School of Economics

Working Paper Series

Family Size Effects on Child Health: Evidence on the Quantity-Quality Trade-off using the NLSY

Kabir Dasgupta and Keisha T.-Solomon

Working Paper – 17/04

Family Size Effects on Child Health: Evidence on the Quantity-Quality Trade-off using the NLSY

Kabir Dasgupta* & Keisha T.-Solomon✓
May 2017

Abstract
In this study, we use matched mother-child data from the National Longitudinal Surveys to study family size effects on child health. Focussing on body weight indicators as our health outcome of interest, we examine the effects of exogenous variations in family size on child health. We find no significant empirical support in favor of the quantity-quality trade-off theory. To combat potential empirical concerns associated with cross-sectional analysis, we make use of the panel aspects of the data and employ child fixed effects. Findings from our panel data analysis suggest that birth of a younger sibling is related to a decline in the likelihood of being overweight. Furthermore, birth of younger siblings at higher parities leads to a significant drop in the likelihood of obesity (but may also lead to an increase in the child's probability of being underweight).

JEL Classification: J13; I10

Keywords: Family Size; Children; Underweight; Overweight; Obesity; Health; Instrumental Variables; Fixed Effects.

*Address: New Zealand Work Research Institute, Auckland University of Technology, 5th Floor, 120 Mayoral Drive, Auckland 1010, New Zealand, e-mail- kabir.dasgupta@aut.ac.nz.
✓Address: Department of Economics, Temple University, 1301 Cecil B. Moore Avenue, Ritter Annex, Philadelphia, PA 19122, e-mail- k.solomon@temple.edu.
We are grateful to Gail Pacheco, Keshar Ghimire, and all participants of CSDA's 2017 workshop on children and youth for their valuable comments on this study. We are solely responsible for all errors committed in this study.

1. Introduction

The primary objective of this study is to investigate family size effects on child health. Family size effects on children's human capital outcomes have been studied extensively in the economic literature following the quantity-quality trade-off theory developed by Becker and his co-authors (Becker 1960; Willis 1973; Becker & Lewis 1973; Becker & Tomes 1976). The theory predicts a negative relationship between child quantity and quality. However, very few studies have focussed on the relationship between family size and child health. In this study, we present a comprehensive empirical analysis on the effects of changes in family size on children's health outcomes. To perform our analysis, we match mothers' information from the original cohort of the National Longitudinal Survey of Youth (NLSY 1979) to their children's data from NLSY Child and Young Adult surveys (NLSY CYA). Our analysis is restricted to children of ages 2 to 17 years. To the best of our knowledge, this is the first paper to utilize a US-based child sample to study the impact of family size variations on children's health outcomes.

The majority of empirical studies on family size effects tend to focus on children's educational outcomes, labor market consequences, and/or parental investments required to develop children's future cognitive skills (Rosenzweig & Wolpin 1980; Blake 1981; Hanushek 1992; Black, Devereux, & Salvanes 2005; Cáceres-Delpiano 2006; Lee 2008; Li, Zhang, & Zhu 2008; Angrist, Lavy, & Schlosser 2010; De Haan 2010; Ponczek & Souza 2012; Juhn, Rubinstein, & Zuppann 2015). Empirical results in most of these studies support the validity of a quantity-quality trade-off. In comparison, shortage of research and lack of consensus among researchers with respect to the nature of relationship between family size and child health provide an important scope to contribute to the existing child development literature based on quantity-quality trade-off theory (relevant studies include papers by Rosenzweig & Zhang 2009; Millimet & Wang 2009; Lundborg, Ralsmark, & Rooth 2015;

Peters, Rees, & Rey 2015; Zhong 2016). Further, some existing studies do not find any evidence of a trade-off between child quantity and child health outcomes (Lundborg et al 2015; Peters et al. 2015). A non-negative relationship between child quantity and child health is consistent with the 'Hygiene' theory proposed by Strachan (1989). The theory discusses specific long-term child health benefits that are associated with a large family size.

Child body weight is an important health indicator. Excess body weight is closely associated with serious long-term health risks including diabetes, heart disorders, hypertension, and cancer (Must et al. 1992; Power, Lake, & Cole 1997; Leonard et al. 2008; Reilly & Kelly 2011). On the other hand, being underweight during childhood can elevate the risk of mortality and affect children's cognitive development (Walker et al. 2000; De Onis et al. 2004). Economic hardships and food insecurity put children at the risk of having extreme body weight. Based on the quantity-quality trade-off hypothesis, changes in allocation of fixed parental resources, induced by an increase in family size, can potentially affect children's body weight outcomes. We try to use accurate estimates of children's height and weight from NLSY CYA to compute children's body mass index (BMI). We refer to the age and sex-specific body weight standards (BMI-for-age) of Centers for Disease Control's (CDC) growth charts[1] and use children's BMI values to construct our binary health indicators. The binary indicators are created for being- overweight (BMI greater than equal to 85th percentile of age and sex-specific BMI distribution), obese (BMI greater than equal to 95th percentile of age and sex-specific BMI distribution), and underweight (BMI less than equal to 5th percentile of the BMI distribution).

