Is the Quantity-Quality Trade-off Really a Trade-off for All?

Daniel L. Millimet*

Southern Methodist University

Le Wang

Southern Methodist University

August 2005

Abstract

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, empirical studies that have been undertaken typically focus on education as a measure of child quality and have been predominantly limited to linear regression analysis, thereby focusing on the impact of the quantity of children on the (conditional) mean. In contrast, this paper uses two measures of child health as well as recently developed nonparametric tests for stochastic dominance to assess whether the quantity-quality trade-off holds across the entire distribution, or whether the benefits of smaller families are only experienced by some. Using data from the Indonesia Family Life Survey and controlling for the potential endogeneity of fertility, we find evidence that the trade-off exists over the majority of the distribution. However, robust rankings of distributions by sibship size are only possible if one accounts for 'dispersion' in child health. Moreover, the magnitude of the trade-off is not always uniform; individuals in the lower tail of the distribution may face a greater trade-off.

JEL: C14, D10, I12, O12

Keywords: Intrahousehold Allocation, Health, Human Capital, Fertility, Stochastic Dominance

*Corresponding address: Daniel Millimet, Department of Economics, Box 0496, Southern Methodist University, Dallas, TX 75275-0496. Tel: (214) 768-3269. Fax: (214) 768-1821. E-mail: millimet@mail.smu.edu.

1 Introduction

Researchers and policymakers have long suspected a negative relationship between child outcomes and the number of children in a household; the so-called *quantity-quality trade-off* (Becker and Tomes 1976; Becker and Lewis 1973; Willis 1973; Becker 1960).[1] The trade-off is typically assumed to originate from parental preferences for equal levels of quality across children (Rosenzweig and Wolpin 1980). Empirical tests of this model concentrate on estimating demand equations for child-specific outcomes (e.g., the level of schooling), where the number of children is one potential influence on demand. Such studies document a negative relationship between sibship size and human capital investments (see, e.g., Conley and Glauber 2005; Glick et al. 2005; Lee 2004; Ahn et al. 1998; Parish and Willis 1993; Hanushek 1992; Knodel and Wongsith 1991; Rosenzweig and Wolpin 1980), although a few find either no effect (Black et al. 2005; Kaestner 1997; Mock and Leslie 1984) or even a positive effect (Qian 2005; Hossain 1990; Chernochovsky 1985; Gomes 1984).

Empirical assessments of the quantity-quality trade-off have focused, however, mainly on one specific dimension of quality: education. While education is clearly an important component of child quality, another salient dimension, health, has received less rigorous analysis, presumably owing to the less frequent availability of anthropometric data. Nonetheless, researchers and policymakers are cognizant of the impact of child health on adult health and other economic outcomes. For example, Thomas and Frankenberg (2002) state that adult stature is largely determined during the fetal and early childhood periods, and Thomas et al. (1990, 1991) note the direct relationship between child anthropometric measures and the probability of survival as well as skill development. Other researchers have documented a positive association between adult health and labor market outcomes at both the microeconomic and macroeconomic levels. Specifically, many studies have found that there is a positive impact of height on individual hourly earning (Strauss and Thomas (1998) provide an excellent review), Fogel (1994) documents the parallel historical increases in height and economic growth, Bloom et al. (2001) document a causal connection between life expectancy and economic growth, and López-Casasnovas et al. (2005) offer a detailed theoretical and empirical account of the linkages between health and economic development (see also Wolpin (1997)).

Given the importance of children's health, a small literature has developed investigating its determinants, focusing on variables such as maternal labor supply and parental education (Glick and Sahn 1998; Thomas 1994; Thomas et al. 1990, 1991), health interventions and other public policies (Frankenberg et al.

[1] Throughout the text, we use the terms 'household' and 'family' interchangeably. In the data section, we explicitly define our empirical measures.

2005; Thomas et al. 1996; Muhuri 1995), micro-credit programs (Pitt et al. 2003), and inequality (Sahn and Younger 2005). Behrman and Deolalikar (1988) and Strauss and Thomas (1995) survey the literature. In many of these studies, household size enters the empirical analysis as an important control, although the estimated relationship is not of primary interest and the issue of causation versus correlation is often ignored. A recent, notable exception is Glick et al. (2005). The authors 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 sibship size.

In this paper, we seek to advance the quantity-quality literature in two important ways. First, we assess the empirical relevance of the trade-off using two measures of child health to capture quality: height-for-age and weight-for-age. Both are frequently used measures of health, 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 based on the notion of stochastic dominance (SD). This approach contrasts with previous empirical analyses of the trade-off relying on regression analysis and standard linear specifications. While such analyses yield clear rankings of child quality across households of different sizes that are easily interpretable, it does so at the cost of assuming *implicit* welfare functions that are linear in child quality. In contrast, our distributional approach utilizes all available information, makes explicit the welfare framework being employed, and tests for uniform rankings of distributions that are robust across a wide class of welfare functions, rendering comparisons based on specific indices unnecessary. Although there exist alternative frameworks for comparing distributions (or portions of distributions), the richness of SD analysis has led to an increasing number of applications. For example, Maasoumi and Millimet (2005) examine changes in U.S. pollution distributions over time and across regions at a point in time. Maasoumi and Heshmati (2000) analyze changes in the Swedish income distribution over time as well as across different population subgroups. Particularly relevant to the analysis at hand are previous applications of SD to the analysis of treatment effects. For instance, Amin et al. (2003) analyze the effect of a micro-credit program in Bangladesh on the distribution of consumption of participants versus non-participants. Abadie (2002) analyzes the impact of veteran status on the distribution of civilian earnings. Bishop et al. (2000) compare the distribution of nutrition levels across populations exposed to two different types of food stamp programs.

In addition to having a solid welfare basis, the SD approach also allows one to easily assess potential heterogeneity in the magnitude of the trade-off across the distribution of child quality, enabling one to

[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).

answer: Is the quantity-quality trade-off really a trade-off for all? 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 outcomes for the least healthy children in the population, 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 (obtained from, say, regression analysis) 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 over 7,000 children ten years of age and younger in households with at least two children to conduct our tests for SD. Specifically, we assess the impact of residing in a household with more than two children (versus only two children) on 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 an instrument based on the gender composition of children in the household, as utilized in, for example, Butcher and Case (1994), Angrist and Evans (1998), Cruces and Galiani (2004), and Conley and Glauber (2005). The results are striking. In particular, we reach three main conclusions. First, we find that only the distributions of weight-for-age can be ranked, and then only in the second-degree sense, when fertility is treated as exogenous; the height-for-age distributions cannot be ranked in either the first- or second-degree sense. Second, when fertility is treated as endogenous, we obtain statistically significant rankings favoring smaller households for both health outcomes, although only in the second-degree sense. Thus, only by incorporating the 'dispersion' of child quality can one obtain uniform welfare rankings. Finally, we find some evidence that the adverse impact of fertility is heterogeneous across the distribution; the effects on weight-for-age are greater for the least healthy children.

The remainder of the paper is organized as follows: section 2 provides a simple theoretical framework to justify the distributional analysis; section 3 details the econometric approach; section 4 discusses the data; section 5 presents the results; and section 6 concludes.

2 Theoretical Model

To motivate the empirical study, we present a simple model of the interaction between child quality (health) and quantity. The model extends Becker and Tomes (1976) to discuss distributional implications, as well as clarify the identification strategy. To begin, we assume that each household maximizes the following utility function

$$U(n, q, r, c) \tag{1}$$

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 as well as a household-level health endowment.[3] Thus, the production function for child quality can be expressed as

$$q = q(w, \theta) \tag{2}$$

where w is a vector of market purchased health inputs and θ is the household health endowment. 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

$$p_c c + p_n n + p_w w n - \delta(r - 0.5)^2 = I \tag{3}$$

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 its utility function (1) given the production function for child quality (2) and the budget constraint (3). The equilibrium conditions are:

$$\begin{aligned} \frac{\partial U}{\partial c} &= \lambda p_c = \lambda \pi_c \\ \frac{\partial U}{\partial q} &= \lambda \frac{p_w}{\partial q / \partial w} n = \lambda \pi_q \\ \frac{\partial U}{\partial n} &= \lambda (p_w w + p_n) = \lambda \pi_n \end{aligned} \tag{4}$$

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,

[3]Designation of the health endowment as a household-level attribute implies that the endowment of each child is equal. As Becker and Tomes (1976) discuss, heterogeneous endowments across children do not alter the primary implications of the model.

which reduces the shadow price of child quantity, π_n, further increasing n, and so on. This interaction behavior 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 in (2). 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 $q_{w\theta} > 0$ (e.g., if (2) is a Cobb-Douglas production function), then the change in π_q from an exogenous increase in n is deceasing in θ. Thus, the magnitude of the quantity-quality trade-off will be larger at the lower tail of the distribution of the health endowment θ. Second, if $q_{w\theta} < 0$, then the opposite occurs and the magnitude of the quantity-quality trade-off will be larger in the upper tail of the distribution of the health endowment θ. Finally, if $q_{w\theta} = 0$ (e.g., if (2) is additively separable as in equation (2.2) in Becker and Tomes (1976)), then the change in π_q from an exogenous increase in n is independent of θ. Thus, the magnitude of the quantity-quality trade-off will be independent of the health endowment θ. Assessing heterogeneity in the trade-off is one of the goals of the empirical analysis herein. However, since θ is unobserved, we test for such heterogeneity not across the distribution of θ, but rather across the distribution of q itself. 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 Empirical Methodology

3.1 Test Statistics

Our distributional comparisons are based on the notion of SD. SD tests offer the benefit of not only examining the uniformity of the quantity-quality trade-off across the entire distribution, but also offer the possibility of making robust welfare comparisons of distributions in the event that the distributions being compared are found to differ. Several tests for SD have been proposed in the literature; the approach herein is based on a generalized Kolmogorov-Smirnov test.[4] To begin, let X and Y denote two outcome (health) variables being compared (e.g., X (Y) might refer to health of children in families with two children (more than two children)). $\{x_i\}_{i=1}^{N}$ is a vector of N possibly dependent observations of X; $\{y_i\}_{i=1}^{M}$ is an analogous

[4] Maasoumi and Heshmati (2000) provide a brief review of the development of alternative tests.

vector of realizations of Y. In the spirit of the historical development of such two-sample tests, $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^M$ each constitute one sample. Thus, we refer to dependence between x_i and x_j, $i \neq j$, as *within-sample dependence* (similarly for observations of Y), and dependence between X and Y as *between-sample dependence*.

Assuming general von Neumann-Morgenstern conditions, let $\mathcal{U}_1$ denote the class of (increasing) utility functions u such that utility is increasing in wages (i.e. $u' \geq 0$), and $\mathcal{U}_2$ the class of social welfare functions in $\mathcal{U}_1$ such that $u'' \leq 0$ (i.e. concavity). Concavity represents an aversion to inequality in the health of children; a high concentration of very healthy and unhealthy children is undesirable. Let $F(x)$ and $G(y)$ represent the cumulative density functions (CDF) of X and Y, respectively, which are assumed to be continuous and differentiable.

Under this notation, X First Order Stochastically Dominates Y (denoted X FSD Y) *iff* $\mathrm{E}[u(X)] \geq \mathrm{E}[u(Y)]$ for all $u \in \mathcal{U}_1$, with strict inequality for some u. Equivalently,

$$F(z) \leq G(z) \qquad \forall z \in \mathcal{Z}, \text{ with strict inequality for some } z. \tag{5}$$

where $\mathcal{Z}$ denotes the union of the supports of X and Y. If X FSD Y, then the expected welfare from X is at least as great as that from Y for all increasing welfare functions, with strict inequality holding for some utility function(s) in the class. The distribution of X Second Order Stochastically Dominates Y (denoted as X SSD Y) *iff* $\mathrm{E}[u(X)] \geq \mathrm{E}[u(Y)]$ for all $u \in \mathcal{U}_2$, with strict inequality for some u. Equivalently,

$$\int_{-\infty}^{z} F(v)dv \leq \int_{-\infty}^{z} G(v)dv \qquad \forall z \in \mathcal{Z}, \text{ with strict inequality for some } z. \tag{6}$$

If X SSD Y, then the expected social welfare from X is at least as great as that from Y for all increasing and concave utility functions in the class $\mathcal{U}_2$, with strict inequality holding for some utility function(s) in the class. FSD implies SSD and higher orders.

Now define the following generalizations of the Kolmogorov-Smirnov test criteria:

$$d = \sqrt{\frac{NM}{N+M}} \min \sup_{z \in \mathcal{Z}} \left[F(z) - G(z)\right] \tag{7}$$

$$s = \sqrt{\frac{NM}{N+M}} \min \sup_{z \in \mathcal{Z}} \int_{-\infty}^{z} \left[F(u) - G(u)\right] du \tag{8}$$

where min is taken over $F - G$ and $G - F$, in effect performing two tests in order to leave no ambiguity between the 'equal' and 'unrankable' cases. Our nonparametric tests for FSD and SSD are based on the empirical counterparts of d and s using the empirical CDFs, where the empirical CDF for X is given by

$$\widehat{F}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\,(X \leq x)$$

and $\mathrm{I}(\cdot)$ is an indicator function; $\widehat{G}_M(y)$ is defined similarly for Y. If $\widehat{d} \leq 0$ ($\widehat{s} \leq 0$) to a degree of statistical certainty, then the null hypothesis of FSD (SSD) is not rejected (see Appendix A for details).

To this point X and Y have represented two *unconditional* variables. However, dependence between the quantity of children and other determinants of child health may confound the effect of fertility with the impact of these other attributes. In particular, health may be related to family background variables (such as parental income and education) and regional characteristics (such as social norms or access to public facilities such as schools and medical facilities).

To circumvent this issue, we employ two strategies. First, we control for a lengthy vector of *observable* determinants of child health and perform dominance tests on the *residual* distribution of health outcomes. To obtain these residuals, we control for a host of attributes that may generate a spurious correlation between the quantity and quality of children, and then conduct our SD tests on the distributions of health outcomes purged of the 'average' effects of these attributes. The conditioning covariates (discussed below) represent individual (such as age and gender), family (such as parental education, parental health, and assets), and regional attributes (such as urban/rural status and access to public facilities). Because the residual distributions reflect health outcomes purged of observable attributes, comparisons of such distributions only yield accurate inference regarding the effect of the number of children if one invokes the *selection on observables* assumption (i.e., households opt for larger or smaller families only on the basis of observables attributes).[5]

To proceed, we estimate separate health production function models for children in each household size category, obtain the intercept–adjusted residuals, and perform the dominance tests on these residuals.[6] Specifically, in the first-stage, we estimate

$$q_{ik} = \alpha_k + m_{ik}\beta_k + \widetilde{\varepsilon}_{ik}, \qquad k = 1, 2 \tag{9}$$

where q_{ik} is the health outcome for child i in household size category k (k equals 1 (two children) or 2 (more than two children) in the application), m is a vector of observable attributes, and $\widetilde{\epsilon}$ is the error term.

