument
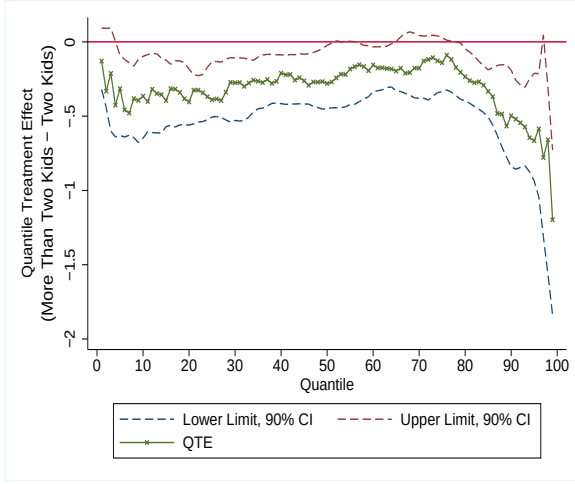
Control Set B
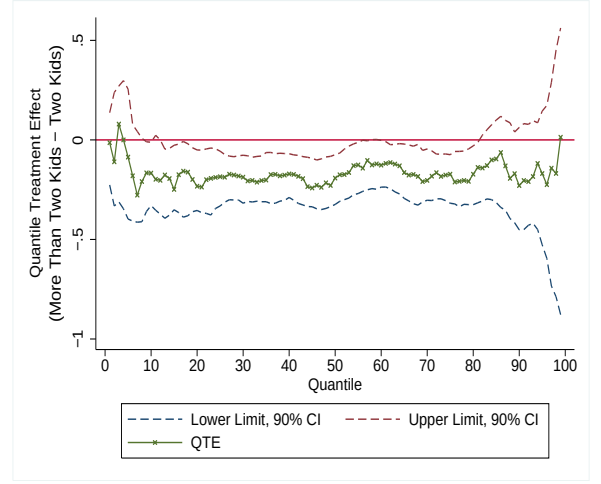
(e) No Instrument

(f) Instrument

Control Set C

Figure 3: Quantile Treatment Effects Adjusted for Covariates: Height-for-age z-scores
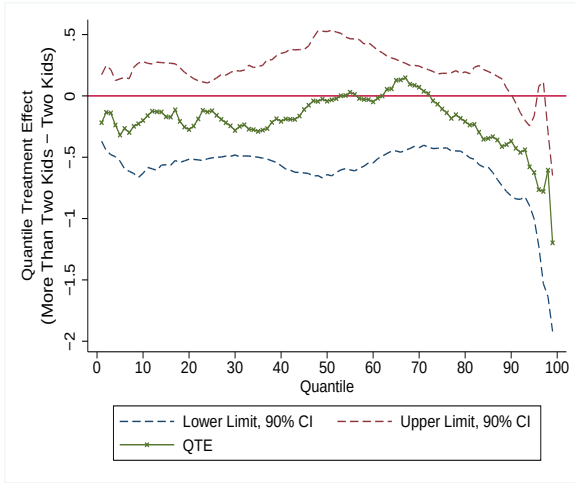
(a) No Instrument   (b) Instrument
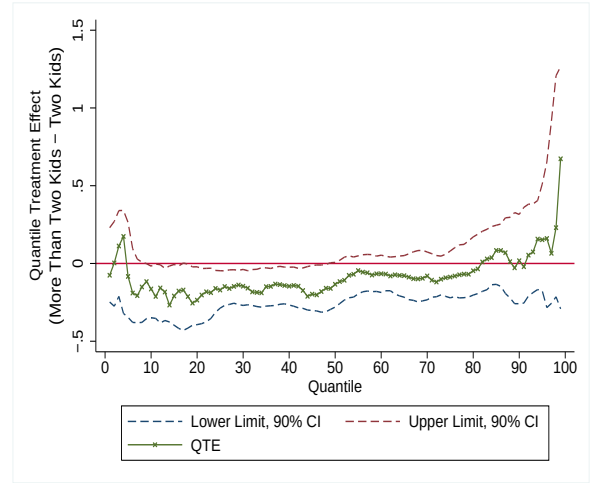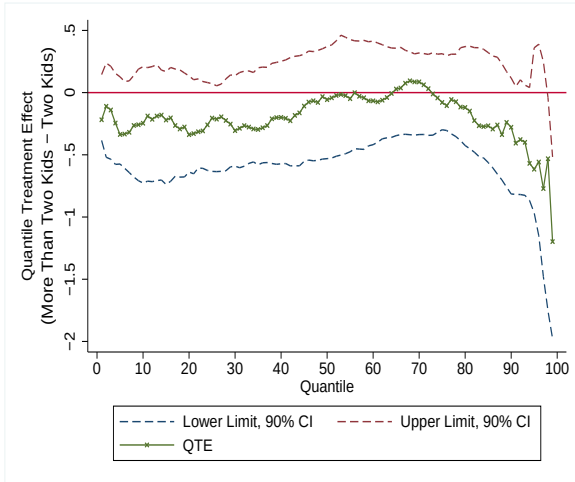
Control Set A
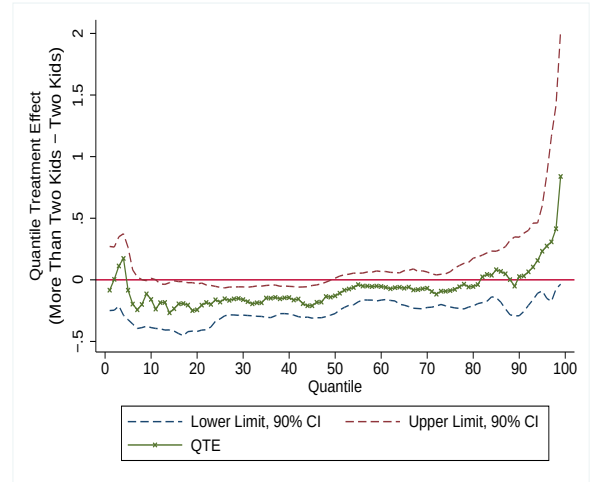
(c) No Instrument   (d) Instrument