We attempt to estimate causal effects by using exogenous variations in family size that are triggered by twin births and parental preference for a mixed sex composition of their

[1] See https://www.cdc.gov/growthcharts/html_charts/bmiagerev.htm; Retrieved on November, 12, 2016.

children. Regression estimates from our instrumental (IV) regressions do not provide any evidence of a trade-off between child quantity and quality. Our results indicate that increase in family size by an additional member may affect children's well-being through some other channels that can potentially offset the negative impact of resource constraints induced by an additional member. In particular, a large sibling size is likely to be associated with some positive health externalities (the hygiene theory) and can also improve sibling relationships (Newman 1996). Hence, to gather deeper understanding of the health outcomes of changes in family size, we study family size effects on two additional child outcomes that are related to child well-being. In particular, we estimate relationship between family size variations and- a) mothers' rating on her child's relationship with siblings, and b) incidence of child's illness that requires medical attention (to test the validity of the hygiene theory). The analysis provides evidence regarding the presence of positive health externalities in families with a large sibship size. Our cross-sectional regression analysis suggests that increase in family size by an additional young member significantly improves child relationship with his/her siblings.

Further, to address some empirical concerns that are likely to be associated with our cross-sectional analysis, we take advantage of the longitudinal format of the NLSY mother-child data to study how birth of a younger sibling affects older children. Controlling for unobserved child fixed effects, regression analysis suggests that birth of a younger sibling leads to a significant decline in the likelihood of being overweight. We also find that birth of a younger sibling significantly decreases the likelihood of falling ill. In addition, the longitudinal data also allows us to capture changes in a child's health outcomes across successive births of younger siblings. Our results show that birth of a younger sibling at higher parities leads to a significant decline in the probability of being obese, but may also lead to an increase in the likelihood of being underweight. Overall, our panel regression

analysis provides useful evidence with respect to quantity-quality trade-off and hygiene hypotheses.

The paper is organized as follows; in Section 2, we review relevant literature to explore existing theories on the relationship between family size and child health and discuss empirical strategies commonly adopted in the family size literature. In Section 3, we describe the NLSY mother-child data used for our empirical analysis. In Section 4, we explain our IV regression strategy. In section 5, we discuss our findings from cross-sectional regression analysis (pooled and IV regressions). In Section 6, we explain our panel regression models and results. Finally, in section 7, we present our concluding remarks.

2. Literature Review

2.1 Mechanism – Evidence from existing theories

The quantity-quality trade-off theory argues that when there is a budget constraint, an increase in the number of children in a family increases the marginal cost of parental investments that determine child quality (Becker 1960; Becker & Lewis 1973; Becker & Tomes 1976). As child quantity increases, fixed parental resources are allocated across a larger number of children. In the context of our analysis, the theory signals towards a decline in child health quality with an increase in child quantity, especially when child health is determined by inputs that require consumption of limited parental resources (such as market-based inputs, time, care, and affection).[2]

Conversely, there is evidence in health literature that indicates that children in larger families tend to have better health outcomes. The association between family size and child health benefits is in line with the hygiene hypothesis proposed by Strachan (1989). Hygiene

[2] Millimet & Wang (2009) present a simplified version of Becker's (1960) quantity-quality trade-off model, where they assume children's (health) quality as a function of market-based health inputs and an initial health endowment. For further reference, we provide the model in Appendix (see Appendix, A.1).

theory argues that compared to smaller families, children in large families develop stronger immune system as they are more likely to be exposed to various diseases during the early years of their childhood. Consistent with the hygiene theory, a number of studies in child health research have shown that children in larger families are less likely to have excess body weight (Edwards & Grossman 1983; Kruger et al. 2006). Some additional studies find that children in large families are also less prone to suffering from diseases and medical conditions such as diarrhoea, respiratory infection, hay fever, asthma, and allergy (Strachan 1989; 1997; Rona et al. 1997; Ponsonby et al. 1998; Jenson & Ahlburg 2002).

2.2 Family size and child health outcomes

The few empirical studies that have analyzed child health effects of family size have primarily focussed on developing regions (such as Bangladesh, China, and Indonesia). Peters et al. (2015) study family size effects on child health in rural Bangladesh. Using multiple health outcomes (including diarrhoea, respiratory infections, eye infections, and colds), the study does not find any statistical evidence in support of the quantity-quality trade-off argument. On the other hand, using data on Chinese children, Rosenzweig and Zhang (2009) and Zhong (2016) find a significant negative relationship between family size and child health.[3] However, in a cross-sectional survey across middle-school children in eastern China, Hesqueth et al. (2003) show children in one-child families are more likely to be overweight and tend to visit doctors and dentists more frequently due to poor health. Millimet and Wang (2009) employ data from Indonesian Family Life Survey to study family size and child health relationship. The study find mixed results for family size effects on height-for-age and BMI-for-age indicators. In particular, their results are in favor of the quantity-quality trade-off theory only at select quantiles of BMI distribution of the child sample used. Finally,

[3] Both studies utilize China's one-child policy to construct an IV for exogenous variations in family size. Health outcomes in the studies are measured by children's self-assessed health, BMI measures, and height-for-age Z-scores.

Lundborg et al. (2015) study the causal effect of family size on children's height using Swedish data on male population from 1965 to 1978 birth cohorts. Their results indicate that an exogenous increase in family size is positively related to children's height.