[5] To be precise, the assumption of selection on observables still holds if households choose family size on the basis of unobservables as long as these are uncorrelated with the health outcomes of children.

[6] The intercepts are included as part of the residuals, otherwise the conditional distributions will all be mean zero, precluding the possibility of first order dominance.

In the second-stage, we analyze the distributions of $\widehat{\varepsilon}_{ik} \equiv \widehat{\alpha}_k + \widehat{\widetilde{\varepsilon}}_{ik}$, which correspond to health outcomes *net of all observable characteristics* (evaluated at the *household size-specific* returns, β_k).

At this point, several comments are warranted. First, it should be emphasized that controls for the number of children are omitted from (9), thereby allowing the error term to capture the residual effect of household size not captured by the included regressors. Second, first-stage variables such as access to pubic facilities are potentially endogenous in the usual sense (see, e.g., Pitt et al. (1993)). However, in the current context, this is not problematic since we are not interested in identifying a causal relationship between these variables and child health. Instead, these variables partially proxy for unobservable geographic attributes that may be correlated with health outcomes and the presence of such facilities.[7]

Third, the intercept-adjusted residuals, $\widehat{\epsilon}_{ik}$, reflect health outcomes net of all observable characteristics evaluated at the household size–specific returns, β_k. Since this approach nets out differences due to observables as well as the household size-specific returns to such observables, we refer to these tests as being based on 'Partial Residuals' (PR). As an alternative, we also conduct tests based on the 'Full Residuals' (FR), where we denote the full residuals as inclusive of differences in the return to observables.[8] To accomplish this, we re-write the first-stage regression (9) for group k as

$$\begin{aligned} q_{i2} &= \alpha_2 + m_{i2}\beta_2 + \widetilde{\epsilon}_{i2} \\ &= \alpha_2 + m_{i2}\beta_2 + \widetilde{\epsilon}_{i2} + (m_{i2}\beta_1 - m_{i2}\beta_1) \\ &= \alpha_2 + m_{i2}\beta_1 + m_{i2}(\beta_2 - \beta_1) + \widetilde{\epsilon}_{i2} \end{aligned} \tag{10}$$

where group $k = 1$ (two children) is implicitly treated as the 'dominant' category (Neuman and Oaxaca 2004). Consequently, we amend the residual tests to compare the previous intercept-adjusted residual distribution of $\widehat{\epsilon}_{i1}$ with $\widehat{\epsilon}_{i2}^{FR} \equiv \left(\widehat{\alpha}_2 + m_{i2}(\widehat{\beta}_2 - \widehat{\beta}_1) + \widehat{\widetilde{\epsilon}}_{i2}\right)$. Comparison of the PR and FR test results provides insight into whether there are important interactions between the quantity of children and the returns to other health determinants.

Despite the lengthy control set utilized above, the selection on observables assumption may still be troublesome. As a result, many previous studies of the quantity-quality trade-off have resorted to instrumental variable (IV) methods. In similar spirit, we employ a second strategy – robust to selection on unobservables – to isolate the causal effect of the number of children on child health. Specifically, we implement the procedure developed in Abadie (2002) to compare the distributions of potential health outcomes

[7]However, it may be the case that such control variables themselves reflect the quantity-quality trade-off (e.g., whether or not the household boils water). If this is the case, our results may represent a lower 'bound' on the trade-off.

[8]It should be noted, however, that most, if not all, empirical studies examining the quantity-quality trade-off do not, to our knowledge, allow the effects of other covariates to depend on the number of children.

for (a subpopulation of) children using an IV method. According to Imbens and Rubin (1997), when a binary instrumental variable is available, the potential distributions of the outcome variable are identified for the subpopulation (referred to as *compliers*) whose treatment assignment (in this case, household size) is affected by variation in the instrument.

Define Q_0 and Q_1 as the distribution of potential outcomes (health) for the untreated (children in households with two children) and treated (children in households with more than two children), with $q_i(0)$ and $q_i(1)$ representing specific values for observation i, $i = 1, ..., N + M$, from the respective distribution. Let D_i be a binary variable equal to zero (one) if the child is from a family with two children (more than two children), and Z_i be a binary instrument (discussed below). Denote $D_i(0)$ the value of D_i if $Z_i = 0$; similarly for $D_i(1)$. Thus, D_i may be written as

$$D_i = \begin{cases} D_i(0) & \text{if } \; Z_i = 0 \\ D_i(1) & \text{if } \; Z_i = 1 \end{cases} \tag{11}$$

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 $\{q_i(0), q_i(1)\}$ are not both observed since only one state of the world – $Z_i = 0$ or $Z_i = 1$ – is observed. Instead, the realized treatment assignment $D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)$ and the realized potential outcome $q_i = q_i(1)D_i + q_i(0)(1 - D_i)$ are observed.

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

$$\begin{aligned} F^c(q) &= \mathrm{E}\left[\mathrm{I}\{q_i(0) \leq q\} | D_i(1) = 1, D_i(0) = 0\right] \\ G^c(q) &= \mathrm{E}\left[\mathrm{I}\{q_i(1) \leq q\} | D_i(1) = 1, D_i(0) = 0\right] \end{aligned} \tag{12}$$

If Z_i satisfies the following three assumptions:

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

(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 the dominance tests defined above conducted on the distributions $F^c(q)$ and $G^c(q)$ identify the causal effect of household size for the subpopulation of compliers, even if there exist unobservable attributes correlated with both the quantity and quality of children (Imbens and Angrist 1994; Angrist et al. 1996). Moreover, as shown in Abadie (2002), SD tests conducted on the distributions $F^c(q)$ and $G^c(q)$ are equivalent to tests conducted on the distributions $F(q)$ and $G(q)$, where F (G) represents the distribution of health outcomes for children with $Z_i = 0$ ($Z_i = 1$). Thus, the test statistics in (7) and (8) are obtained by replacing F and G with their empirical counterparts:

$$\widehat{F}_{N_0}(q) = \frac{1}{N_o}\sum_{i=1}^{N_0} \mathrm{I}\left(Q \leq q\right) \quad (13)$$

$$\widehat{G}_{N_1}(q) = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathrm{I}\left(Q \leq q\right) \quad (14)$$

where N_0 (N_1) is the size of the sample with $Z_i = 0$ ($Z_i = 1$).

In the analysis, we utilize two IVs. The primary instrument, based on Butcher and Case (1994) and Angrist and Evans (1998), is an indicator of whether the first two children are of the same gender. For robustness, we also utilize a second instrument which is unity if the first two children are both girls. As shown in Rosenzweig and Wolpin (2000), for the gender composition of children to be a valid exclusion restriction certain assumptions are required. Specifically, we require that (i) the sex ratio of children, r, enters a family's utility function in (1), (ii) child-rearing costs do not vary with r (i.e., $\delta = 0$ in (3)) and/or consumption, c, and per child quality, q are strongly separable in (1), and (iii) per child quality, q, and r are strongly separable in (1). We assess the validity of these requirements below (through examination of the first-stage regressions and use of overidentification tests).

3.2 Inference

The asymptotic distribution of the test statistics, d and s, depend on the unknown underlying distributions, F and G. In the analysis below, we first approximate the empirical distribution of the test statistics using *simple bootstrap* techniques (Maasoumi and Heshmati 2000; Maasoumi and Millimet 2005). To evaluate the null $\mathrm{H}_o : d \leq 0$, we first report in our tables whether the observed empirical distributions are *seemingly* rankable by FSD or SSD; we present the sample values of $\max\{d_1\}$, $\max\{d_2\}$, $\widehat{d}$, $\max\{s_1\}$, $\max\{s_2\}$, and $\widehat{s}$ (see Appendix A). We then obtain simple bootstrap estimates of the probability that d lies in the non-positive interval (i.e. $\Pr\{d \leq 0\}$) using the relative frequency of $\{\widehat{d}^* \leq 0\}$, where $\widehat{d}^*$ is the bootstrap estimate of d (500 repetitions are used).[9] If this interval has a large probability, say 0.90 or higher, and $\widehat{d} \leq 0$, we may infer dominance to a desirable degree of confidence. If this interval has a low probability, say 0.10 or smaller, and $\widehat{d} > 0$, we may infer the presence of significant crossings of the empirical CDFs, implying an inability to rank the outcomes. Finally, if the probability lies in the intermediate range, say between 0.10

[9]Note, we also report simple bootstrap estimates of the $\Pr\{d^* \geq \widehat{d}\}$. These are provided to facilitate visualization of the simple bootstrap distribution.

and 0.90, there is insufficient evidence to distinguish between equal and unrankable distributions. This is a classic confidence interval test; specifically, we are assessing the likelihood that the event $d \leq 0$ has occurred. Similarly, we estimate $\Pr\{s \leq 0\}$ to evaluate the second order dominance proposition given by $\mathrm{H}_o : s \leq 0$.[10]

As an alternative, we also evaluate the less decisive dominance proposition $\mathrm{H}_o : d = 0$ via the Linton et al. (2005) *recentered bootstrap* procedure, which the authors demonstrate provides a consistent test. It is known that $\widehat{d}$ converges to d under general conditions (likewise for the SSD statistic). However, under the null $\mathrm{H}_o : d = 0$, centering of computations around their corresponding sample values introduces second order errors that are negligible for first order (asymptotic) approximations, but is desirable for removing some uncertainties due to estimation of unknown parameters and distributions. This is the source of improvement in bootstrap power gained from recentering. The other source of improvement arising from recentering pertains to the technique's robustness to within-sample dependence.

Utilizing the algorithm detailed in Appendix A, we obtain recentered bootstrap p-values in the classical sense as the relative frequency of $\{\widehat{d}^{**} > \widehat{d}\}$, where $\widehat{d}^{**}$ is the recentered bootstrap estimate of d. If the $\Pr\{\widehat{d}^{**} > \widehat{d}\}$ is low, say 0.10 or smaller, we reject the null $\mathrm{H}_o : d = 0$; if this p-value is greater than 0.10, we fail to reject the null.[11] It is important to emphasize, however, that while rejection of the null provides valuable insight in the recentered bootstrap case, failure to reject the null provides less information. If we reject the null and $\widehat{d} < 0$, we may infer dominance to a desirable degree of confidence. Conversely, if we reject the null and $\widehat{d} > 0$, we may infer unequal, but unrankable, distributions. These are both strong findings, as the former (latter) indicates that all (not all) increasing social welfare functions will concur on the relative rankings of the distributions in question. On the other hand, failure to reject the null merely implies that we cannot eliminate the possibility that $F = G$; strict dominance also cannot be ruled out to some degree of confidence. Seen in this light, the recentered bootstrap is a conservative test. In

[10]Note, we do not impose and test the Least Favorable Case (LFC) of equality of the distributions. This could be done by combining the data on X and Y and bootstrapping from the combined sample (e.g., Abadie 2002). Our bootstrap samples still contain N (M) observations from X (Y). As argued in Linton et al. (2003), working under LFC has some undesirable power consequences as it can produce biased tests that are not similar on the boundary of the null. This happens when the boundary of the null itself is composite. Following Maasoumi and Heshmati (2000) and Maasoumi and Millimet (2005), we are also reporting the *maximum test sizes associated with our (conservative) critical value of zero*, which is clearly on the boundary of the null that includes the LFC. Thus, the bootstrap probabilities reported represent the critical levels associated with this non-rejection region. Such critical levels can be shown to be conservative since, in the limit, they are at least as large as the corresponding levels for the asymptotic test on the boundary (Linton et al. 2003).

[11]We also report the $\Pr\{\widehat{d}^{**} \leq 0\}$ in the tables. This allows the reader to see the significance level (size) of the test associated with the special critical value 'zero.' In our tables, these are obtained simply as $\Pr\{\widehat{d}^{**} > 0\} = 1 - \Pr\{\widehat{d}^{**} \leq 0\}$.

the discussion of the results, we focus more heavily on the more decisive simple bootstrap for inference. Similarly, we report the relative frequency of $\{\widehat{s}^{**} > \widehat{s}\}$ and $\{\widehat{s}^{**} > 0\}$ to evaluate the null $\text{H}_o : s = 0$.

A final, necessary comment pertains to inference in the FR tests (i.e., those incorporating the Oaxaca-Blinder decomposition). Due to the usage of a common set of coefficient estimates in obtaining both residual distributions being compared, there *necessarily* exists between-sample dependence. For example, the FR test compares the distributions of $\widehat{\epsilon}_{i1}$ and $\widehat{\epsilon}_{i2}^{FR}$. The former depends on $\{q_{i1}, m_{i1}, \beta_1(q_1, m_1)\}$, where q_1 and m_1 represent the full data vector for q and m for the $k = 1$ sample; the latter, $\widehat{\epsilon}_{i2}^{FR} = q_{i2} - m_{i2}\beta_1$, depends on $\{q_{i2}, m_{i2}, \beta_1(q_1, m_1)\}$. This source of dependence is atypical. Between-sample dependence usually arises when the same individuals appear in the two samples being compared (e.g., distributions of pre- and post-tax incomes for a sample of individuals). To handle this more common type of between-sample dependence, pairwise (or 'clustered') bootstrap samples are drawn in order to maintain the dependence in the resampled data (Linton et al. 2005). In the current situation, the between-sample dependence is maintained by re-estimating the first-stage equations (9) and (10) on each bootstrap resample. Specifically, by resampling N observations $\{q_{i1}^*, m_{i1}^*\}$ and M observations $\{q_{i2}^*, m_{i2}^*\}$ nonparametrically and re-estimating (9), we obtain the resampled distributions of $\widehat{\epsilon}_{i1}^*$ and $\widehat{\epsilon}_{i2}^{FR*}$, where the former depends on $\{q_{i1}^*, m_{i1}^*, \beta_1^*(q_1^*, m_1^*)\}$ and the latter depends on $\{q_{i2}^*, m_{i2}^*, \beta_1^*(q_1^*, m_1^*)\}$. Thus, as in the usual pairwise bootstrap case, the source of between-sample dependence is maintained in the resampling procedure.

4 Data

The data are obtained from Indonesian Family Life Survey (IFLS), a large-scale longitudinal survey conducted jointly by RAND and the Center for Population and Policy Studies (CPPS) at the University of Gadjah Mada. The IFLS provides a rich data source based on a sample of households representing about 83% of the Indonesian population living in 13 of the nation's 26 provinces in 1993. Three waves exist presently: 1993, 1997, and 2000 (see Strauss et al. (2004a, 2004b) for a complete description of the surveys). We utilize the 2000 wave, which is the most current survey containing physical assessments of children, and form a sample of roughly 7,300 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.