Control Set B

(e) No Instrument   (f) Instrument

Control Set C

Figure 4: Quantile Treatment Effects Adjusted for Covariates: Weight-for-age z-scores

A   Econometric Details

A.1   Calculation and Inference of the KS Test of Equality

The test is based on the following KS statistic:

$$d_{eq} = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \sup |F_1 - F_0| \tag{12}$$

Specifically, our procedure calls for:

(i) obtaining the empirical CDF for Q_0 and Q_1, defined as

$$\widehat{F}_{jN_j}(q) = \frac{1}{N_j} \sum_{i=1}^{N_j} \mathrm{I}(Q_j \leq q), \qquad j = 0, 1 \tag{13}$$

by computing the values of $\widehat{F}_{0N_0}(q_k)$ and $\widehat{F}_{1N_1}(q_k)$, where $\mathrm{I}(\cdot)$ is an indicator function and q_k, $k = 1, \ldots, K$, denotes points in the support that are utilized ($K = 500$ in the application),

(ii) and computing

$$\widehat{d}_{eq} = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \max_k \{|\widehat{F}_1(q_k) - \widehat{F}_0(q_k)|\} \tag{14}$$

Inference is conducted using the bootstrap procedure applied in Abadie (2002). Specifically, we pool the two samples, resample (with replacement) from the combined sample, split the new sample into two samples, where the first N_0 represent Q_0 and the remainder represent Q_1, and compute the KS statistic. This process is repeated B times, and the p-value is given by

$$p - value = \frac{1}{B} \sum_{b=1}^{B} \mathrm{I}(\widehat{d}_{eq}^* > \widehat{d}_{eq}) \tag{15}$$

The null hypothesis is rejected if the p-value is less than the desired significance level, say 0.10.

A.2 Implementation of the Inverse Propensity Score Weighting Procedure

To control for covariates, we utilize the inverse propensity score weighting procedure as applied in Bitler et al. (2006) (see also Firpo 2007). This entails altering the method for estimating the empirical CDFs. Specifically, the empirical CDF for Q_j, $j = 0, 1$, is now computed as

$$\widehat{F}_{jN_j}(q) = \frac{\sum_{i=1}^{N_j} \widehat{\omega}_i \, \mathrm{I}(Q_j \leq q)}{\sum_{i=1}^{N_j} \widehat{\omega}_i} \tag{16}$$

where the weights, $\widehat{\omega}_i$, are given by

$$\widehat{\omega}_i = \frac{D_i}{\widehat{p}_i(X_i)} + \frac{1 - D_i}{1 - \widehat{p}_i(X_i)} \tag{17}$$

and $\widehat{p}_i(X_i)$ is the propensity score (i.e., the predicted probability that D_i equals one given the observables, X_i, from a first-stage probit model). Inference is conducted using the same bootstrap procedure discussed above. The only difference is that the first-stage probit model, and resulting weights, are estimated anew during each bootstrap replication.