Children's body weight has long-term health implications. Economic hardships, especially in developing countries, often restrict children from receiving adequate nutritional inputs, thereby increasing the chances of being underweight (Ravallion, Huppi, & Mundial 1989; Nandy et al. 2005). At the same time, poverty and lack of education are often associated with risks of obesity (Drewnowski & Specter 2004; Meyers, Karp, & Kral 2006; Tanumihardjo et al. 2007; Lee, Harris, & Gordon-Larsen 2009). High incidence of obesity among lower socio-economic groups are likely to be due to food insecurity, unhealthy dietary practices, and higher levels of consumption of high-caloric fast foods and beverages. Exogenous increases in family size may impose financial and resource restrictions on families. Consequently, a rise in family size may elevate the risk of having extreme body weight for young children. In US, National Health and Nutrition Survey's recent estimates indicate that prevalence of childhood obesity (17 percent in 2011-2012) tend to be much higher than prevalence of childhood underweight (3.5 percent in 2011-2012) (Fryar & Ogden 2014; Ogden et al. 2016).

Investigation of causal relationship between family size and child health has important family and child welfare policy implications. In a few studies, economic researchers have evaluated the benefits of family planning interventions on human capital outcomes and parental investments in developing regions where population density is a large socio-economic concern (Hossain 1989, Bangladesh; Rosenzweig & Wolpin 1986, Philippines; Rosenzweig & Zhang 2009, China). Given the institutional and demographic differences across regions, empirical results from our analysis shall provide important evidence on the relevance of family planning policies in the context of the US economy.

2.3 Sources of exogenous variation in family size – Instrumental variables

Family size can be endogenously determined by unobserved parental (or family) characteristics that may also affect child outcomes. Empirical research in recent family size literature has used twin births and parental preference for a mixed sex composition of their children as two sources of exogenous variations in family size to estimate causal effects (a few examples include Angrist, et al. 2010; Black, et al. 2005; De Haan 2010). Rosenzweig and Wolpin (1980) are the first to use twin births as a natural experiment for exogenous increase in family size. Angrist and Evans (1998) considered sex composition of children born earlier in a family as an important determining factor for sibling size. They argued that parents with children of the same sex are more likely to have an additional child in order to have a mixed sex-composition of their children. An important consideration for the validity of their instrumental variable is that children's sex is randomly determined.

It is important to note however, that family size interventions generated by the two IV's differ in nature. Hence, average causal effects of family size are likely to vary across IV regressions (see details in Angrist et al. 2010). For example, Black et al. (2005) find a significant negative relationship between family size and male IQ scores when they use twin birth as an IV, but they do not find any effect when family size is instrumented by an IV based on older siblings' sex composition. In general, multiple sources of exogenous variations allow researchers to check for consistency of their findings and compare family size effects across alternative model specifications.

3. Data

3.1 The NLSY sample

We match mother's data from the original NLSY 79 cohort with their children from the Child and Young Adult Surveys (NLSY CYA). Children in NLSY CYA have been surveyed

biennially from 1986 to 2012. To prepare our sample for empirical analysis, first, we consider families with at least two children (Angrist et al. 2010; Black et al. 2005). Since we are interested in looking at the family size effects on children's health, we restrict our sample to children aged from 2 to 17. Our explanatory variable is the number of members in a child's household who are aged under 18 (non-adult household members). Next, we restrict our regression sample to children who reside with both their parents to study family size effects in regular households.[4]

Further, height and weight information reported by children and their mothers may suffer from reporting errors, which can bias our regression coefficients (Cawley et al. 2015). Overweight people often tend to under-report their weight and underweight people are likely to over-report their weight. In order to minimize reporting errors and construct more accurate health measures, we restrict our samples to objective measurements (scale and tape-measured information) rather than relying on self-reported data.[5] The above corrections limit our data to 7342 observations. We provide descriptive information of the variables used in our regression analysis in Table 1 below.

[4] NLSY CYA includes information on child's usual residence and whether child's father lives in her/ his household. We also perform regressions using an unrestricted child sample that includes all children who reside with their mothers, while their fathers' presence in households may vary (due to various reasons such as mother's marital status or parents' occupational choices). Controlling for mother's marital status in regressions using unrestricted child sample, regressions results are consistent with our main findings.
[5] See Appendix A.4 for the specific NLSY CYA questions that were used to identify the samples.

Table 1

Descriptive information of regression sample from NLSY's (1986-2012)

Variables	Mean/Proportion (SD)
Child characteristics	
Body Mass Index (BMI)	18.083 (5.675)
Underweight	0.114 (0.318)
Overweight	0.277 (0.448)
Obese	0.140 (0.347)
Incidence of child illness	0.339 (0.474)
Sibling relationship	3.291 (0.683)
Household Size (under 18 years)	2.712 (1.100)
Hispanic	0.199 (0.399)
White	0.613 (0.487)
Females	0.491 (0.500)
Age	8.460 (3.366)
Weight at birth (in ounces)	120.085 (23.962)
Mother Characteristics	
Highest grade completed (HGCM)	13.364 (2.641)
AFQT Scores	48466.200 (28772.720)
Rotter Scale	8.632 (2.361)
Body Mass Index	26.666 (6.036)
Age at child birth	28.057 (4.705)
Age	36.544 (4.723)
Family and birth-order characteristics	
Poverty status	0.117 (0.321)
First child	0.323 (0.467)
Second child	0.382 (0.486)
Third child	0.192 (0.394)
Fourth child	0.066 (0.248)
Fifth child +	0.038 (0.191)
IV information	
Family with twins (later births)	0.004 (0.062)
First two children of same sex	0.461 (0.499)
First two children are girls	0.220 (0.414)
First two children are boys	0.242 (0.428)
Number of families (pooled regression)	1540
Number of children (pooled regression)	3052
Pooled regression sample size	7342
Twin IV regression sample size	7057
Sex ratio IV sample size (first child)	2369
Sex ratio IV sample size (first two children)	5177

Note: The mean (standard deviation) of the variables used in regression analyses are based on mother-child sample used in pooled regression analyses. The summary statistics of the variables in our IV regression samples are similar. The IV information in the above table are with respect to respective samples used. The proportions of same-sex children (same sex, both girls, and both boys) are based on child sample of first two children of families with at least two children. The mean value of HGCM represents that on average, mothers have a little over one year of college education. Sibling relationship is a categorical variable coded 1-4 (a rating of 1 represents poor relationship and rating 4 implies excellent relationship).The mean value of the variable (approximately equal to 3) implies that on average, children have very good relationship with their siblings (as reported by mothers).