Two outcomes of interest are examined: child's height and child's weight. However, since children are still growing, comparing anthropometric data from children of different ages is complicated (Vidmar et al. 2004). Thus, we standardize the raw data on height and weight to the reference population for the child's age and sex utilizing the 1990 British Growth Reference data. The quantity of children is defined as the number of children 'belonging' to a given set of parents. Assuming that couples 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 from the previous union, then these children are used when computing the number of siblings. Our definition *excludes* children 'belonging' to other couples who may share a common residence. 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 1993 or 1997 wave. Once the number and identity of siblings is established, we create a dummy variable, *MoreThan2*, equal to one if the household has more than two children and zero otherwise (recall, the sample is limited to households with at least two children). To implement the IV SD tests, we also create two binary variables based on the gender of children: *SameSex2* is equal to one if the first two children are of the same sex (zero otherwise), and *Daughter2* is equal to one if the first two children are female (zero otherwise).

To obtain the partial and full residuals, we utilize an extensive set of individual, parental, and household controls. Specifically, we include the following variables (in addition to a constant term):

Individual: age in months, a gender dummy, birth order dummies, days in bed due to poor health, dummy variables for illness (cough, fever, and diarrhea) reported in the past four weeks prior to survey (to capture fluctuations in health in the very short-term);

Parental: separate dummy variables for mother's and father's education (less than primary school, junior high school, senior high school, university, and other), mother's and father's height, mother's and father's weight, mother's and father's age, mother's and father's work status dummy variables (work or not), separate dummy variables for mother's and father's religion (Islam, Protestant, Catholic, Hindu), 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 (own or not), dummy variable indicating if the household head is female, dummy variables for decisionmaking powers concerning children's health (jointly made by husband and wife, only made by husband, only made by wife, and other), dummy variables for type of dwelling (single unit single level, single unit multiple levels, duplex single level, duplex multiple level, multiple unit single level, and other), 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 if water is boiled prior to consumption, dummy variables for type of sanitary conditions of household, region dummy (rural or urban), province dummy variables (North Sumatra, West Sumatra, South Sumatra, Lapung, Jakarta, West Java, Central Java, Yogyakarta, East Java, Bali, West Nusa Tenggara, South Kalimantan, and South Sulawesi), and a dummy variable indicating whether or not house size is missing.[12]

The inclusion of parental height and weight is noteworthy. As noted in Thomas et al. (1990, 1996), parental health status, as reflected in anthropometric measures, may have a direct effect on child health through their impact on birthweight. In addition, parental measures may partially capture household unobservables, lending further credibility to our causal interpretation of the results reported below.

The IFLS also collected very extensive information at the community level. However, this community information is only available for some communities. As a result, rather than limiting our sample size, we perform the PR and FR SD tests twice, once conditioning on just the previous individual, parental, and household characteristics, and once conditioning on the previous sets of variables plus community characteristics. Specifically, we include the number of various health facilities (hospitals, private practitioners, integrated health posts) and schools (elementary schools, junior high schools, and senior high schools) in the community, as well as the minimum distance to each type of health facility and school.

Table B1 (in Appendix B) presents summary statistics. On average, children in a household with two children are taller and weigh more than those in a household with more than two children, consonant with the quantity-quality trade-off, although they are also nearly a year older. Before turning to the distributional results, Table B2 presents the results from standard parametric OLS and IV regressions. This is useful not only for comparison to the SD results, but also to assess the representativeness of the sample. 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. Treating fertility as endogenous, but excluding the control variables, we find a much larger adverse impact of fertility on both height-for-age and weight-for-age, although the coefficients are only statistically significant for weight-for-age. Moreover, we also find that the instruments are positive and statistically significant, as expected, in the first-stage regressions (especially *SameSex2*), and that when we utilize both instruments, the instruments easily pass the Sargan overidentification test. This is comforting given the restrictions required for the instruments to be valid discussed previously. Finally, when we estimate the parametric models by IV and include the control variables, the adverse impact of fertility becomes even larger in magnitude, continues to be statistically significant for weight-for-age, and the instruments

[12]Missing values are replaced with sample means.

continue to fair well in the first-stage and according to the Sargan test. Given these baseline results, we now turn to discussion of the SD results.

5 Results

5.1 Unconditional Tests

The first set of tests compares the unconditional distributions, treating fertility (*MoreThan2*) as either exogenous or endogenous. The results are displayed in Tables 1 (height-for-age) and 2 (weight-for-age), with Panel A containing the results treating fertility as exogenous and Panels B and C utilize *SameSex2* and *Daughter2* as an instrument, respectively. The corresponding CDFs, integrated CDFs, and differences between the CDFs at each quantile are plotted in Figures 1 (height-for-age) and 2 (weight-for-age).

Treating fertility as exogenous (Panel A in Tables 1 and 2), the distributions of height-for-age and weight-for-age for children in households with only two children are observed to first order dominate the corresponding distribution for children in larger households.[13] However, the simple bootstrap indicates that only the distributions for weight-for-age can be ranked to a degree of statistical certainty, and then only in the second order sense (height-for-age: $\Pr(s \leq 0) = 0.846$; weight-for-age: $\Pr(s \leq 0) = 0.930$) The recentered bootstrap fails to reject the null $H_o : d = 0$ or $s = 0$ for both health measures, indicating that strict dominance or equal distributions cannot be rejected. Taken altogether, then, the results treating fertility as exogenous suggest a statistically significant, second order ranking in favor of smaller households for weight-for-age, and fail to provide evidence of a statistically significant ranking for height-for-age.

The observed FSD ranking combined with the statistically significant SSD ranking for weight-for-age is an extremely powerful result. First, it implies that the quantity-quality trade-off does hold at the distributional level (when fertility is treated as exogenous and no covariates are included). Second, the results show that any policymaker whose welfare function is increasing in child health and averse to health inequality will favor reductions fertility. Third, the fact that there are differences between the results for height and weight suggest that the quantity-quality trade-off may differentially operate in short- and long-term given what each health measure captures.

Examining the actual plots (Figures 1 and 2, top row) provides additional detail, especially with respect to the magnitude and uniformity of the quantity-quality trade-off. In particular, the plots reveal little actual

[13] While the CDFs for height-for-age cross in the extreme lower tail in the top row of Figure 1 (favoring larger households and thereby precluding an FSD or SSD ranking in favor of small households), Table 1 reports a finding of FSD (and, hence, SSD) because of the procedure utilized, which ignores the extreme lower and upper tails (see Appendix A).

difference between the distributions for height-for-age and weight-for-age; in both cases, larger household size is associated with a roughly 0.2 to 0.3 decline in health, and this effect is approximately uniform across the distribution. In light of the theoretical model, this is consonant with the production function in (2) being additively separable for both anthropometric measures, as in Becker and Tomes (1976).

Although interesting, the previous results treat fertility as exogenous (and fail to control for any covariates). Thus, we now turn to Panels B (*SameSex2*) and C (*Daughter2*) in Tables 1 and 2 which make use of the two instruments. Using the preferred instrument, *SameSex2*, we observe a second order ranking favoring small households for both height-for-age and weight-for-age.[14] However, in both cases, the simple bootstrap indicates that this ranking is not statistically significant (height-for-age: $\Pr(s \leq 0) = 0.798$; weight-for-age: $\Pr(s \leq 0) = 0.858$), and the recentered bootstrap fails to reject the null $H_o : s = 0$ (height-for-age: p-value = 0.996; weight-for-age: p-value = 0.998), indicating that strict second order dominance or equal distributions cannot be rejected. Examination of the plots in Figures 1 and 2 (second row) continues to indicate that the differences in the CDFs are fairly uniform across the entire health distribution, consonant with the assumption of additive separability for the health production function. Finally, when using the second instrument, *Daughter2*, we do not observe any ranking in the first or second degree sense. Moreover, the recentered bootstrap rejects the null $H_o : d = 0$ and $H_o : s = 0$ for both health outcomes (height-for-age: FSD p-value = 0.022; SSD p-value = 0.012; weight-for-age: FSD p-value = 0.040; SSD p-value = 0.020), implying the presence of significant crossings. Examination of the plots (Figures 1 and 2, bottom row), however, suggests that any differences in the health distributions across large and small households is minimal.

Just as the statistically significant second order ranking for weight-for-age when sibship size is treated as exogenous is a strong result, the lack of any statistically significant rankings according to the IV tests is also strong; conclusions regarding the empirical relevance of the quantity-quality trade-off are not robust across all social welfare functions in the class $\mathcal{U}_2$. Moreover, the fact that some differences arise – although they may not be statistically meaningful – across the results based on *SameSex2* versus the *Daughter2* instruments suggests that the impact of household size is heterogeneous in the population. Specifically, the trade-off does not appear to be identical in the subpopulation of households induced to have a third child because the first two children are both daughters, as opposed to both sons.

To assess the sensitivity of these unconditional results, we turn to the residual-based SD test results. In particular, while the sex composition of children is arguably uncorrelated with most other determinants

[14] We consider *SameSex2* the preferred instrument for the unconditional tests since it yielded higher first-stage F-statistics in the parametric models with no other covariates (see Table B2 in Appendix B).

of child health, the instrument may be correlated with the gender of the first two children unless the probability of having a son is exactly one-half (Angrist and Evans 1998). Since this is not always the case in the sample (see Table B1), we analyze the distributions of child health purged of the effects of the gender of the first two children, as well as the other potential determinants of child health discussed in the previous section.

5.2 Partial Residual Tests

Tables 3 (height-for-age) and 4 (weight-for-age) display the PR test results, which ignore differences in the returns to health inputs across household size.[15] The differences in the CDFs are displayed in column 1 of Figures 3 (height-for-age) and 4 (weight-for-age). Surprisingly, the results for the two health outcomes are now in total opposition. In terms of height-for-age, we observe at least a marginally statistically significant FSD ranking in favor of smaller households according to the simple bootstrap regardless of whether fertility is treated as exogenous ($\Pr(d \leq 0) = 0.916$) or endogenous (*SameSex2* IV: $\Pr(d \leq 0) = 0.930$; *Daughter2* IV: $\Pr(d \leq 0) = 0.890$). In all three cases, the recentered bootstrap fails to reject the null $\text{H}_o : d = 0$, indicating that strict dominance or equal distributions cannot be rejected. Finally, in all three cases, the plots in Figure 3 indicate the impact of fertility continues to be approximately uniform across the entire distribution.

Turning to the results for weight-for-age, we now find strong results in favor of *larger* households, suggesting the possibility of a short-run gain to more siblings. Specifically, we obtain a statistically significant FSD ranking in favor of households with more than two children according to the simple bootstrap regardless of whether fertility is treated as exogenous ($\Pr(d \leq 0) = 0.912$) or endogenous (*SameSex2* IV: $\Pr(d \leq 0) = 0.934$; *Daughter2* IV: $\Pr(d \leq 0) = 0.964$). As with height-for-age, the less decisive recentered bootstrap fails to reject the null $\text{H}_o : d = 0$, indicating that strict dominance or equal distributions cannot be rejected. Lastly, as with height-for-age, the plots in Figure 4 indicate the apparent beneficial impact of fertility is roughly uniform across the entire distribution. Again, these results are consistent with a health production function that is additively separable, as in Becker and Tomes (1976).

Although not shown, the contrasting results is attributable to the change in rankings of the intercepts across the first-stage regressions for the two health outcomes. For height-for-age (weight-for-age), the intercept is larger for smaller (larger) households, and this drives the rankings. However, because the PR

[15] The first-stage regressions used to obtain the results in Tables 3 – 6 do not control for community attributes. We discuss the results obtained controlling for community characteristics below. In addition, the first-stage regression results are not reported, but are available upon request.

tests only account for differences in the intercept, while ignoring differences in the slope coefficients, we now turn to the more complete FR test results.

5.3 Full Residual Tests

The FR test results are displayed in Tables 5 (height-for-age) and 6 (weight-for-age); the corresponding differences in the CDFs are displayed in column 1 of Figures 5 (height-for-age) and 6 (weight-for-age). Accounting for differences in all the first-stage parameters, the results for the two health outcomes are once again in virtual agreement, implying little difference between relatively long- versus short-term measures of health. In terms of height-for-age, we are unable to rank the distributions when fertility is treated as exogenous; the recentered bootstrap marginally rejects the null $\text{H}_o : s = 0$ (p-value = 0.114), indicating the presence of significant crossings. However, when fertility is treated as endogenous, we observe a statistically significant SSD ranking in favor of smaller households using the *SameSex2* instrument and the simple bootstrap ($\Pr(s \leq 0) = 0.970$). While we also observe a SSD ranking using the *Daughter2* instrument, the result is not statistically significant according to the simple bootstrap ($\Pr(s \leq 0) = 0.696$), confirming the previous unconditional results that suggest differences in the trade-off across the subpopulations affected by each of the two instruments. Finally, in all instances, the plots in Figure 5 indicate that the beneficial impact of lower fertility is approximately uniform across the entire distribution.

In terms of weight-for-age, two differences arise. First, when fertility is treated as exogenous, we now observe a statistically significant SSD ranking in favor of smaller households according to the simple bootstrap ($\Pr(s \leq 0) = 0.956$). Second, while the IV results confirm the previous statistically significant SSD ranking for height-for-age using the *SameSex2* instrument, the impact of fertility appears more pronounced at the lower end of the distribution, suggesting a larger effect of fertility reductions on the least healthy children (as measured by age appropriate weight). This non-uniformity provides some evidence that the health production function may not be additively separable for weight-for-age. Moreover, since weight-for-age may reflect short-term changes in health status, whereas height-for-age only reflects long-term nutritional status, these results suggest important differences in the production of child health and, hence, the quantity-quality trade-off between the short- and the long-term.

5.4 Sensitivity Analysis

As discussed in Section 4, community attributes that are likely to influence fertility and human capital investments are only available for a portion of the sample utilized thus far. While these attributes are most likely uncorrelated with sibling sex composition, we nonetheless assess the robustness of our results by conducting the PR and FR tests adding these variables to the first-stage covariate set. The results are displayed in Tables B3 – B6 in Appendix B; the corresponding plots are given in column 2 of Figures 3 – 6. The results are qualitatively unchanged. In particular, the preferred results based on the FR SD tests and using the *SameSex2* instrument continue to indicate a statistically significant SSD ranking in favor of smaller households. As such, any policymaker with a social welfare function that is increasing and concave in child health will prefer reductions in fertility. This is a much more robust finding that previous parametric investigations of the quantity-quality trade-off based solely on the (conditional) mean. Moreover, the plots continue to affirm the uniformity of the impact of fertility across the entire health distribution for height-for-age, and a marginally larger trade-off in the lower tail for weight-for-age. As noted in the theoretical model, this suggests that the Becker and Tomes (1976) assumption of additive separability between the health endowment and other health inputs in the health production function may not hold for weight.