A.3 Implementation of Stochastic Dominance Tests

To test for FSD and SSD, we use the following generalizations of the Kolmogorov-Smirnov test criteria:

$$d = \sqrt{\frac{N_0 N_1}{N + N_1}} \min \sup_{z \in \mathcal{Z}} [F_0(z) - F_1(z)] \tag{18}$$

$$s = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min \sup_{z \in \mathcal{Z}} \int_{-\infty}^{z} [F_0(v) - F_1(v)] \, dv \tag{19}$$

where min is taken over $F_0 - F_1$ and $F_1 - F_0$, in effect performing two tests in order to leave no ambiguity between the equal and unrankable cases. Specifically, our procedure calls for:

(i) computing the empirical CDFs using either (13) and (16), depending on if one wishes to adjust

for covariates, at q_k, $k = 1, \ldots, K$,

(ii) computing the differences $d_1(q_k) = \widehat{F}_{0N_0}(q_k) - \widehat{F}_{1N_1}(q_k)$ and $d_2(z_j) = \widehat{F}_{1N_1}(q_k) - \widehat{F}_{0N_0}(q_k)$,

(iii) obtaining $\widehat{d} = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min\{\max\{d_1\}, \max\{d_2\}\}$,

(iv) calculating the sums $s_{1j} = \sum_{k=1}^{j} d_1(q_k)$ and $s_{2j} = \sum_{k=1}^{j} d_2(q_k)$, $j = 1, ..., J$, and

(v) obtaining $\widehat{s} = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min\{\max\{s_{1j}\}, \max\{s_{2j}\}\}$.

If $\widehat{d} \leq 0$ and $\max\{d_1\} < 0$, then Q_0 is observed to first-order dominate Q_1; if $\widehat{d} \leq 0$ and $\max\{d_2\} < 0$, then the reverse is observed. If $\widehat{d} > 0$, then there is no observed ranking in the first-order sense. Similar interpretations are given to $\widehat{s}$, $\max\{s_{1j}\}$, $\max\{s_{2j}\}$ with respect to second order dominance.

There exist two inferential approaches based on different bootstrap procedures to evaluate the null of FSD (SSD), which is equivalent to $H_o : d \leq 0$ ($H_o : s \leq 0$): equal bootstrap and simple bootstrap. The first follows Abadie (2002), and is identical to the approach described above for the test of equality. However, as noted in Linton et al. (2005), the boundary between the null and alternative hypotheses is much larger than the LFC region. As such, bootstrap-based tests imposing the LFC are not asymptotically similar on the boundary, implying that the test is biased. In particular, if $d = 0$ or $s = 0$ is true, but the LFC fails to hold, the test will not have the appropriate asymptotic size. Thus, we utilize the second procedure, following Maasoumi and Heshmati (2000). Here, one resamples (with replacement) from each individual sample, Q_0 and Q_1. Thus, this procedure does not impose the LFC (or any other portion of the null). Consequently, one does not form p-values using (15). Instead, under this resampling scheme, if $\Pr\{\widehat{d}^* \leq 0\}$ is large, say 0.90 or higher, and $\widehat{d} \leq 0$, one can infer FSD to a desirable degree of confidence. This is a classic confidence interval test; one is assessing the likelihood that the event $d \leq 0$ has occurred. $\Pr\{\widehat{s}^* \leq 0\}$ is interpreted in similar fashion.

B   Additional Results Using the Full Sample of All Children

Table B1: Regression Results

	Height-for-age			Weight-for-age		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	(1)	(2)	(3)	(4)	(5)	(6)
Control Set	A	B	C	A	B	C
	Panel A: OLS Results					
Morethan 2	-0.147***	-0.098**	-0.088*	-0.164***	-0.152***	-0.149***
	[0.048]	[0.047]	[0.047]	[0.050]	[0.050]	[0.050]
	Panel B: IV Results					
Morethan 2	-1.159	-0.990	-0.825	-2.191*	-1.621	-1.466
	[0.961]	[1.055]	[0.978]	[1.169]	[1.252]	[1.149]
	Panel C: Weak-IV Robust Inference					
Anderson-Rudin F	1.589	0.947	0.740	4.768	2.015	1.872
p-value	0.207	0.331	0.390	0.029	0.156	0.171
Stock-Wright S stat	1.590	0.959	0.753	4.764	2.041	1.902
p-value	0.207	0.327	0.386	0.029	0.153	0.168
	Panel D: Reduced Form					
Samesex 2	-0.048	-0.036	-0.032	-0.086**	-0.055	-0.053
	[0.038]	[0.037]	[0.037]	[0.039]	[0.039]	[0.039]
Number of Obs	5122	4355	4355	5140	4374	4374

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[2] Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

[3] Panel D reports reduced form estimates where morethan2 is excluded from these regressions.