We compute children's BMI using their height and weight measures (in inches and pounds). We use the estimated BMI to construct our indicators of child health outcomes. We refer to the age and sex-specific BMI-for-age distribution from 2000 CDC Growth Charts for the United States (CDC, 2000) to construct three binary indicators for being overweight, obese, and underweight. The construction of our binary health indicators is similar to the approach employed by Cawley, Moran, and Simon (2010) and Cawley, Frisvold, and Meyerhoefer (2013). In Table 1, we find that the average child BMI is 18. 28 percent of the sample are overweight and 14 percent are obese. Further, 11 percent of the sample are underweight.

In our cross-sectional regression analysis, we control for a number of relevant child, mother, and family-specific characteristics. Child-specific controls include information on birth weight (in pounds), age, sex, race, and ethnicity. Children's average birth weight is close to 8 pounds. While 61 percent of our sample are Whites, 20 percent belong to Hispanic ethnicity.

We include mothers' information on highest grade completed, Armed Forces Qualification Test raw scores (AFQT scores), and scores on Rotter's scale (locus of control measure) to control for mothers' academic achievement, cognitive ability, and non-cognitive characteristics, respectively. On average, mothers in our regression sample have completed at least one year of college education (mean value equal to 13) and tend to have childbirth at the age of 28. Finally, approximately 12 percent of our sample are reported to be living in poverty.

4. Cross-sectional analysis

4.1 IV regression strategy

In our cross-sectional regression analysis, we estimate both pooled and IV regressions

for our health outcomes. The IV regression model (Black et al. 2005) is represented by-

$$Y_i = \beta_0 + \beta_1 \widehat{HHSIZE}_i + X_i\beta_2 + \varepsilon_i \qquad (1)$$

$$HHSIZE_i = \alpha_0 + \alpha_1 Z_i + X_i\alpha_2 + \nu_i \qquad (2)$$

where Y_i represents child i's health outcome (BMI and indicators for being overweight, obese, and underweight). $HHSIZE_i$ is the explanatory variable given by number children (aged under 18) in i's household. X_i is a vector of child, mother, and family controls related to child i. β_1 captures child health effects of changes in household size.

First, we estimate equation (1) and perform ordinary least squares regression (OLS) for BMI (continuous measure) and probit regressions for the three binary health indicators. In equation (2), we describe our first stage of the IV regression analysis, where Z_i is the instrumental variable for household size. In regression analysis using twin births as an instrumental variable, Z_i is a dummy variable that equals 1 for child i belonging to household that currently (at the time of survey) has twins. In regression analyses where we use sibling sex composition as a source of exogenous variation in family size, Z_i is a binary indicator for whether the first two siblings in i's family are of the same sex.

Twin births at later parities (twin IV) and sex composition of the first two siblings (same-sex siblings IV) are more likely to generate exogenous changes in family size for the older children (especially for the first two children) in the family. To ensure precision of our IV regression estimates of family size effects, we incorporate some further modifications in our IV regression samples. In families with twin births, we restrict our sample to children born prior to the twins for our twin IV regression analysis (Juhn et al. 2015). Therefore, the twin IV regression analysis compares older siblings born prior to twins to children in non-twin families.

In models where we use same-sex siblings IV, we perform separate regressions using samples of first-born children and first two children. Causal effects of family size in the sample of first-born children is given by the the difference between health outcomes of the first-born children in- same-sex siblings' families and families where the first two children are of different sex (Angrist et al. 2010; Black et al. 2005; De Haan 2010). Further, to check for consistency in our results, we perform similar analysis for the sample of first two children.

Past studies have shown that it is important to control for birth-order effects in family size research (Hanushek 1992). However, there is ambiguity in the literature whether birth order has a favorable impact on child well-being and development. While some papers argue that later-born children might benefit from parents' child-rearing experiences, others suggest that children born earlier have access to a greater share of parental endowments (Booth & Kee 2009). Following these arguments, we additionally control for children's birth order using dummies for the second child, the third child, the fourth child and for the fifth (or more) child in our regressions. The first child is the excluded category. The standard errors of our regression estimates are corrected for clustering at the child level. In our IV regression analyses, we estimate a two-stage least squares (2-SLS) model for BMI and probit regressions in the second stage for the binary health indicators.

4.2 IV regression samples

In NLSY CYA data, twin births are rare events. In the overall NLSY CYA sample, there are 39 families where twins are born at the second birth parity (the second child and the third child are twins), 21 families where twins are born at the third parity, 11 families with twins at the fourth parity, 7 families with twins at fifth parity, 3 families with twins at the sixth parity, and 1 family with twins at the seventh parity. In our relevant twin IV sample, only 0.4 percent

of the observations are related to twin families (see Table 1). One of the potential limitations of using twin IV is that the limited number of observations may restrict our ability to precisely estimate the family size effects on child health. With respect to same-sex siblings IV regression sample (sample of first two children), in over 46 percent of our sample, the first two children are of the same sex. In 22 percent of the same sample, the first two children are girls (see Table 1).