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, empirical studies that have been undertaken typically focus on education as a measure of child quality and have been predominantly limited to linear regression analysis. While such results are easily interpretable, at best they provide evidence of the impact of fertility on measures of child-specific human capital investments only at the (conditional) mean. In contrast, this study analyzes the impact of fertility over the entire distribution of two measures of health using data from Indonesia, while accounting for the potential endogeneity of fertility. Our distributional approach utilizes all available information in the data, makes explicit the welfare framework being employed, and assesses several potential forms of heterogeneity in the trade-off. The results yield three main conclusions. First, we find that the distributions of weight-for-age can be ranked, but only in the second-degree sense, when fertility is treated as exogenous; the height-for-age distributions cannot be ranked in either the first- or second-degree sense. Second, when fertility is treated as endogenous, we obtain statistically significant rankings favoring smaller households for both health outcomes, although only in the second-degree sense. Thus, only by incorporating the 'dispersion' of child health can one obtain uniform welfare rankings across a wide class of social welfare functions. Finally, we find evidence consistent with several types of heterogeneity in the magnitude of the quantity-quality trade-off: across the distribution of weight-for-age, between the short-term and long-term, and across different subpopulations. Although insightful, 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 sub-populations of compliers? Second, utilizing other instruments for identification, are there gains from reducing fertility from, say, six children in a family to five, or from two children to only one? Nonetheless, answering these questions within a distributional framework is necessary for understanding 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] Ahn, T.S., J. Knodel, D. Lam, and J. Freidman (1998), "Family Size and Children's Education in Vietnam," *Demography*, 35, 57-70.

[3] Amin, S., A.S. Rai, and G. Topa (2003), "Does Microcredit Reach the Poor and Vulnerable? Evidence From Northern Bangladesh," *Journal of Development Economics*, 70, 59-82.

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

[5] Becker, G.S. (1960), "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries*, Universities-National Bureau Conference Series, No. 11, Princeton, NJ: Princeton University Press.

[6] 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.

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

[8] Behrman, J. and A. Deolalikur (1988), "Health and Nutrition," Chapter 14 in H. Chenery and T.N. Srinivasan (eds.), *Handbook of Development Economics*, Vol. 1, Elsevier Science Publishers B.V.

[9] Bishop, J.A., J.P. Formby, and L.A. Zeager (2000), "The Effect of Food Stamp Cashout on Undernutrition," *Economics Letters*, 67, 75-85.

[10] 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.

[11] Bloom, D.E., D. Canning, J. Sevilla (2001), "The Effect of Health on Economic Growth: Theory and Evidence," NBER Working Paper No. 8587.

[12] 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.

[13] Chernochovsky, D. (1985), "Socioeconomic and Demographic Aspects of School Enrollment and Attendance in Rural Botswana," *Economic Development and Cultural Change*, 33, 319-32.

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

[15] 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.

[16] Cruces, G. and S. Galiani (2004), "Fertility and Female Labor Supply in Latin America: New Causal Evidence," unpublished manuscript, London School of Economics.

[17] 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.

[18] Frankenberg, E., W. Suriastini, D. Thomas (2005), "Can Expanding Access to Basic Healthcare Improve Children's Health Status? Lessons from Indonesia's 'Midwife in the Village' Programme," *Population Studies*, 59, 5-19.

[19] Glick, P., A. Marini, and D.E. Sahn (2005), "Estimating the Consequences of Changes in Fertility on Child Health and Education in Romania: An Analysis Using Twins Data," Cornell University.Food and Nutrition Policy Program Working Paper No. 183.

[20] Glick, P. and D.E. Sahn (1998), "Maternal Labour Suppy and Child Nutrition in West Africa," *Oxford Bulletin of Economics and Statistics,* 60, 325-355.

[21] Gomes, M. (1984), "Family Size and Educational Attainment in Kenya," *Population and Development Review*, 10, 647-660.

[22] Hanushek, E. (1992), "The Trade-off Between Child Quantity and Quality," *Journal of Political Economy*, 100, 84-117.

[23] Hossain, S.I. (1990), "Interrelations between Child Education, Health, and Family Size: Evidence from a Developing Country," *Economic Development and Cultural Change*, 38, 763-781.

[24] Kaestner, R. (1997), "Are Brothers Really Better? Sibling Sex Composition and Educational Achievement Revisited," *Journal of Human Resources*, 32, 250-284.

[25] Knodel, J. and M. Wongsith (1991), "Family Size and Children's Education in Thailand: Evidence from a National Sample," *Demography*, 28, 119-131.

[26] Lee, J. (2004), "Sibling Size and Investment in Children's Education: An Asian Instrument," IZA Discussion Paper No. 1323.

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

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

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

[30] Maasoumi, E. and D.L. Millimet (2005), "Robust Inference Concerning Recent Trends in U.S. Environmental Quality," *Journal of Applied Econometrics*, 20, 55-77.

[31] Mock, P.R. and J. Leslie (1984), "Childhood Malnutrition and Schooling in the Terai Region of Nepal," *Journal of Development Economics*, 20, 33-52.

[32] Muhuri, P.K. (1995), "Health Programs, Maternal Education, and Differential Child Mortality in Matlab, Bangladesh," *Population and Development Review*, 21, 813-834.

[33] Neuman, S. and R.L. Oaxaca (2004), "Wage Decompositions with Selectivity Corrected Wage Equations: A Methodological Note," *Journal of Economic Inequality*, 2, 3-10.

[34] Parish, W. and R. Willis (1993), "Daughters, Education, and Family Budgets: Taiwan Experiences," *Journal of Human Resources*, 28, 863-898.

[35] Pitt, M.M., S. Khandkur, O.H. Chowdhury, and D.L. Millimet (2003), "Credit Programs for the Poor and the Health Status of Children in Rural Bangladesh," *International Economic Review*, 44, 87-118.

[36] Pitt, M.M., M. Rosenzweig, D. Gibbons (1993), "The Determinants and Consequences of the Placement of Government Programs in Indonesia," *World Bank Economic Review*, 7, 319-348.

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

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

[39] Qian, N. (2005), "Quantity-Quality: The Positive Effect of Family Size on Scholl Enrollment in China," unpublished manuscript, Department of Economics, MIT.

[40] Sahn, D. E. and S. D. Younger (2005), "Improvements in Children's Health: Does Inequality Matter?," unpublished manuscript, Department of Economics, Cornell University.

[41] 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.

[42] 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.

[43] 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.

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

[45] Thomas, D. (1994), "Like Father, Like Son; Like Mother, Like Daughter: Parental Resources and Child Height," *Journal of Human Resources*, 29, 951-988.

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

[47] Thomas, D., V. Lavy, and J. Strauss (1996), "Public Policy and Anthropometric Outcomes in the Cŏte d'Ivoire," *Journal of Public Economics*, 61, 155-192.

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

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

[50] Vidmar, S., J. Carlin, K. Hesketh and T. Cole (2004), "Standardizing Anthropometric Measures in Children and Adolescents with New Functions for egen," *Stata Journal,* 4, 50-55.

[51] Ware, J.E. (1987), "Standards for Validating Health Measures: Definition and Content," *Journal of Chronic Disease*, 40, 473-80.

[52] Willis, R.J. (1973), "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 81, S14-64.

[53] Wolpin, K.I. (1997), "Determinants and Consequences of the Mortality and Health of Infants and Children," Chapter 10 in M.R. Rosenzweig and O. Stark (eds.) *Handbook of Population and Family Economics*, Vol. 1A, Elsevier Science B.V.

A Appendix: Technical Details

A.1 Computation of $\widehat{d}$ and $\widehat{s}$

The test for FSD requires:

(i) computing the values of $\widehat{F}(z_j)$ and $\widehat{G}(z_j)$ for z_j, $j = 1, ..., J$, where J denotes the number of points in the support $\mathcal{Z}$ that are utilized ($J = 500$ in the application, where the points are equally spaced beginning at the first percentile and ending at the 99^{th} percentile of the empirical support, $\mathcal{Z}$, to focus attention away from extreme values),

(ii) computing the differences $d_1(z_j) = \widehat{F}(z_j) - \widehat{G}(z_j)$ and $d_2(z_j) = \widehat{G}(z_j) - \widehat{F}(z_j)$, and

(iii) finding $\widehat{d} = \sqrt{\frac{NM}{N+M}} \min \{\max\{d_1\}, \max\{d_2\}\}$.

If $\widehat{d} \leq 0$ (to a degree of statistical certainty), then the null of FSD is not rejected. Furthermore, if $\widehat{d} \leq 0$ and $\max\{d_1\} < 0$, then X FSD Y. On the other hand, if $\widehat{d} \leq 0$ and $\max\{d_2\} < 0$, then Y FSD X. If $\widehat{d} = \max\{d_1\} = \max\{d_2\} = 0$, then the (estimated) distributions of X and Y are identical. The test for SSD requires the following additional steps:

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

(ii) finding $\widehat{s} = \sqrt{\frac{NM}{N+M}} \min \{\max\{s_{1j}\}, \max\{s_{2j}\}\}$.

If $\widehat{s} \leq 0$ (to a degree of statistical certainty), then the null of SSD is not rejected. Moreover, if $\widehat{s} \leq 0$ and $\max\{s_{1j}\} < 0$, then X SSD Y; otherwise, if $\max\{s_{2j}\} < 0$, then Y SSD X.

A.2 The Recentered Bootstrap

To obtain recentered bootstrap p-values, we compute the relative frequency of $\{\widehat{d}^{**} > \widehat{d}\}$, where $\widehat{d}^{**}$ is the recentered bootstrap estimate of d. The recentering algorithm requires:

(i) generating bootstrap samples of size N (M) from X (Y),

(ii) computing the values of $\widehat{F}^*(z_j)$ and $\widehat{G}^*(z_j)$ for z_j, $j = 1, ..., J$, where the values of z_j used to analyze the original sample are utilized,

(iii) computing the differences $d_1^c(z_j) = \left[\widehat{F}^*(z_j) - \widehat{G}^*(z_j)\right] - \left[\widehat{F}(z_j) - \widehat{G}(z_j)\right]$ and $d_2^c(z_j) = \left[\widehat{G}^*(z_j) - \widehat{F}^*(z_j)\right] - \left[\widehat{G}(z_j) - \widehat{F}(z_j)\right]$, and

(iv) finding $\widehat{d}^{**} = \sqrt{\frac{NM}{N+M}} \min \{\max\{d_1^c\}, \max\{d_2^c\}\}$.

We then compute the relative frequency of $\{\widehat{d}^{**} > \widehat{d}\}$, where $\widehat{d}$ is the sample estimate of d.

Table B1. Summary Statistics.

Variable	More than Two Children			Two Children		
	Mean	SD	Obs	Mean	SD	Obs
Height for age (*z*-score)	-1.78	1.26	3486	-1.51	1.32	2276
Weight for age (*z*-score)	-1.79	1.28	3490	-1.57	1.33	2292
First two children are same sex (1 = yes)	0.53	0.50	4409	0.49	0.50	2863
First two children are daughters (1 = yes)	0.26	0.44	4409	0.24	0.42	2863
Age in months	74.81	37.70	4409	63.95	37.11	2768
Gender (1 = male)	0.52	0.50	4108	0.51	0.50	2863
First Child's Gender (1 = male)	0.50	0.50	4409	0.50	0.50	2863
Second Child's Gender (1 = male)	0.50	0.50	4409	0.52	0.50	2863
Father's Education						
Less than elementary school	0.53	0.50	4409	0.36	0.48	2863
Junior high school	0.14	0.34	4409	0.16	0.36	2863
Senior high school	0.17	0.37	4409	0.26	0.44	2863
University	0.09	0.28	4409	0.11	0.32	2863
Other	0.02	0.14	4409	0.02	0.13	2863
Mother's Education						
Less than elementary school	0.66	0.47	4409	0.45	0.50	2863
Junior high school	0.12	0.33	4409	0.19	0.39	2863
Senior high school	0.14	0.35	4409	0.27	0.44	2863
University	0.05	0.22	4409	0.08	0.27	2863
Other	0.01	0.10	4409	0.01	0.10	2863
Father's Height	161.41	5.18	4409	162.22	5.43	2863
Mother's Height	150.49	5.13	4409	150.87	4.84	2863
Father's Weight	56.39	8.17	4409	57.06	8.17	2863
Mother's Weight	52.03	9.19	4409	51.52	8.11	2863
Father's Age	40.98	7.48	4409	34.97	6.09	2863
Mother's Age	35.85	6.14	4409	30.14	5.37	2863
Father's Work Status (1 = Work)	0.91	0.28	4409	0.89	0.31	2863
Mother's Work Status (1 = Work)	0.55	0.50	4409	0.49	0.50	2863
Own Farm (1 = yes)	0.37	0.48	4409	0.37	0.48	4409
Father's Religion						
Islam	0.82	0.39	4409	0.81	0.40	2863
Protestant	0.06	0.23	4409	0.03	0.16	2863
Catholic	0.01	0.11	4409	0.02	0.13	2863
Hindu	0.04	0.19	4409	0.05	0.22	2863
Mother's Religion						
Islam	0.87	0.34	4409	0.88	0.32	2863
Protestant	0.06	0.23	4409	0.03	0.17	2863
Catholic	0.01	0.11	4409	0.02	0.13	2863
Hindu	0.04	0.19	4409	0.05	0.22	2863
Female Household Head (1 = yes)	0.09	0.29	4409	0.12	0.33	2863
Decision on Children's Health						
Jointly made by husband and wife	0.01	0.11	4409	0.02	0.13	2863
Only made by husband	0.11	0.31	4409	0.09	0.29	2863
Only made by wife	0.06	0.24	4409	0.05	0.23	2863
Otherwise	0.56	0.50	4409	0.56	0.50	2863
Days in Bed Due to Health Problem	0.36	1.39	4409	0.36	1.43	2863
Cough (1 = yes)	0.31	0.46	4409	0.37	0.48	2863
Fever (1 = yes)	0.01	0.11	4409	0.01	0.12	2863
Diarrhea (1 = yes)	0.19	0.39	4409	0.20	0.40	2863
Type of Dwelling						
Single unit single level	0.73	0.45	4409	0.76	0.42	2863
Single unit multiple levels	0.07	0.26	4409	0.07	0.26	2863
Duplex single level	0.06	0.24	4409	0.07	0.26	2863
Duplex multiple level	0.01	0.08	4409	0.01	0.08	2863
Multiple unit single level	0.03	0.16	4409	0.03	0.17	2863
Others	0.10	0.30	4409	0.05	0.22	2863

Table B1 (cont.). Summary Statistics.