Table B2: The Effects of IV on More Than 2 (First Stage)

	Height-for-Age				Weight-for-age			
	Model 1 (1)	Model 2 (2)	Model 3 (3)	Model 4 (4)	Model 1 (5)	Model 2 (6)	Model 3 (7)	Model 4 (8)
	A. Coefficients of Instrument Variables							
Same Sex 2	0.059*** [0.015]	0.041*** [0.012]	0.036*** [0.013]	0.038*** [0.013]	0.055*** [0.015]	0.039*** [0.012]	0.034*** [0.012]	0.036*** [0.012]
	B. Identification Tests							
F-test	15.254	12.303	8.349	9.426	13.551	11.127	7.444	8.429
p-value	0.000	0.000	0.004	0.002	0.000	0.001	0.006	0.004
Kleibergen-Paap rk Wald Stat	15.259	12.315	8.465	9.589	13.556	11.138	7.547	8.574
p-value	0.000	0.000	0.004	0.002	0.000	0.001	0.006	0.003
Number of Obs	5122	5122	4355	4355	5140	5140	4374	4374

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[2] Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

Table B3: The Effects of Alternative IVs on Fertility (First Stage)

		Height-for-age		Weight-for-age	
		Model (1)	No. of obs (2)	Model (3)	No. of obs (4)
Panel A	Endog Var: Morethan 1 IV: Daughter 1	-0.007 [0.009]	5718	-0.007 [0.009]	5760
Panel B	Endog Var: Morethan 2 IV: Daughter 1	-0.011 [0.013]	4355	-0.013 [0.013]	4374
Panel C	Endog Var: Morethan 2 IV: Daughter 2	0.016 [0.015]	4355	0.013 [0.015]	4374

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[2] Columns (1)-(3) report results for Height-for-age using Control set A, B, and C, respectively; Columns (4)-(6) report results for Weight-for-age using Control set A, B, and C, respectively. See text for detail of the variables in each control set.

Table B4: Distributional Tests of Unconditional Children's Health.

Distributions		Observed Ranking	Test of Equality	First Order Dominance				Second Order Dominance			
X	Y			$d_{1,MAX}$	$d_{2,MAX}$	d	$Prd^* \leq 0$ (Simple Boot)	$s_{1,MAX}$	$s_{2,MAX}$	s	$Prd^* \leq 0$ (Simple Boot)
No IV											
A. Height-for-Age											
Two Children	More than Two	None	p=0.000	0.052	4.017	0.052	0.054	0.013	265.026	0.013	0.392
B. Weight-for-Age											
Two Children	More than Two	X SSD Y	p=0.000	0.014	3.507	0.014	0.342	0.000	285.013	0.000	0.962
With IV											
A. Height-for-Age											
Two Children	More than Two	None	p=0.072	0.479	1.692	0.479	0.000	0.434	80.841	0.434	0.210
B. Weight-for-Age											
Two Children	More than Two	X SSD Y	p=0.004	0.035	2.562	0.035	0.114	0.000	134.964	0.000	0.918

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
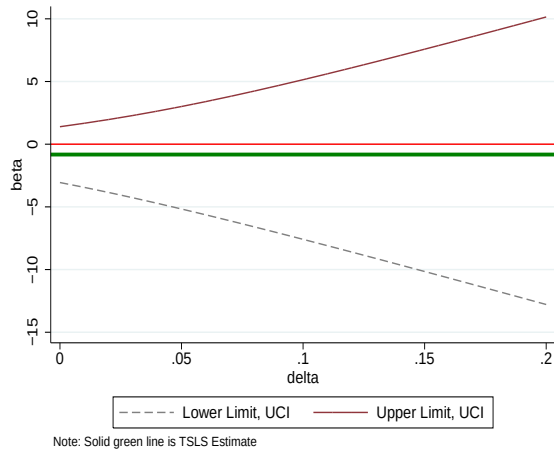
[2] P-values obtained using 500 bootstrap repetitions.