5. Probit/OLS and IV regression Analysis

5.1 BMI and child body weight indicators

In Table 2 below, we report OLS and IV (2-SLS) regression estimates of the impact of household size[6] on children's BMI. Our OLS regression result suggests that an increase in household size has a strongly significant and negative association with child BMI. More specifically, increase in household size by an additional member is related to a decrease in a child's BMI by 0.2 units (column 1). Further, children's birth weight has a positive and significant association with future BMI. We do not find any significant relationship between household size and BMI when we allow for exogenous changes in household size using twin births and same-sex siblings as IV's. Household size and BMI are negatively related when exogenous variation in family size is triggered by twin births. In regressions using same-sex sibling IV, household size and BMI are positively related in both the samples of first children (column 3) and first two children (column 4).

[6] For our cross-sectional regression analysis, based on the specific NLSY CYA information, we are able to look at the effects of child's household size (rather than family size). Hence, to describe our results in Section 5, we will be using the term 'household size' instead of 'family size'. The effects of family size variations are captured in section 6.

Table 2

Relationship between household size and body mass index (BMI)

	OLS	Twin IV	Sex ratio IV	
			First child	Two children
	(1)	(2)	(3)	(4)
BMI				
Household size (<18 years)	-0.234**	-0.305	0.499	0.793
	(0.081)	(0.747)	(1.936)	(1.367)
Birth weight	0.012***	0.013***	0.014**	0.010**
	(0.003)	(0.004)	(0.006)	(0.004)
2nd child	0.074	0.072		-0.115
	(0.176)	0.242		(0.325)
3rd child	0.437*	0.534		
	(0.252)	0.686		
4th child	0.996*	1.248		
	(0.488)	(1.236)		
5th child +	0.775	0.867		
	(0.484)	(1.847)		
First stage coefficients		1.067***	0.141***	0.125***
		(0.246)	(0.047)	(0.034)
Partial F-statistic first stage		71.88	21.08	26.15
Sample size	7342	7057	2369	5177

*p=0.10 **p=0.05 *** p=0.01

Note: We report linear regression coefficients in the above table. We estimate two-stage least squares models for our IV regression analysis. Standard errors are corrected for clustering at the child level and are provided in parentheses. Child controls include information on birth weight (in ounce), age, sex, race, and birth order. Mother controls include BMI, AFQT scores, Rotter's scale, highest grade completed, and current age. Family control includes family's poverty status. The mother-child matched regression sample includes children who reside with both their parents. In column (3), we perform IV regression analysis (sex ratio) using a sample of first-born children, and in column (4) we use a sample of first two children.

The first stage regression coefficients in Table 2 are strongly significant for both the IV's. Twin birth leads to an increase in household size by more than one child (column 2). The first-stage regression coefficient on twin IV is within the range of estimates found in recent literature. Regression coefficient on twin IV in Juhn et al.'s (2015) study is 1.8 (NLSY CYA data), and De Haan's (2010) first-stage regression coefficient equals 0.9 (Wisconsin Longitudinal Study data). Our first stage regression coefficients on same-sex siblings IV are marginally higher than estimates found in recent studies (coefficients in studies by Angrist et

al. 2010, Black et al. 2005, and De Haan 2010 vary in the range 0.07-0.11). In the sample of first-born children, our point estimate is 0.14 (column 3) and in the sample of first two children, the point estimate equals 0.13 (column 4). However, the 95 percent confidence intervals around our point estimates in the two samples ([0.05, 0.23] for column 3 estimate and [0.06, 0.19] for column 4 estimate) are in the range of coefficient values obtained in previous studies that employ same-sex IV for exogenous changes in family size. Validity of both the IV's is further supported by large F-values in the first stage regressions.

BMI threshold for having a healthy body weight varies by child's age. So next, we provide relatively more specific evidences on child health outcomes of family size variations. In Table 3, we report estimated marginal effects from pooled probit and IV regressions with respect to indicators for being overweight and obese. Probit regression results suggest that increase in household size by an additional member is associated with a decrease in the probability of being overweight and in the probability of being obese by approximately 2 percentage points (column 1). The effects are statistically significant at the conventional levels. However, although household size appears to be negatively related to the likelihood of having excess body weight in the IV regressions (columns 2-4), the marginal effects are not statistically significant. Further, child's birth weight has significant positive association with the likelihood of having excess body weight (in almost all our regression models).

Next, in Table 4, we report marginal effects from the regression analysis using binary indicator of being underweight as child health outcome. Household size does not have any effect on the probability of being underweight (columns 1-4). Further, child's birth weight has a statistically significant negative association with the likelihood of being underweight. Overall, we do not find particularly significant birth order effects on child outcomes in our regression models.