Variable	More than Two Children			Two Children		
	Mean	SD	Obs	Mean	SD	Obs
House Size	76.26	68.89	4409	82.44	175.82	2863
Number of Rooms	5.41	2.57	4409	5.42	2.62	2863
Floor Materials						
Ceramic/Marble/Granite/Stone	0.14	0.34	4409	0.18	0.39	2863
Tiles/Terrazzo	0.21	0.40	4409	0.21	0.41	2863
Cement/Bricks	0.36	0.48	4409	0.36	0.48	2863
Lumber/Board	0.16	0.37	4409	0.12	0.32	2863
Dirt	0.12	0.32	4409	0.12	0.33	2863
Wall Materials						
Masonry (cement/brick)	0.59	0.49	4409	0.66	0.47	2863
Lumber/Board	0.29	0.45	4409	0.25	0.43	2863
Bamboo/Woven/Mat	0.12	0.32	4409	0.09	0.29	2863
Water sources for drinking						
Aqua/ Air mineral etc.	0.23	0.42	4409	0.26	0.44	2863
Pipe water	0.24	0.43	4409	0.30	0.46	2863
Well/Pump	0.34	0.47	4409	0.29	0.46	2863
Well water	0.08	0.26	4409	0.07	0.25	2863
Rain water	0.05	0.23	4409	0.03	0.17	2863
Others	0.02	0.14	4409	0.03	0.16	2863
Boil Water (1 = yes)	0.89	0.31	4409	0.91	0.28	2863
House surrounded by human/animal waste (1 = yes)	2.79	0.61	4406	2.86	0.51	2863
House surrounded by piles of trash (1 = yes)	2.71	0.70	4409	2.76	0.65	2863
House surrounded by stagnant water (1 = yes)	2.77	0.64	4409	2.81	0.59	2863
Stable under/next to house (1 = yes)	2.52	0.85	4406	2.60	0.80	2863
Sufficient ventilation (1 = yes)	1.51	0.87	4409	1.42	0.81	2863
Yard is cleaned up (1 = yes)	1.71	0.96	4409	1.57	0.90	2863
House has a moderately sized yard (1 = yes)	1.78	0.98	4409	1.77	0.97	2863
House has kitchen outside (1 = yes)	2.45	0.89	4406	2.48	0.88	2863
Region (1 = urban)	1.57	0.50	4409	1.51	0.50	2863
Number of Schools						
Elementary school	8.26	3.95	3501	8.76	4.45	2066
Junior high school	7.18	3.11	3501	7.66	3.29	2066
Senior high school	7.75	4.28	3496	8.59	4.72	2059
Minimum Distance of Schools						
Elementary school	0.38	0.56	3463	0.33	0.50	2050
Junior high school	1.86	2.86	3485	1.55	1.97	2054
Senior high school	4.03	4.71	3396	3.42	4.37	2028
Number of Health Facilities						
Hospital	2.54	1.70	3258	2.80	1.91	1969
Integrated health post	8.36	6.34	3493	9.36	6.35	2064
Private practice	27.21	11.07	3501	28.78	11.31	2066
Minimum Distance of Health Facilities						
Hospital	19.57	32.19	3128	15.42	25.84	1903
Integrated health post	0.62	3.82	3424	0.37	1.63	2035
Private practice	0.45	0.73	3475	0.44	0.71	2064

Notes: Data from Indonesia Family Life Survey 3 wave. Appropriate sample weights utilized. Additional controls not listed include 13 province dummies, ten birth order dummies, and dummy variables indicating missing values for father's height, mother's height, father's age, and house size.

Table B2. Estimates of Average Effects: OLS and Instrumental Variables.

	Height-for-Age		Weight-for-Age	
Model	**Coefficient**	**SE**	**Coefficient**	**SE**
OLS	-0.101†	(0.048)	-0.103†	(0.050)
IV (No Controls):				
Same Sex 2	-1.087	(0.674)	-1.509†	(0.750)
F-test	F = 16.40 [p = 0.00]		F = 15.30 [p = 0.00]	
Daughter 2	-0.673	(1.306)	-1.438	(1.573)
F-test	F = 4.05 [p = 0.04]		F = 3.40 [p = 0.07]	
Same Sex 2, Daughter 2	-1.106	(0.673)	-1.513†	(0.746)
F-test	F = 8.25 [p = 0.00]		F = 7.75 [p = 0.00]	
Sargan Test	$\chi^2 = 0.14$ [p = 0.71]		$\chi^2 = 0.003$ [p = 0.96]	
IV (With Controls):				
Same Sex 2	-2.328	(1.428)	-2.996‡	(1.569)
F-test	F = 8.67 [p = 0.00]		F = 9.13 [p = 0.00]	
Daughter 2	-2.328	(1.428)	-2.996‡	(1.569)
F-test	F = 8.67 [p = 0.00]		F = 9.13 [p = 0.00]	
Same Sex 2, Daughter 2	-2.278	(1.417)	-2.975‡	(1.563)
F-test	F = 4.35 [p = 0.01]		F = 4.57 [p = 0.01]	
Sargan Test	$\chi^2 = 0.46$ [p = 0.50]		$\chi^2 = 0.15$ [p = 0.70]	

NOTES: Control variables include:age in months, gender, gender of the first child, gender of the second child (except in the models that use Same Sex 2 and Daughter 2 as instruments), birth order, days in bed, cough in the past four weeks, fever in the past four weeks, diaarrhea in the past four weeks, parents' education, parents' height, parents' weight, parents' working status, parents' religion, whether or not household head is female, household decision making on children's health, household sanitary conditions, dwelling, house size, floor, wall, water sources, whether or not boil water, whether or not own a farm, urban/rural dummy, province dummies, whether or not parents' height are missing, whether or not father's age is missing, and whether or not house size is missing. F-test refers to the test of significance of the instrument(s) in the first-stage. Sargan test is an overidentification test. Appropriate sample weights used. † (‡) indicates statistical significance at the 5% (10%) level.

Table B3. Partial Residual Stochastic Dominance Tests: Height-for-Age, Alternative Control Set.

	A. Without Instrument				**B. IV 1 (*Same Sex 2*)**				**C. IV 2 (*Daughter 2*)**			
	Simple Bootstrap		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***X FSD Y***				***X FSD Y***				***X FSD Y***			
$d_{1,MAX}$	-0.408				-0.228				-0.389			
$d_{2,MAX}$	20.759				14.096				25.579			
d	-0.408				-0.228				-0.389			
$Pr\{d_1^*\leq 0\}$	0.606		0.744		0.582		0.686		0.750		0.578	
$Pr\{d_2^*\leq 0\}$	0.334		0.742		0.320		0.686		0.164		0.762	
$Pr\{d^*\leq 0\}$	0.940		0.758		0.902		0.702		0.914		0.772	
$Pr\{d_1^*\geq d_1\}$	0.586		0.998		0.482		0.996		0.824		1.000	
$Pr\{d_2^*\geq d_2\}$	0.456		0.094		0.526		0.146		0.404		0.076	
$Pr\{d^*\geq d\}$	0.414		0.996		0.198		0.988		0.824		0.998	
$s_{1,MAX}$	-0.478				-0.434				-0.824			
$s_{2,MAX}$	4032.949				2689.212				5889.007			
s	-0.478				-0.434				-0.824			
$Pr\{s_1^*\leq 0\}$	0.630		0.824		0.656		0.794		0.756		0.664	
$Pr\{s_2^*\leq 0\}$	0.346		0.856		0.328		0.792		0.220		0.842	
$Pr\{s^*\leq 0\}$	0.976		0.952		0.984		0.916		0.976		0.938	
$Pr\{s_1^*\geq s_1\}$	0.804		0.930		0.518		0.918		0.570		0.970	
$Pr\{s_2^*\geq s_2\}$	0.452		0.060		0.526		0.102		0.402		0.026	
$Pr\{s^*\geq s\}$	0.548		0.860		0.280		0.820		0.570		0.738	

Notes: First-stage regressions include controls for:age in months, gender, gender of first two children, birthorder, days in bed, cough in the past four weeks, fever in the past four weeks, diaarrhea in the past four weeks, parents' education, parents' height, parents' weight, parents' working status, parents' religion, whether or not household head is female, household decision making on children's health, household sanitary conditions, dwelling type, house size, number of rooms in house, floor type, wall type, drinking water source, whether or not boil water, whether or not own a farm, urban/rural dummy, province dummies, whether or not parents' height are missing, whether or not father's age is missing, whether or not house size is missing, number of elementary schools, junior high schools, and senior high schools, minimum distance of elementary schools, junior high schools, and senior high schools, number of hospitals, integrated health posts, and private practictians, and minimum distance of hospitals, integrated health posts, and private practictians. See Table 1 and text for further details.

Table B4. Partial Residual Stochastic Dominance Tests: Weight-for-Age, Alternative Control Set.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***Y SSD X***				***Y FSD X***				***Y FSD X***			
$d_{1,MAX}$	0.122				29.896				13.815			
$d_{2,MAX}$	3.227				-0.569				-0.367			
d	0.122				-0.569				-0.367			
$\Pr\{d_1^*\leq 0\}$	0.410		0.764		0.072		0.740		0.268		0.746	
$\Pr\{d_2^*\leq 0\}$	0.468		0.774		0.900		0.744		0.620		0.652	
$\Pr\{d^*\leq 0\}$	0.878		0.792		0.972		0.752		0.888		0.756	
$\Pr\{d_1^*\geq d_1\}$	0.548		0.198		0.552		0.068		0.558		0.088	
$\Pr\{d_2^*\geq d_2\}$	0.490		0.142		0.732		0.998		0.844		1.000	
$\Pr\{d^*\geq d\}$	0.076		0.136		0.726		0.998		0.816		0.998	
$s_{1,MAX}$	-0.169				7325.597				2927.062			
$s_{2,MAX}$	556.370				-0.628				-0.367			
s	-0.169				-0.628				-0.367			
$\Pr\{s_1^*\leq 0\}$	0.504		0.840		0.080		0.812		0.276		0.832	
$\Pr\{s_2^*\leq 0\}$	0.474		0.860		0.912		0.842		0.692		0.752	
$\Pr\{s^*\leq 0\}$	0.978		0.954		0.992		0.922		0.968		0.942	
$\Pr\{s_1^*\geq s_1\}$	0.512		0.918		0.540		0.032		0.552		0.052	
$\Pr\{s_2^*\geq s_2\}$	0.486		0.098		0.600		0.944		0.824		0.912	
$\Pr\{s^*\geq s\}$	0.048		0.808		0.592		0.936		0.576		0.770	

Notes: See Table B3.

Table B5. Full Residual Stochastic Dominance Tests: Height-for-Age, Alternative Control Set.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap	
	X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
	Two Kids	More than Two	Two Kids	More than Two	Two Kids	More than Two	Two Kids	More than Two	Two Kids	More than Two	Two Kids	More than Two
Observed Ranking	***X SSD Y***				***X SSD Y***				***X SSD Y***			
$d_{1,MAX}$	1.190				0.397				0.038			
$d_{2,MAX}$	1.364				1.756				2.251			
d	1.190				0.397				0.038			
$Pr\{d_1^* \leq 0\}$	0.000		0.396		0.000		0.710		0.036		0.726	
$Pr\{d_2^* \leq 0\}$	0.000		0.750		0.000		0.712		0.000		0.722	
$Pr\{d^* \leq 0\}$	0.000		0.750		0.000		0.712		0.036		0.726	
$Pr\{d_1^* \geq d_1\}$	0.788		0.142		0.976		0.252		0.932		0.270	
$Pr\{d_2^* \geq d_2\}$	0.890		0.136		0.914		0.170		0.600		0.242	
$Pr\{d^* \geq d\}$	0.706		0.102		0.976		0.204		0.932		0.266	
$s_{1,MAX}$	-0.101				-0.265				-0.111			
$s_{2,MAX}$	129.275				196.651				264.084			
s	-0.101				-0.265				-0.111			
$Pr\{s_1^* \leq 0\}$	0.572		0.428		0.908		0.744		0.834		0.748	
$Pr\{s_2^* \leq 0\}$	0.000		0.810		0.000		0.748		0.004		0.760	
$Pr\{s^* \leq 0\}$	0.572		0.842		0.908		0.782		0.838		0.786	
$Pr\{s_1^* \geq s_1\}$	0.446		0.992		0.340		0.998		0.276		1.000	
$Pr\{s_2^* \geq s_2\}$	0.870		0.090		0.880		0.078		0.580		0.082	
$Pr\{s^* \geq s\}$	0.446		0.986		0.340		0.986		0.274		0.978	

Notes: See Tables 5 and A3.

Table B6. Full Residual Stochastic Dominance Tests: Weight-for-Age, Alternative Control Set.

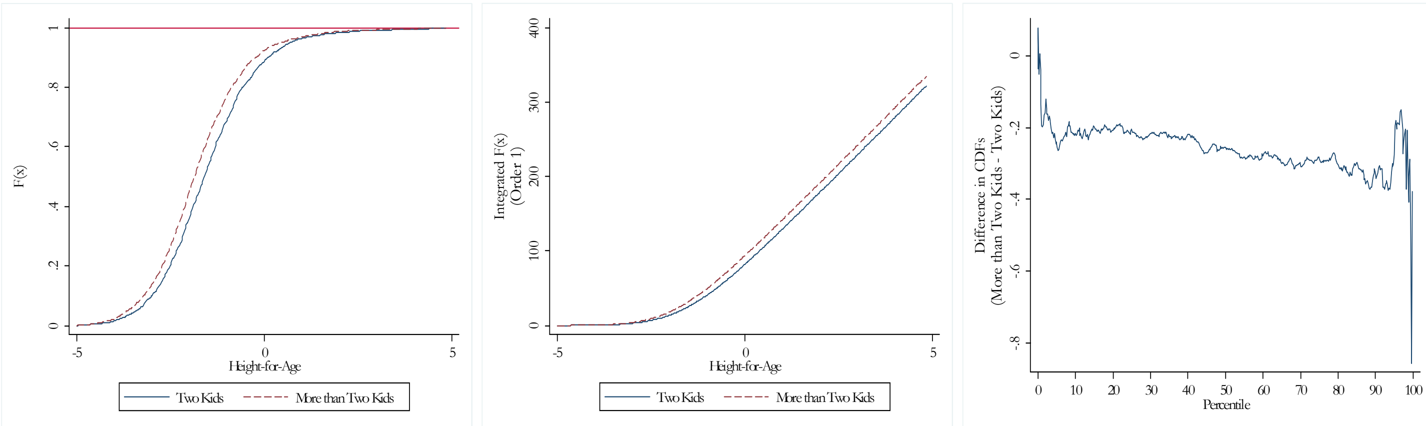
	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**		**Simple Bootstrap**		**Recentered Bootstrap**	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***X FSD Y***				***None***				***X FSD Y***			
$d_{1,MAX}$	0.543				0.168				0.003			
$d_{2,MAX}$	2.230				2.102				2.762			
d	0.543				0.168				0.003			
$Pr\{d_1^* \leq 0\}$	0.000		0.428		0.000		0.716		0.152		0.716	
$Pr\{d_2^* \leq 0\}$	0.000		0.732		0.000		0.716		0.000		0.720	
$Pr\{d^* \leq 0\}$	0.000		0.732		0.000		0.720		0.152		0.720	
$Pr\{d_1^* \geq d_1\}$	0.960		0.202		0.994		0.266		0.846		0.284	
$Pr\{d_2^* \geq d_2\}$	0.828		0.130		0.930		0.156		0.576		0.248	
$Pr\{d^* \geq d\}$	0.960		0.158		0.994		0.238		0.846		0.280	
$s_{1,MAX}$	-0.286				0.001				-0.119			
$s_{2,MAX}$	298.900				314.350				484.223			
s	-0.286				0.001				-0.119			
$Pr\{s_1^* \leq 0\}$	0.798		0.466		0.946		0.754		0.924		0.746	
$Pr\{s_2^* \leq 0\}$	0.000		0.774		0.000		0.754		0.000		0.748	
$Pr\{s^* \leq 0\}$	0.798		0.812		0.946		0.796		0.924		0.778	
$Pr\{s_1^* \geq s_1\}$	0.342		0.998		0.054		0.246		0.204		0.986	
$Pr\{s_2^* \geq s_2\}$	0.756		0.076		0.800		0.066		0.550		0.068	
$Pr\{s^* \geq s\}$	0.342		0.986		0.054		0.204		0.204		0.980	

Notes: See Table B5.