Table B5: Distributional Tests of Children's Health Adjusted for Covariates
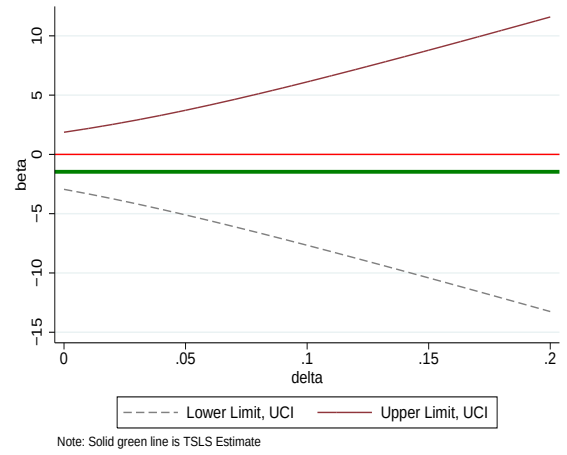
| Distributions | | Control Set | Observed Ranking | Test of Equality | First Order Dominance | | | | Second Order Dominance | | | |
X	Y				$d_{1,MAX}$	$d_{2,MAX}$	d	$Prd^* \leq 0$ (Simple Boot)	$s_{1,MAX}$	$s_{2,MAX}$	s	$Prd^* \leq 0$ (Simple Boot)
No IV												
A. Height-for-Age												
Two Children	More than Two	A	None	p=0.000	0.051	3.860	0.051	0.012	0.057	211.818	0.057	0.310
Two Children	More than Two	B	None	p=0.000	0.058	6.073	0.058	0.006	0.174	440.860	0.174	0.208
Two Children	More than Two	C	None	p=0.000	0.206	5.134	0.206	0.002	0.107	305.725	0.107	0.246
B. Weight-for-Age												
Two Children	More than Two	A	X SSD Y	p=0.000	0.000	3.802	0.000	0.120	0.000	267.474	0.000	0.870
Two Children	More than Two	B	X SSD Y	p=0.000	3.848	2.386	2.386	0.012	0.000	98.426	0.000	0.608
Two Children	More than Two	C	X SSD Y	p=0.000	3.395	3.252	3.252	0.010	0.000	116.151	0.000	0.686
With IV												
A. Height-for-Age												
Two Children	More than Two	A	None	p=0.084	0.437	1.689	0.437	0.000	0.482	83.107	0.482	0.196
Two Children	More than Two	B	None	p=0.062	0.549	1.666	0.549	0.000	0.200	68.606	0.200	0.268
Two Children	More than Two	C	None	p=0.018	0.639	1.752	0.639	0.000	0.238	72.708	0.238	0.254
B. Weight-for-Age												
Two Children	More than Two	A	X SSD Y	p=0.002	0.025	2.631	0.025	0.100	0.000	140.439	0.000	0.898
Two Children	More than Two	B	X SSD Y	p=0.012	0.191	2.140	0.191	0.002	0.000	93.558	0.000	0.800
Two Children	More than Two	C	X SSD Y	p=0.000	0.231	2.236	0.231	0.002	0.000	95.969	0.000	0.828

[1] Note – Robust standard errors in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
[2] P-values obtained using 500 bootstrap repetitions.
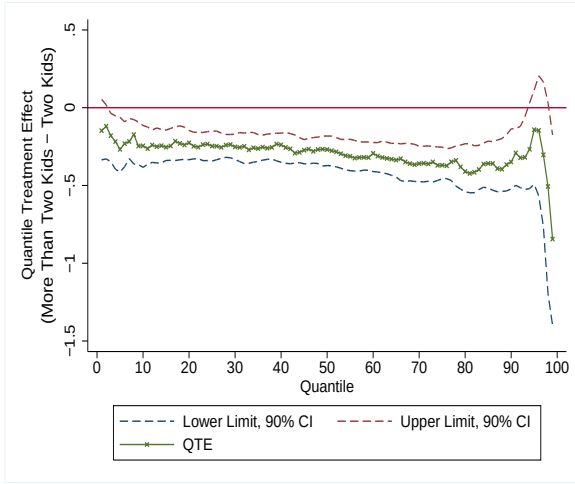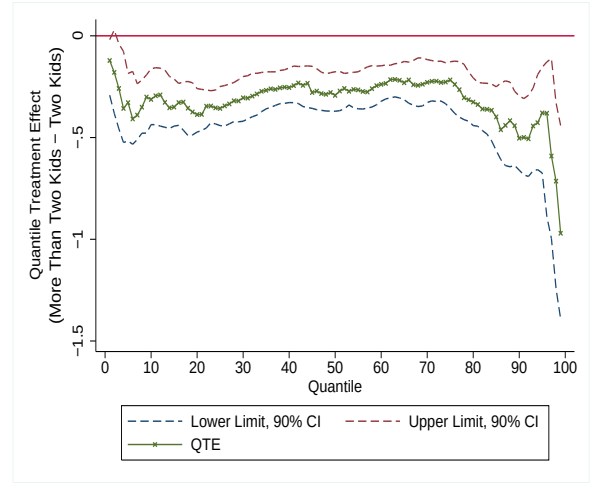
(a) Height-for-age   (b) Weight-for-age