Table 3

Relationship between household size and likelihood of having excess body weight

	Probit	Twin IV	Sex ratio IV	
			First child	Two children
	(1)	(2)	(3)	(4)
Overweight				
Household size (<18 years)	-0.015**	-0.004	-0.172	-0.093
	(0.007)	(0.085)	(0.144)	(0.121)
Birth weight	0.001***	0.001**	0.002**	0.001*
	(0.000)	(0.000)	(0.001)	(0.000)
2nd child	-0.001	-0.004		0.010
	(0.018)	(0.025)		(0.030)
3rd child	0.024	0.015		
	(0.023)	(0.076)		
4th child	0.042	0.041		
	(0.033)	(0.135)		
5th child +	0.049	0.025		
	(0.042)	(0.208)		
Obese				
Household size (<18 years)	-0.017***	-0.046	-0.066	-0.025
	(0.005)	(0.050)	(0.136)	(0.135)
Birth weight	0.001**	0.000*	0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
2nd child	0.003	0.010		0.001
	(0.014)	(0.018)		(0.024)
3rd child	0.034*	0.061		
	(0.019)	(0.046)		
4th child	0.040	0.096		
	(0.025)	(0.081)		
5th child +	0.059	0.136		
	(0.033)	(0.124)		
First stage coefficients		1.067***	0.141***	0.125***
		(0.246)	(0.047)	(0.034)
Partial F-statistic first stage		71.88	21.08	26.15
Sample size	7342	7057	2369	5177

*p=0.10 **p=0.05 *** p=0.01

Note: We report marginal effects from our probit regressions in the above table. In IV regression analysis, we estimate probit regressions in the second stage. Standard errors are corrected for clustering at the child level and are provided in parentheses. Child controls include information on birth weight (in ounce), age, sex, race, ethnicity, and birth order. Mother controls include BMI, AFQT scores, Rotter's scale, highest grade completed, and current age. Family control includes family's poverty status. The mother-child matched regression sample includes children who reside with both their parents. In column (3), we perform IV regression analysis (sex ratio) using a sample of first-born children, and in column (4) we use a sample of first two children.

Table 4

Relationship between household size and likelihood of being underweight

	Probit	Twin IV	Sex ratio IV	
			First child	Two children
	(1)	(2)	(3)	(4)
Underweight				
Household size (<18 years)	0.003	0.037	-0.034	0.008
	(0.005)	(0.061)	(0.112)	(0.108)
Birth weight	-0.001***	-0.001***	-0.001**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
2nd child	-0.006	-0.011		-0.007
	(0.011)	(0.018)		(0.019)
3rd child	-0.000	-0.026		
	(0.014)	(0.054)		
4th child	-0.013	-0.078		
	0.020	(0.098)		
5th child +	-0.022	-0.096		
	(0.027)	(0.150)		
First stage coefficients		1.067***	0.141***	0.125***
		(0.246)	(0.047)	(0.034)
Partial F-statistic first stage		71.88	21.08	26.15
Sample size	7342	7057	2369	5177

*p=0.10 **p=0.05 *** p=0.01

Note: We report marginal effects from our probit regressions in the above table. In IV regression analysis, we estimate probit regressions in the second stage. Standard errors are corrected for clustering at the child level and are provided in parentheses. Child controls include information on birth weight (in ounce), age, sex, race, ethnicity, and birth order. Mother controls include BMI, AFQT scores, Rotter's scale, highest grade completed, and current age. Family control includes family's poverty status. The mother-child matched regression sample includes children who reside with both their parents. In column (3), we perform IV regression analysis (sex ratio) using a sample of first-born children, and in column (4) we use a sample of first two children.

To summarize, our pooled probit regression analysis shows that increase in household size by an additional young member decreases the likelihood of being overweight or obese with no effect on the probability of being underweight. However, controlling for endogeneity in household size, there is no significant relationship between change in number of children in a household and child health outcomes. Further, differences in marginal effects across IV regression models are potentially due to differences in exogenous variations generated by the two IV's (Black et al. 2005) and variations in child samples used in the analyses. The

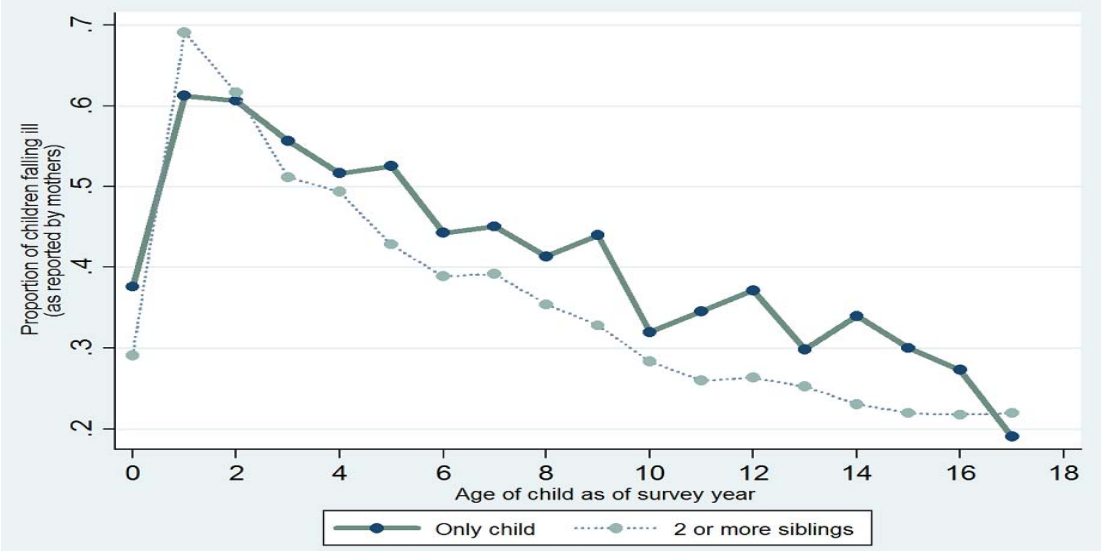
insignificant IV regression results do not provide any empirical evidence in support of the quantity-quality trade-off theory.