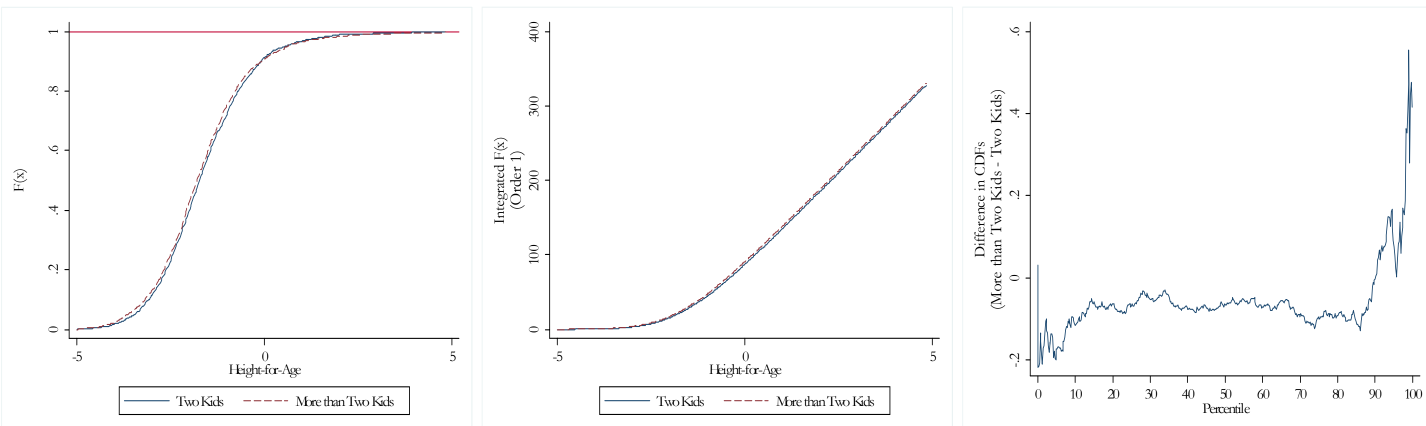
Table 1. Unconditional Stochastic Dominance Tests: Height-for-Age.

	A. Without Instrument		B. IV 1 (*Same Sex 2*)		C. IV 2 (*Daughter 2*)	
	Simple Bootstrap	Recentered Bootstrap	Simple Bootstrap	Recentered Bootstrap	Simple Bootstrap	Recentered Bootstrap
	X Two Kids; ***Y*** More than Two	***X*** Two Kids; ***Y*** More than Two	***X*** Two Kids; ***Y*** More than Two	***X*** Two Kids; ***Y*** More than Two	***X*** Two Kids; ***Y*** More than Two	***X*** Two Kids; ***Y*** More than Two
Observed Ranking	***X FSD Y***		***X SSD Y***		***None***	
$d_{1,MAX}$	-0.110		0.403		0.517	
$d_{2,MAX}$	3.824		1.309		1.186	
d	-0.110		0.403		0.517	
$\Pr\{d_1^* \le 0\}$	0.492	0.776	0.000	0.748	0.000	0.786
$\Pr\{d_2^* \le 0\}$	0.000	0.776	0.000	0.748	0.000	0.784
$\Pr\{d^* \le 0\}$	0.492	0.780	0.000	0.750	0.000	0.788
$\Pr\{d_1^* \ge d_1\}$	0.902	1.000	0.632	0.160	0.754	0.112
$\Pr\{d_2^* \ge d_2\}$	0.614	0.006	0.726	0.078	0.600	0.044
$\Pr\{d^* \ge d\}$	0.902	1.000	0.630	0.062	0.694	0.022
$s_{1,MAX}$	-0.110		-0.108		26.260	
$s_{2,MAX}$	707.865		204.947		56.001	
s	-0.110		-0.108		26.260	
$\Pr\{s_1^* \le 0\}$	0.846	0.826	0.798	0.810	0.064	0.822
$\Pr\{s_2^* \le 0\}$	0.000	0.840	0.000	0.810	0.142	0.844
$\Pr\{s^* \le 0\}$	0.846	0.894	0.798	0.874	0.206	0.884
$\Pr\{s_1^* \ge s_1\}$	0.574	0.998	0.516	0.998	0.566	0.104
$\Pr\{s_2^* \ge s_2\}$	0.526	0.002	0.530	0.038	0.526	0.074
$\Pr\{s^* \ge s\}$	0.574	0.994	0.516	0.996	0.238	0.012

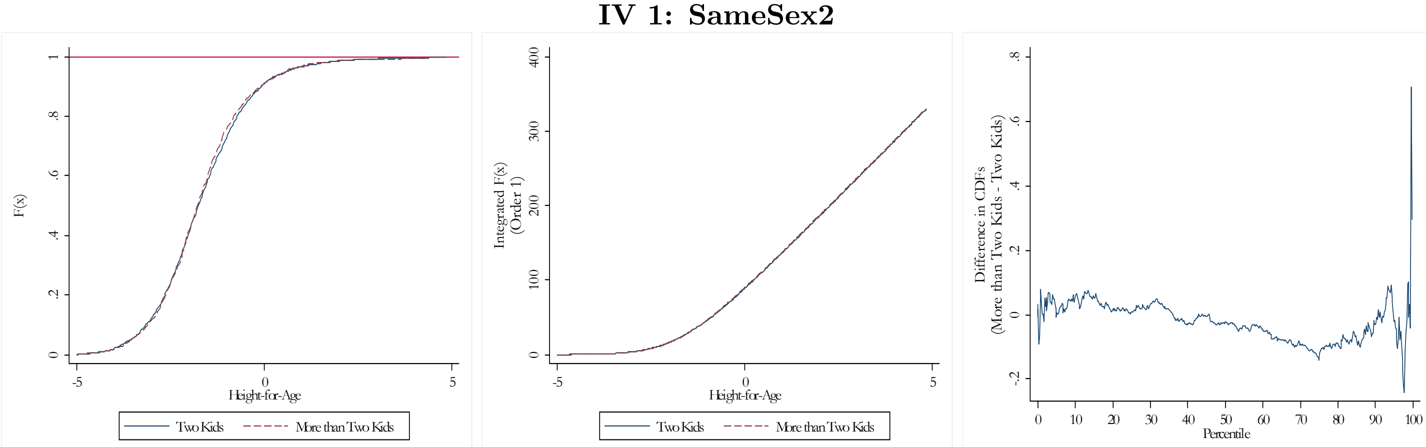
Notes: Bootstrap results based on 500 repetitions. Appropriate sample weights utilized. See text for further details.



No Instrument



IV 1: SameSex2



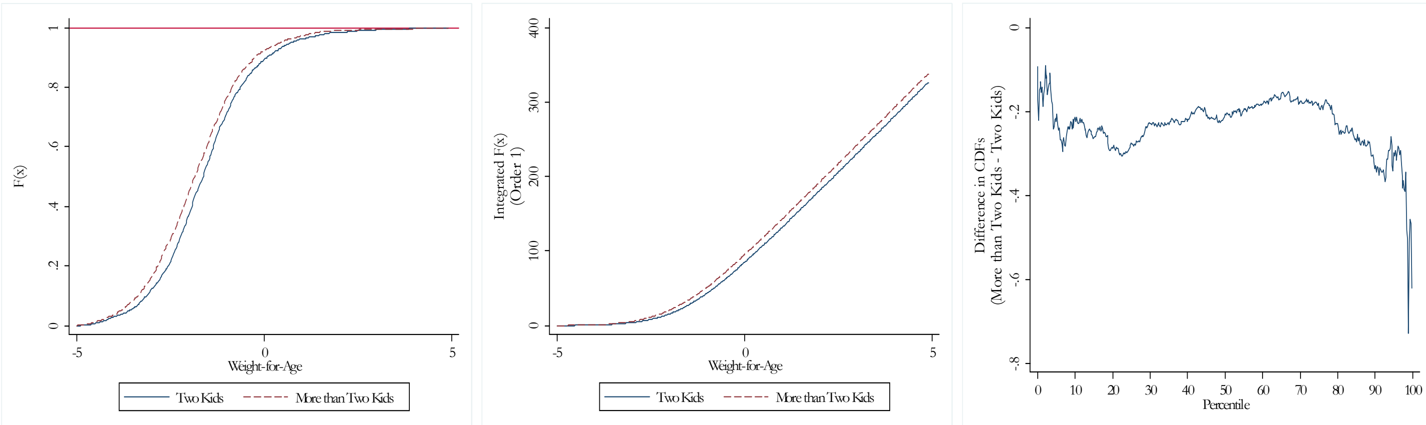
IV 2: Daughter2

Figure 1. Unconditional Empirical Cumulative Distribution Functions: Height-for-Age z-scores.
Note: First row splits the sample using *MoreThan2*; second (third) row splits the sample using *SameSex2* (*Daughter2*). In each row, column 1 plots the CDF, column 2 plots the integrated CDF, and column 3 plots the difference in the CDFs at each percentile. Appropriate sample weights used. See text for further details.

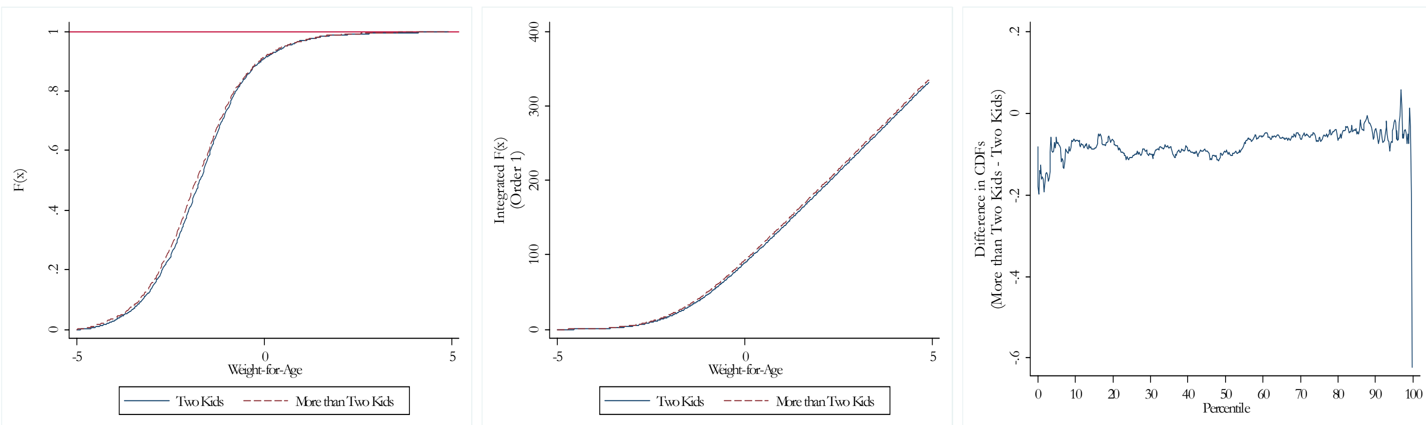
Table 2. Unconditional Stochastic Dominance Tests: Weight-for-Age.

	A. Without Instrument		B. IV 1 (*Same Sex 2*)		C. IV 2 (*Daughter 2*)	
	Simple Bootstrap	Recentered Bootstrap	Simple Bootstrap	Recentered Bootstrap	Simple Bootstrap	Recentered Bootstrap
	X Two Kids / ***Y*** More than Two	***X*** Two Kids / ***Y*** More than Two	***X*** Two Kids / ***Y*** More than Two	***X*** Two Kids / ***Y*** More than Two	***X*** Two Kids / ***Y*** More than Two	***X*** Two Kids / ***Y*** More than Two
Observed Ranking	***X FSD Y***		***X SSD Y***		***None***	
$d_{1,MAX}$	-0.173		0.106		0.347	
$d_{2,MAX}$	2.903		1.560		0.758	
d	-0.173		0.106		0.347	
$Pr\{d_1^*\leq 0\}$	0.700	0.788	0.070	0.780	0.008	0.794
$Pr\{d_2^*\leq 0\}$	0.000	0.790	0.000	0.786	0.000	0.788
$Pr\{d^*\leq 0\}$	0.700	0.792	0.070	0.786	0.008	0.794
$Pr\{d_1^*\geq d_1\}$	0.800	1.000	0.784	0.174	0.822	0.114
$Pr\{d_2^*\geq d_2\}$	0.750	0.012	0.548	0.046	0.782	0.096
$Pr\{d^*\geq d\}$	0.800	1.000	0.784	0.116	0.808	0.040
$s_{1,MAX}$	-0.203		-0.230		17.639	
$s_{2,MAX}$	636.727		208.046		84.654	
s	-0.203		-0.230		17.639	
$Pr\{s_1^*\leq 0\}$	0.930	0.856	0.858	0.840	0.182	0.836
$Pr\{s_2^*\leq 0\}$	0.000	0.832	0.000	0.830	0.056	0.820
$Pr\{s^*\leq 0\}$	0.930	0.902	0.858	0.890	0.238	0.868
$Pr\{s_1^*\geq s_1\}$	0.436	0.996	0.516	1.000	0.558	0.096
$Pr\{s_2^*\geq s_2\}$	0.490	0.004	0.510	0.030	0.520	0.090
$Pr\{s^*\geq s\}$	0.436	0.994	0.516	0.998	0.362	0.020

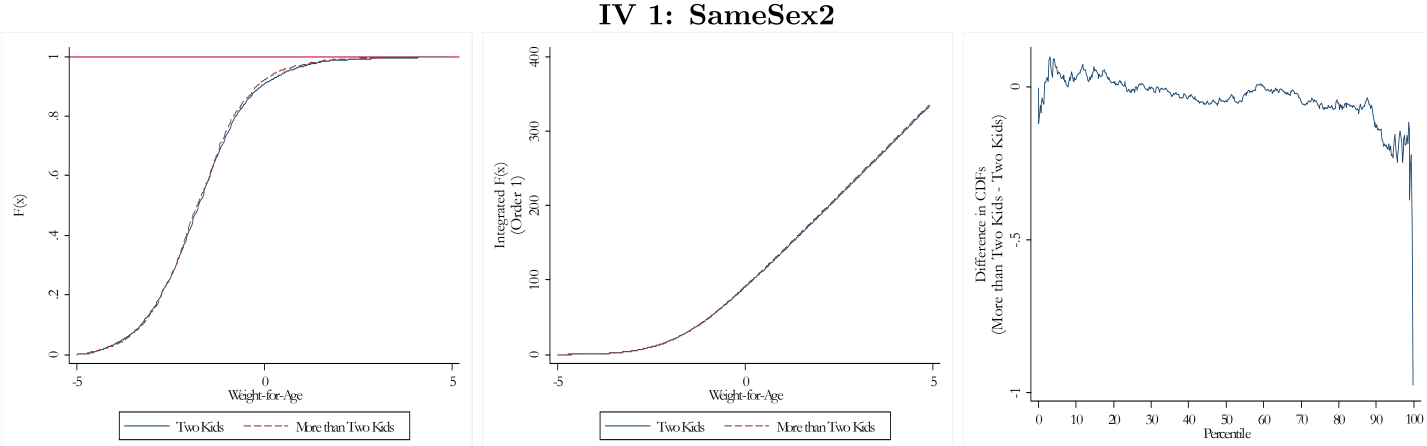
Notes: See Table 1.