Figure B1: Sensitivity Analysis of Instrument Variable
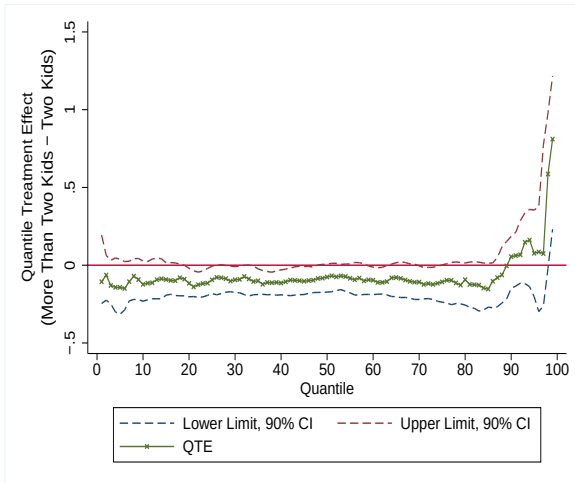
Note: Control set C is utilized.
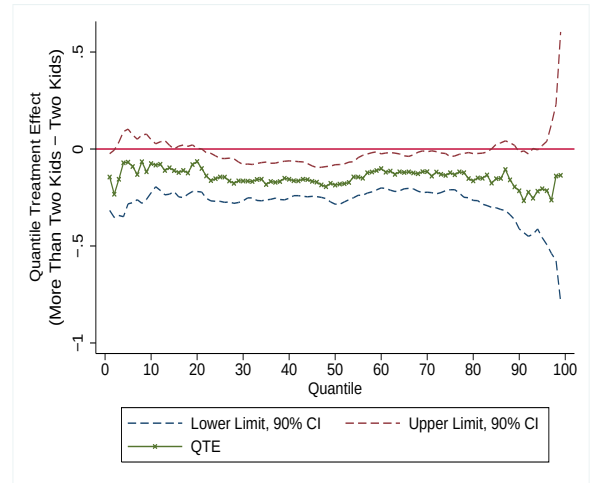
(a) Height-for-age

(b) Weight-for-age
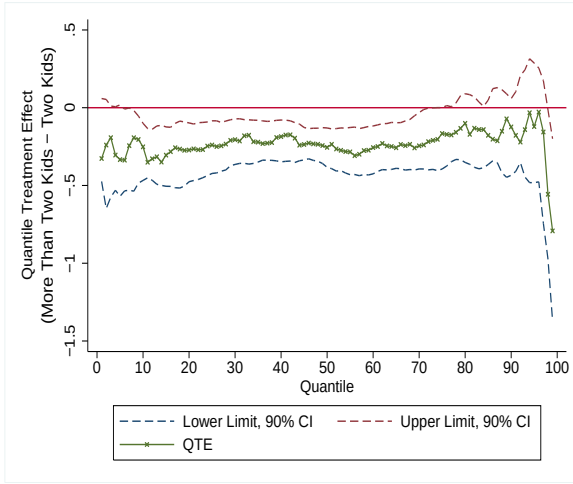
No Instrument
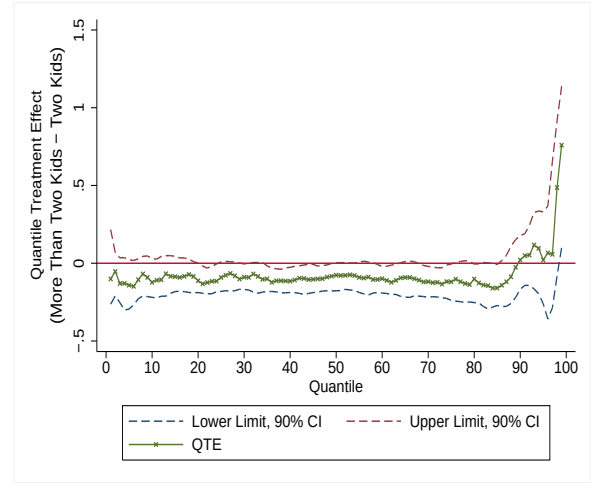
(c) Height-for-age