5.2 Additional measures of child health-related outcomes

The results discussed in the previous section indicate that certain health benefits of having an additional child member can potentially mitigate the effect of an increase in shadow price of child quality. Next, we consider some additional health-related measures that are likely to be associated with children's health outcomes studied in the previous section. In particular, we draw information on- a) child's relationship with her siblings (rated by mother), and b) whether child had illness that required medical attention in the year prior to survey (reported by mother). The details on the particular NLSY information used in the analysis are reported in Appendix Table A.2. Strong sibling bonds and emotional ties have positive influence on children's well-being (Newman 1996; McHale et al. 2012). In addition, estimating relationship between household size and incidence of child illness shall provide useful evidence into validity of the hygiene theory. First, we utilize our data to see age-specific trends in incidence of child illness across households with different sibling size and present our findings in Figure 1 below.

Figure 1

Age-specific trends in incidence of child illness by sibling size



Note: The above graph compares proportion of first-born children falling ill (that requires medical attention) in the year prior to survey (as reported by mothers) in households with two or more siblings with proportion of children falling ill in the year prior to survey in one-child households.

In Figure 1, we compare children with no siblings with first-born children in households with two or more siblings. We find that in both the household types, children are more likely to fall ill during the early childhood years followed by a declining trend in the incidence of illness. However interestingly, we find that after reaching a maximum, the trend in incidence of illness for children with at least one sibling consistently lies below the line representing trend for children with no sibling, until the age of 17. The observed trends in incidence of illness support the hygiene theory.

Table 5 below presents our regression results for the two additional health-related outcomes. OLS and 2-SLS regressions with respect to sibling relationships show that an addition of a young member in the household leads to a significant improvement in child's relationship with his/her siblings. Consistent with previous family size literature, IV regression estimates are larger than OLS coefficients. We estimate probit regressions for binary indicator for whether child was ill in the period prior to survey in our pooled regression. The marginal effect in column 1 is negative and statistically significant at the 1

percent level. In particular, in pooled regression, increase in household size by an additional member is associated with a decline in the probability of falling ill by 3 percentage points. However, marginal effects from IV regressions indicate that household size does not have any significant effect on the probability of falling ill. Overall, Table 5 results indicate that increase in household size may have some positive influences on children's well-being.

Table 5

Effects of household size on additional child outcomes

	OLS/ Probit	Twin IV	Sex ratio IV	
			First child	Two children
	(1)	(2)	(3)	(4)
Sibling relationship				
Household size (<18 years)	0.047***	0.238*	0.484**	0.343**
	(0.011)	(0.132)	(0.229)	(0.158)
Sample size	6611	6399	2384	4892
Child Illness				
Household size (<18 years)	-0.032***	0.071	-0.033	-0.001
	(0.006)	(0.084)	(0.140)	(0.089)
Sample size	12501	12020	4131	8888

*p=0.10 **p=0.05 *** p=0.01

Note: We report linear regression coefficients (from OLS/ 2-SLS)/ marginal effects (from Probit/ IV Probit) in the above table. Standard errors are corrected for clustering at the child level and are provided in parentheses. Child controls include information on birth weight (in ounce), age, sex, race, ethnicity, and birth order. Mother controls include BMI, AFQT scores, Rotter's scale, highest grade completed, and current age. Family control includes family's poverty status. The mother-child matched regression sample includes children who reside with both their parents. In column (3), we perform IV regression analysis (sex ratio) using a sample of first-born children, and in column (4) we use a sample of first two children.

6. Child Fixed Effects Regression Analysis

Recent family size literature has discussed the possibility that both the IV's may suffer from omitted variable biases. However, performing additional empirical tests, Angrist et al. (2010) and De Haan (2010) do not find evidence against the exclusion restriction assumption of the two IV's. In particular, it is argued that closely spaced younger siblings tend to receive a larger share of parental attention. Hence, twin births may adversely affect older siblings' outcomes (see Angrist et al. 2010; Black et al. 2005). On the other hand, children's sex

composition is likely to be associated with economies of scale (such as sharing of room, clothes, and other resources among same sex siblings) (Rosenzweig & Zhang 2009; Angrist et al. 2010). Hence, before reaching to a specific conclusion regarding the child health effects of family size, we correct for some empirical concerns in our cross-sectional regression analysis by controlling for unobserved time-invariant child-specific characteristics that may potentially bias our results discussed in the previous section.

6.1 Health effects of arrival of younger siblings

Effects of increase in family size in large families may differ from effects of having an additional member in a small family. However, in addition to empirical concerns discussed in the previous section, the cross-sectional regression analysis does not address this potential selection issue. Further, in our IV regression analysis, we use NLSY's information on children's household size as our explanatory variable, which may include members who are not a part of a children's immediate family. Hence, we may not be able to capture the true effect of change in a child's family size in our IV regressions.