No Instrument



IV 1: SameSex2



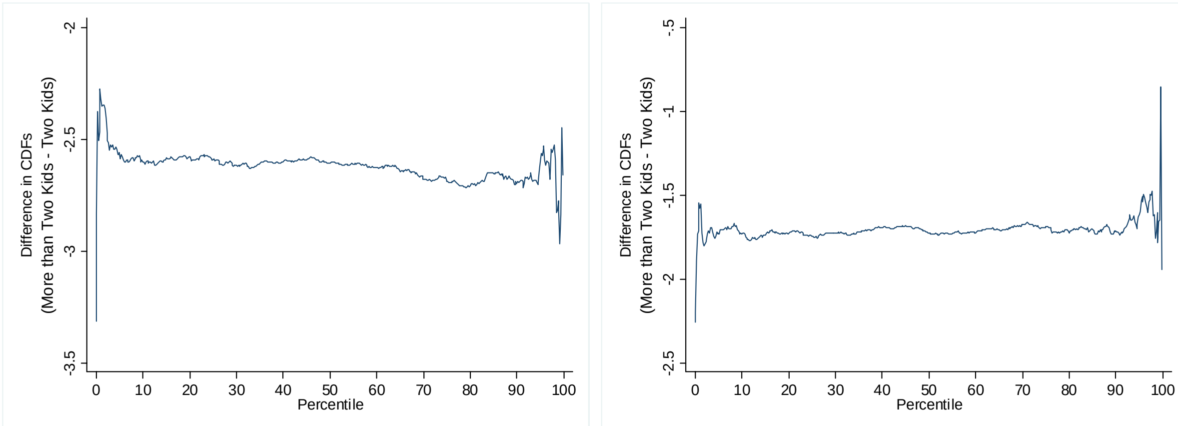
IV 2: Daughter2

Figure 2. Unconditional Empirical Cumulative Distribution Functions: Weight-for-Age z-scores.
Note: See Figure 1.

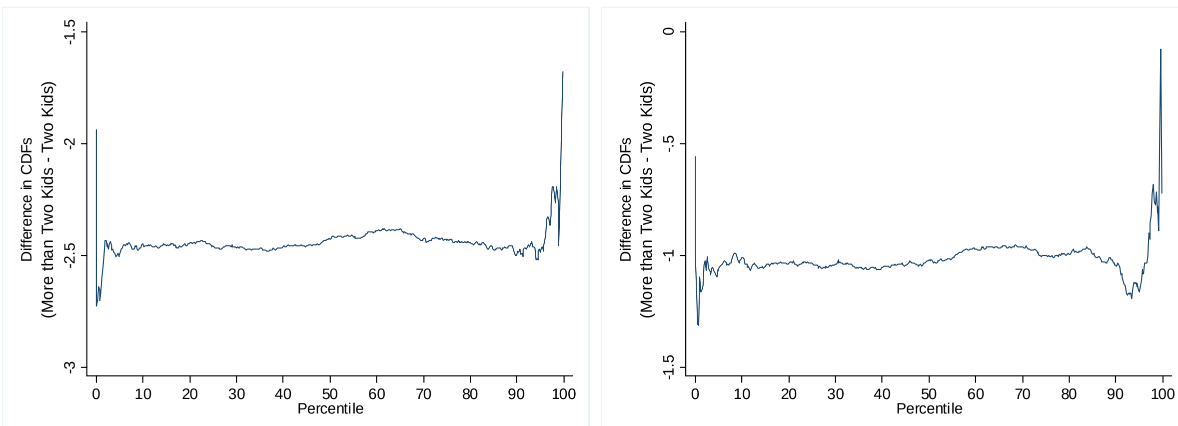
Table 3. Partial Residual Stochastic Dominance Tests: Height-for-Age.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***X FSD Y***				***X FSD Y***				***X FSD Y***			
$d_{1,MAX}$	-0.584				-0.554				-0.107			
$d_{2,MAX}$	30.484				30.379				6.461			
d	-0.584				-0.554				-0.107			
$Pr\{d_1^*\leq 0\}$	0.844		0.774		0.826		0.790		0.592		0.762	
$Pr\{d_2^*\leq 0\}$	0.072		0.770		0.104		0.616		0.298		0.768	
$Pr\{d^*\leq 0\}$	0.916		0.790		0.930		0.800		0.890		0.774	
$Pr\{d_1^*\geq d_1\}$	0.842		1.000		0.536		1.000		0.446		0.994	
$Pr\{d_2^*\geq d_2\}$	0.490		0.090		0.494		0.266		0.584		0.118	
$Pr\{d^*\geq d\}$	0.840		1.000		0.504		1.000		0.164		0.992	
$s_{1,MAX}$	-0.584				-0.640				-0.235			
$s_{2,MAX}$	6582.045				6322.549				1202.067			
s	-0.584				-0.640				-0.235			
$Pr\{s_1^*\leq 0\}$	0.850		0.834		0.860		0.878		0.606		0.848	
$Pr\{s_2^*\leq 0\}$	0.114		0.858		0.110		0.666		0.344		0.860	
$Pr\{s^*\leq 0\}$	0.964		0.938		0.970		0.938		0.950		0.952	
$Pr\{s_1^*\geq s_1\}$	0.826		0.968		0.618		0.958		0.454		0.928	
$Pr\{s_2^*\geq s_2\}$	0.490		0.044		0.490		0.032		0.576		0.098	
$Pr\{s^*\geq s\}$	0.776		0.902		0.580		0.938		0.150		0.838	

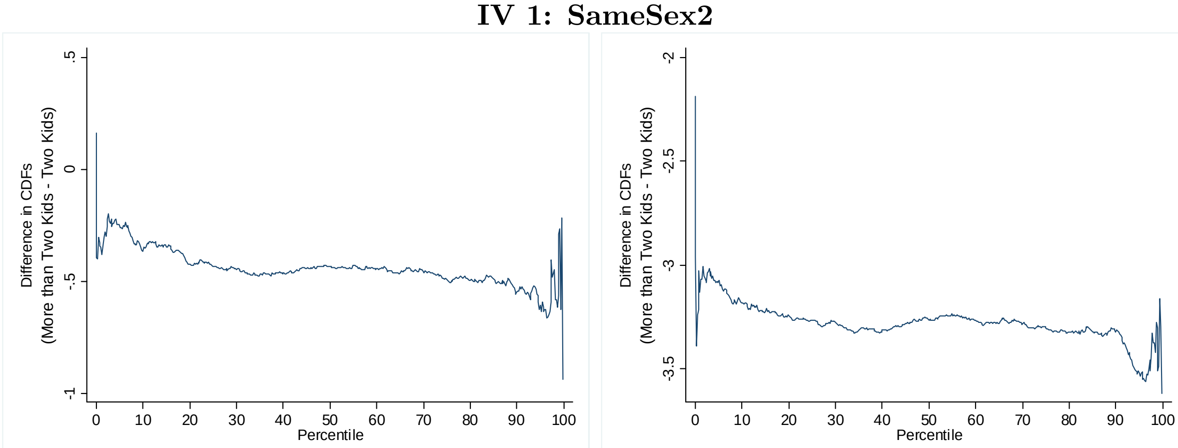
Notes: First-stage regressions include controls for:age in months, gender, gender of first two children, birthorder, days in bed, cough in the past four weeks, fever in the past four weeks, diaarrhea in the past four weeks, parents' education, parents' height, parents' weight, parents' working status, parents' religion, whether or not household head is female, household decision making on children's health, household sanitary conditions, dwelling, house size, floor, wall,water sources, whether or not boil water, whether or not own a farm, urban/rural dummy, province dummies, whether or not parents' height are missing, whether or not father's age is missing, and whether or not house size is missing.
See Table 1 and text for further details.



No Instrument



IV 1: SameSex2



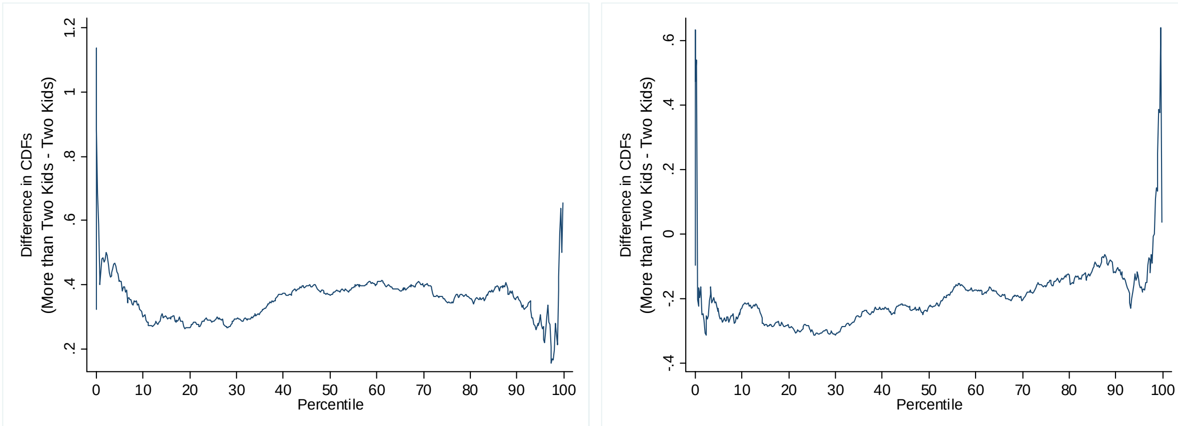
IV 2: Daughter2

Figure 3. Partial Residual Empirical Cumulative Distribution Functions: Height-for-Age z-scores.
Note: Distributions obtained using equation (9) in the text. In each rown, the controls used to obtain the plots in column 1 includes those listed in the Appendix, Table B2; the contols used to obtain the plots in column 2 is the same as in column 1 except the number and minimum distance to schools and health facilities are included (see the Appendix, Table B3). See text and Figure 1 for further details.

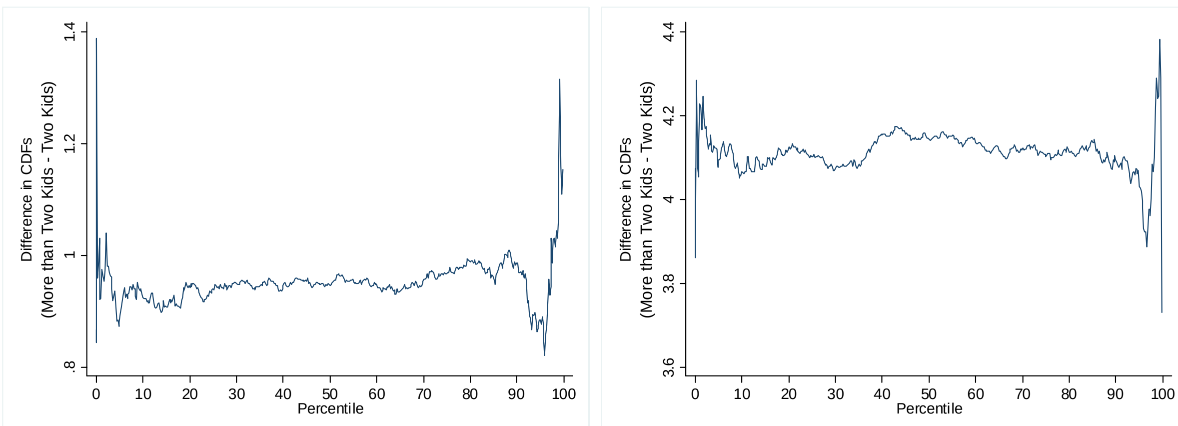
Table 4. Partial Residual Stochastic Dominance Tests: Weight-for-Age.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***Y FSD X***				***Y FSD X***				***Y FSD X***			
$d_{1,MAX}$	5.711				13.531				31.908			
$d_{2,MAX}$	-0.160				-0.448				-0.441			
d	-0.160				-0.448				-0.441			
$\Pr\{d_1^* \le 0\}$	0.376		0.776		0.280		0.756		0.058		0.760	
$\Pr\{d_2^* \le 0\}$	0.536		0.792		0.654		0.764		0.906		0.754	
$\Pr\{d^* \le 0\}$	0.912		0.804		0.934		0.774		0.964		0.768	
$\Pr\{d_1^* \ge d_1\}$	0.518		0.132		0.514		0.112		0.458		0.060	
$\Pr\{d_2^* \ge d_2\}$	0.472		0.986		0.516		0.998		0.624		1.000	
$\Pr\{d^* \ge d\}$	0.128		0.976		0.368		0.996		0.622		1.000	
$s_{1,MAX}$	1083.104				2825.991				8831.956			
$s_{2,MAX}$	-0.482				-0.667				-0.441			
s	-0.482				-0.667				-0.441			
$\Pr\{s_1^* \le 0\}$	0.396		0.850		0.306		0.828		0.058		0.858	
$\Pr\{s_2^* \le 0\}$	0.572		0.854		0.668		0.854		0.932		0.832	
$\Pr\{s^* \le 0\}$	0.968		0.940		0.974		0.936		0.990		0.944	
$\Pr\{s_1^* \ge s_1\}$	0.518		0.112		0.510		0.082		0.454		0.016	
$\Pr\{s_2^* \ge s_2\}$	0.486		0.934		0.500		0.910		0.612		0.968	
$\Pr\{s^* \ge s\}$	0.224		0.870		0.358		0.854		0.568		0.886	

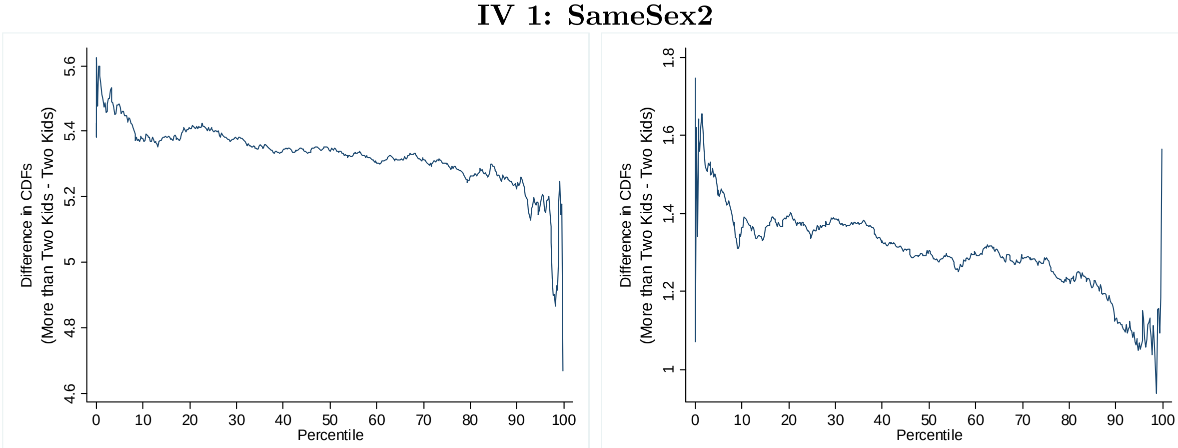
Notes: See Table 3.