(d) Weight-for-age

Instrument: SameSex2

Figure B2: Unconditional Quantile Treatment Effects
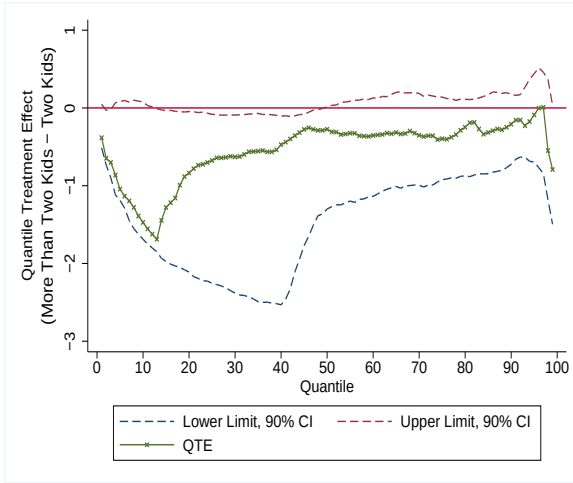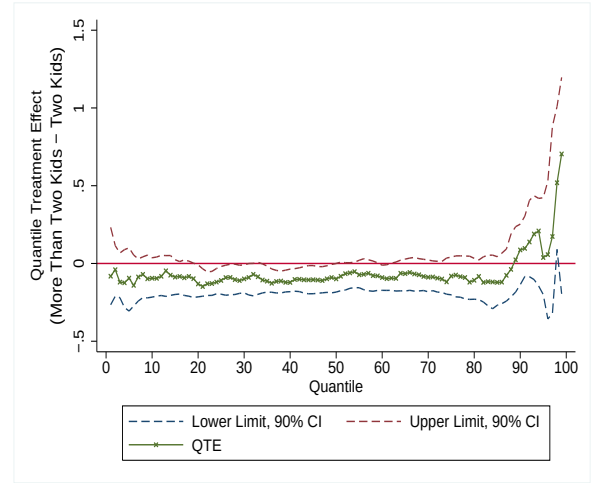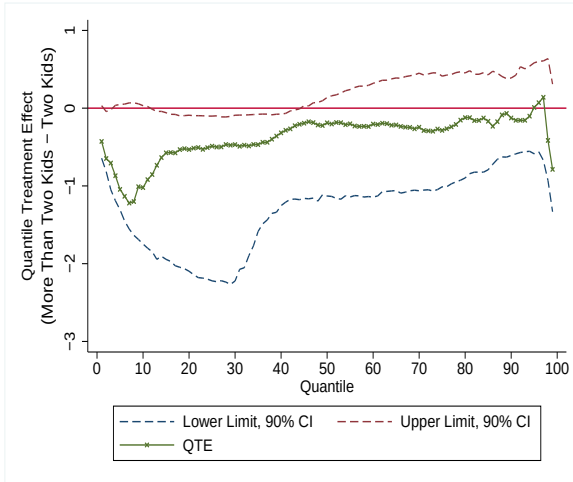
(a) No Instrument

(b) Instrument

Control Set A
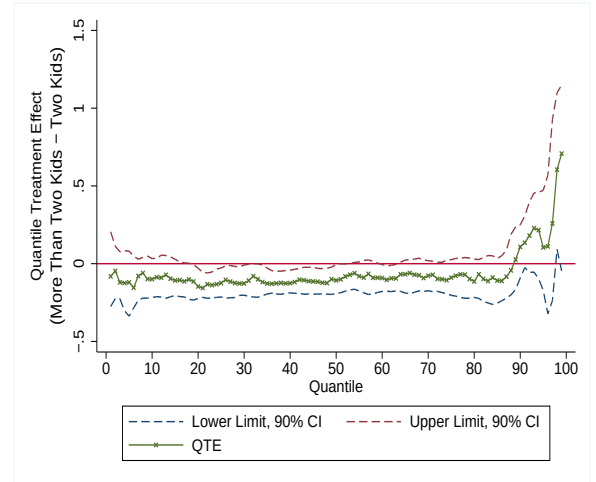
(c) No Instrument

(d) Instrument
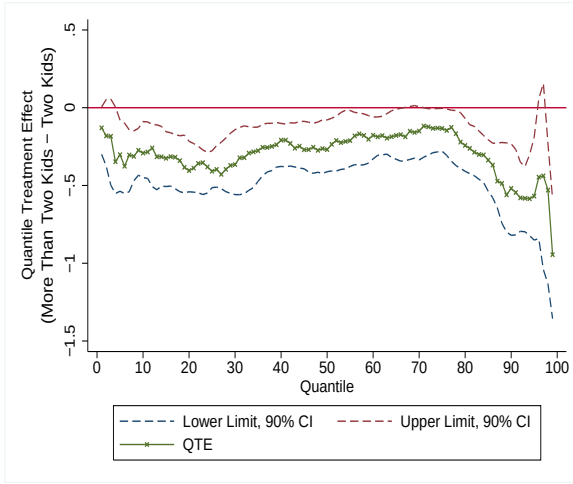
Control Set B
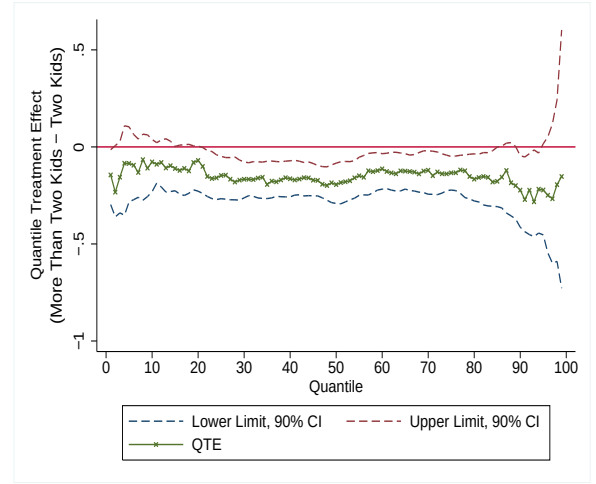
(e) No Instrument