No Instrument



IV 1: SameSex2



IV 2: Daughter2

Figure 4. Partial Residual Empirical Cumulative Distribution Functions: Weight-for-Age z-scores.
Note: See Figure 3.

Table 5. Full Residual Stochastic Dominance Tests: Height-for-Age.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***None***				***X SSD Y***				***X SSD Y***			
$d_{1,MAX}$	0.094				0.238				0.017			
$d_{2,MAX}$	2.383				1.988				2.358			
d	0.094				0.238				0.017			
$Pr\{d_1^* \leq 0\}$	0.040		0.448		0.000		0.612		0.120		0.784	
$Pr\{d_2^* \leq 0\}$	0.000		0.750		0.000		0.614		0.000		0.784	
$Pr\{d^* \leq 0\}$	0.040		0.750		0.000		0.616		0.120		0.784	
$Pr\{d_1^* \geq d_1\}$	0.884		0.534		0.896		0.364		0.852		0.216	
$Pr\{d_2^* \geq d_2\}$	0.726		0.096		0.784		0.306		0.622		0.202	
$Pr\{d^* \geq d\}$	0.884		0.196		0.896		0.336		0.852		0.216	
$s_{1,MAX}$	0.773				-0.222				-0.035			
$s_{2,MAX}$	385.064				243.636				354.058			
s	0.773				-0.222				-0.035			
$Pr\{s_1^* \leq 0\}$	0.646		0.502		0.970		0.628		0.696		0.790	
$Pr\{s_2^* \leq 0\}$	0.000		0.796		0.000		0.634		0.000		0.790	
$Pr\{s^* \leq 0\}$	0.646		0.850		0.970		0.652		0.696		0.796	
$Pr\{s_1^* \geq s_1\}$	0.180		0.494		0.380		0.998		0.362		1.000	
$Pr\{s_2^* \geq s_2\}$	0.570		0.028		0.636		0.064		0.534		0.052	
$Pr\{s^* \geq s\}$	0.180		0.114		0.380		0.994		0.362		0.998	

Notes: See Table 3 and text for further details.

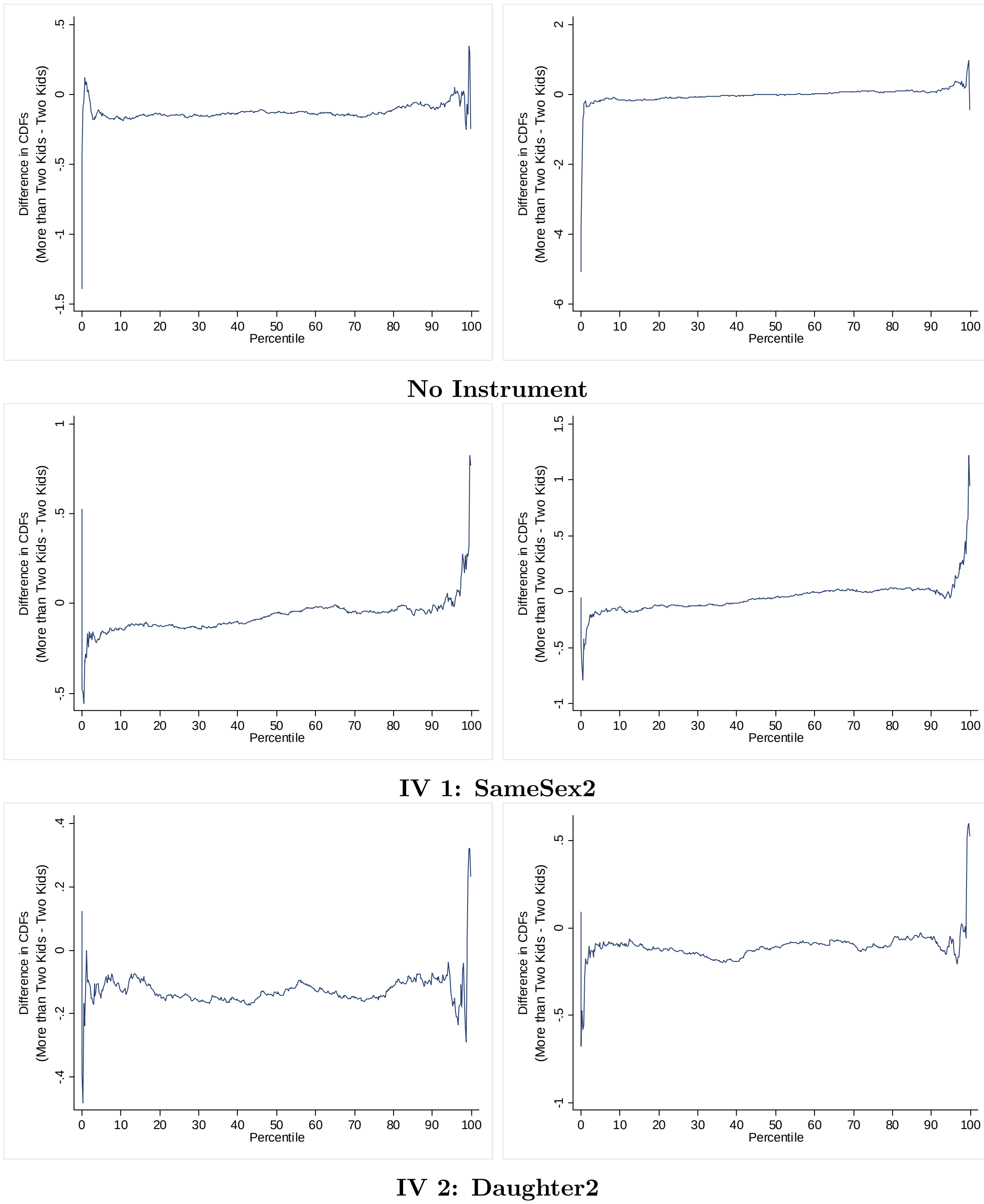
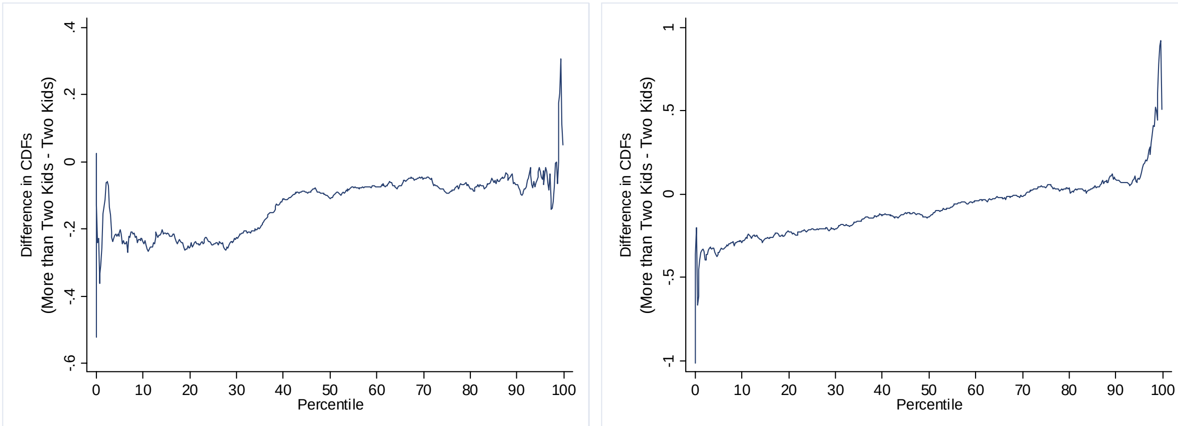


Figure 5. Full Residual Empirical Cumulative Distribution Functions: Height-for-Age z-scores.
Note: Distributions obtained using equation (10) in the text. See text and Figures 1 and 3 for further details.

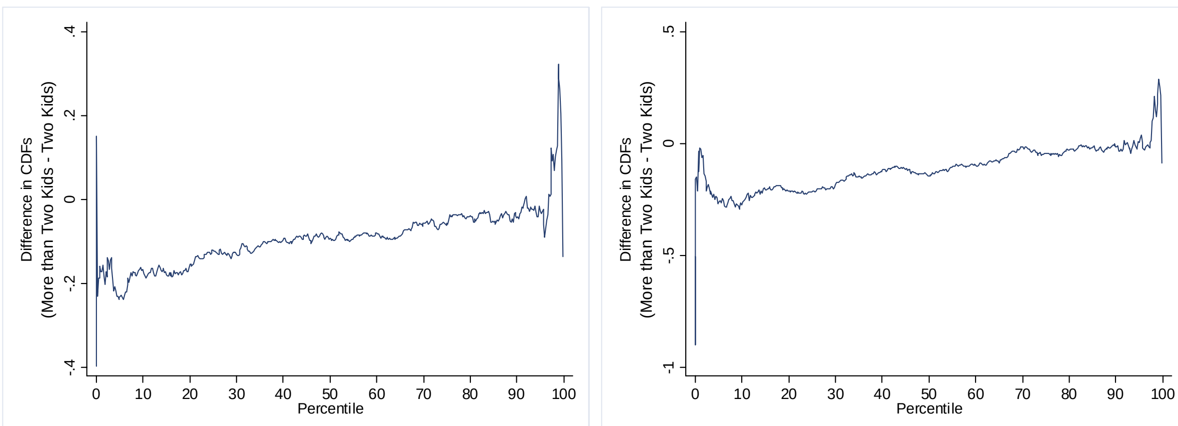
Table 6. Full Residual Stochastic Dominance Tests: Weight-for-Age.

	A. Without Instrument				B. IV 1 (*Same Sex 2*)				C. IV 2 (*Daughter 2*)			
	Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap		Simple Bootstrap		Recentered Bootstrap	
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
	Two Kids	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**	**Two Kids**	**More than Two**
Observed Ranking	***X SSD Y***				***X SSD Y***				***X SSD Y***			
$d_{1,MAX}$	0.092				0.229				0.028			
$d_{2,MAX}$	2.867				1.693				2.512			
d	0.092				0.229				0.028			
$Pr\{d_1^*\leq 0\}$	0.036		0.480		0.002		0.738		0.300		0.762	
$Pr\{d_2^*\leq 0\}$	0.000		0.746		0.000		0.740		0.000		0.764	
$Pr\{d^*\leq 0\}$	0.036		0.748		0.002		0.742		0.300		0.764	
$Pr\{d_1^*\geq d_1\}$	0.906		0.504		0.880		0.214		0.608		0.238	
$Pr\{d_2^*\geq d_2\}$	0.652		0.110		0.898		0.172		0.570		0.200	
$Pr\{d^*\geq d\}$	0.906		0.204		0.880		0.174		0.608		0.230	
$s_{1,MAX}$	-0.233				-0.169				-0.038			
$s_{2,MAX}$	391.432				305.363				466.253			
s	-0.233				-0.169				-0.038			
$Pr\{s_1^*\leq 0\}$	0.956		0.500		0.986		0.770		0.864		0.794	
$Pr\{s_2^*\leq 0\}$	0.000		0.786		0.000		0.798		0.000		0.812	
$Pr\{s^*\leq 0\}$	0.956		0.808		0.986		0.832		0.864		0.844	
$Pr\{s_1^*\geq s_1\}$	0.374		0.996		0.292		0.996		0.184		0.986	
$Pr\{s_2^*\geq s_2\}$	0.584		0.026		0.576		0.046		0.518		0.030	
$Pr\{s^*\geq s\}$	0.374		0.990		0.292		0.958		0.184		0.946	

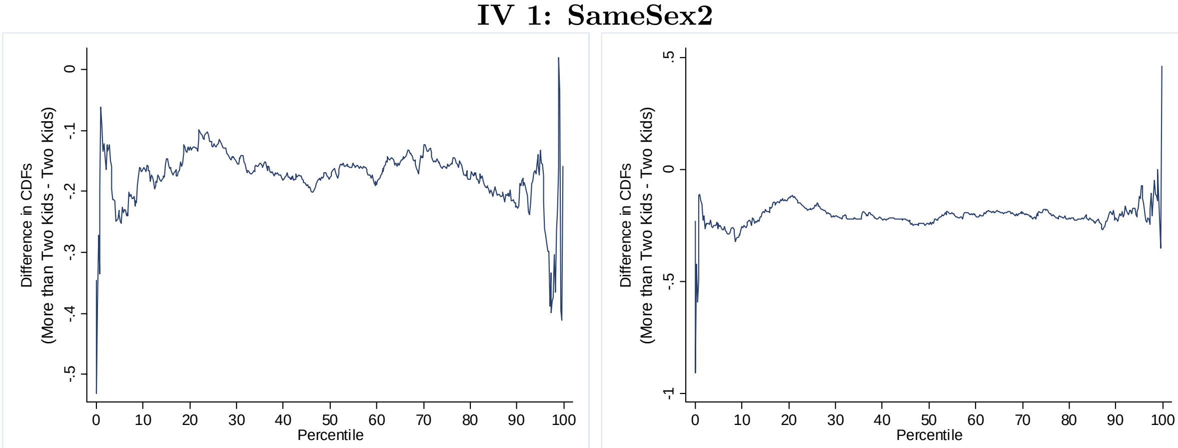
Notes: See Table 5.



No Instrument



IV 1: SameSex2



IV 2: Daughter2

Figure 6. Full Residual Empirical Cumulative Distribution Functions: Weight-for-Age z-scores.
Note: See Figure 5.