(f) Instrument

Control Set C

Figure B3: Quantile Treatment Effects Adjusted for Covariates: Height-for-age z-scores
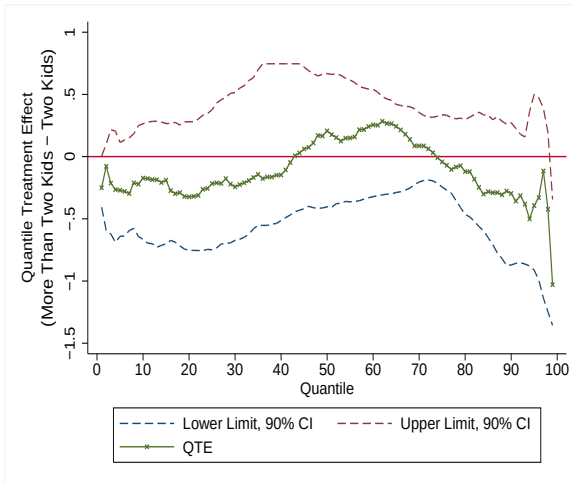
50
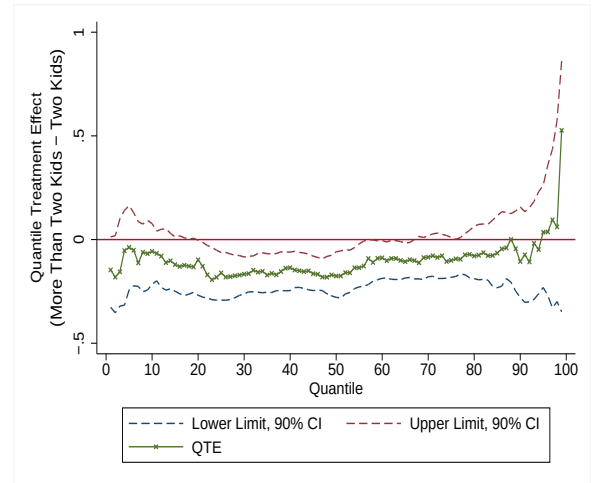
(a) No Instrument

(b) Instrument
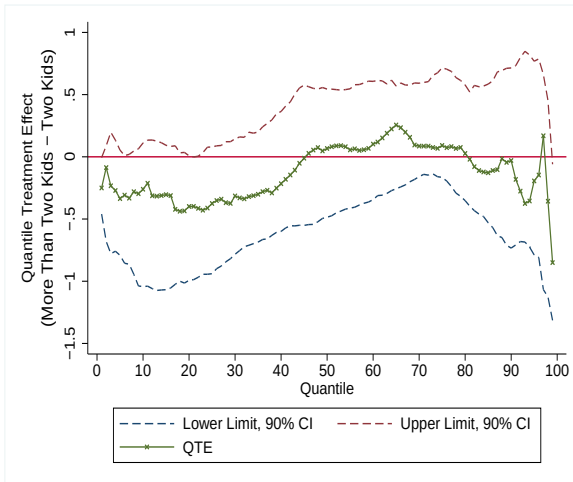
Control Set A
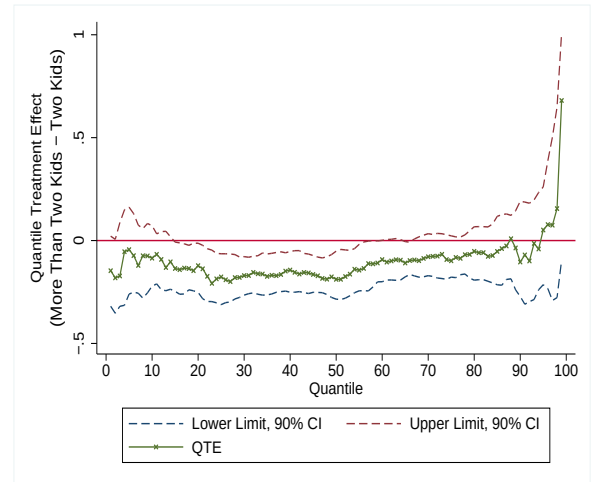
(c) No Instrument

(d) Instrument

Control Set B

(e) No Instrument

(f) Instrument

Control Set C

Figure B4: Quantile Treatment Effects Adjusted for Covariates: Weight-for-age z-scores