To study the family size effects more accurately, we utilize the longitudinal format of the NLSY CYA child samples to perform child fixed effects regression analysis. The fixed effects regressions control for child-specific time-invariant effects that may be correlated with family size and also affect child health outcomes. Our empirical approach is similar to the strategy adopted by Juhn et al. (2015). More specifically, we study how birth of a younger sibling affects older children's health outcomes. The fixed effects model is

$$Y_{ijt} = \gamma_0 + \gamma_1 1.\{after\}_{ijt} + H_{it}\gamma_2 + a_i + u_{ijt} \qquad (3)$$

where Y_{ijt} is the health outcome of child i with a younger sibling j (i's immediate next sibling) at time t, H_{it} is a vector of time-varying controls. a_i represents child fixed effects. The sample for estimation of equation (3) includes children whose mothers gave birth to at least

one younger sibling during the study period. $1.\{after\}_{ijt}$ is an indicator for arrival of a younger sibling *j*. The variable is constructed by matching birth year of younger sibling *j* with NLSY CYA survey years and equals 1 for survey years greater than equal to *j*'s birth year. γ_1 captures the impact of having an additional sibling on children's health outcome.

Table 6 below reports our linear fixed effects regression estimates of impact of birth of a younger sibling on older sibling's body weight outcomes. We find that arrival of a younger sibling does not have any statistically significant effect on child BMI and the likelihood of having extreme body weight (being underweight or obese) (columns 1, 5, and 7). However, having a younger sibling leads to a decline in child's probability of being overweight by 4 percentage points (column 3). This effect is statistically significant at the 10 percent level. Additionally, we estimate short-term and long-term impacts of birth of a younger sibling on children's body weight indicators. To perform our analysis, we estimate a model similar to equation (3) and include two binary indicators to capture a short-term effect (for 0 to 3 years after younger sibling *j* is born in child *i*'s family) and a long-term effect (for 3+ years). Our results in Table 6 suggest that arrival of a younger sibling is negatively related to- child BMI and the likelihood of having excess body weight both in the short-term and in the long-term. However, we find statistically significant effects only for the coefficients related to the overweight indicator (column 4).

Table 6

Effects of birth of a younger sibling on child health using children's fixed effects

	BMI		Overweight		Obese		Underweight	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After birth of younger sibling	-0.343 (0.308)		-0.044* (0.024)		-0.017 (0.020)		-0.001 (0.020)	
Short-term (0-3 years)		-0.338 (0.306)		-0.043* (0.024)		-0.016 (0.019)		0.000 (0.021)
Long-term (3+ years)		-0.357 (0.346)		-0.048* (0.027)		-0.021 (0.022)		-0.003 (0.023)

*p=0.10 **p=0.05 ***p=0.01

Note: Linear child fixed effect regression coefficients are reported in the above table. Standard errors are corrected for clustering at the child level and reported in parentheses. The models control for time-varying characteristics including child age, mother's age, mother's BMI, and family's poverty status. The above regression analyses include 10611 observations. Performing OLS regressions to study the impact of arrival of a younger sibling on child health outcomes, we find that regression estimates are lower than the fixed effects regression estimates for all the outcomes, although the nature of relationship is consistent. OLS regression results are available upon request.

In addition to the above analysis, we estimate equation (3) to study the effects of how birth of a younger sibling affects child's relationship with his/her siblings and the likelihood of falling ill. The linear child fixed effects regression coefficients are presented in Table 7 below. Controlling for unobserved time-invariant child characeristics, regression estimates suggest that while arrival of a younger sibling does not have any effect of child's relationship with his/her siblings, the same leads to a siginficant (at the 1 percent level) decline in the probability of falling ill by 5 percentage points (column 3). For both the outcomes, the long-term effects appear to be larger in absolute terms than the respective short-term effects. The regression results with respect to additional child outcomes differ from our findings in the IV regression analyses. In particular, we find that child relationship with siblings worsens in the long-run upon arrival of a younger sibling (column 2). However, it is important to note the difference between the two analyses. The child fixed effects regressions capture the effect of change in the size of immediate family (sibling size), while our IV regressions analyze the effect of changes in overall household size and may also suffer from omitted variable biases as discussed earlier. Performing OLS with the same set of right-hand side variables (as used

in equation 3) indicate that unobserved child-specific characteristics that determine sibling bond are inversely correlated with family size.

Table 7

Relationship between younger sibling's birth and additional child outcomes using children's fixed effects

	Sibling relationship		Child illness	
	(1)	(2)	(3)	(4)
After birth of younger sibling	-0.084 (0.069)		-0.049*** (0.016)	
Short-term (0-3 years)		-0.070 (0.070)		-0.047*** (0.016)
Long-term (3+ years)		-0.141* (0.074)		-0.054*** (0.018)
Sample size	7554	7554	19597	19597

*p=0.10 **p=0.05 *** p=0.01

Note: Linear child fixed effect regression coefficients are reported in the above table. Standard errors are corrected for clustering at the child level and reported in parentheses. The models control for time-varying characteristics including child age, mother's age, mother's BMI, and family's poverty status.

6.2 Child Health Outcomes across Successive Sibling Births

The NLSY child sample allows us to capture changes in child health outcomes across successive births of younger siblings. In particular, we modify equation (3) by introducing multiple indicators for successive births of younger siblings. The estimated model is represented by-

$$Y_{ijt} = \rho_0 + \sum_{j=i+1}^{N} \lambda_j 1.\{after\}_{ijt} + H_{it}\rho_1 + a_i + e_{ijt} \qquad (4)$$

In equation (4), for each child *i* we capture variations in *i*'s health outcomes generated by births of *i+1* up to the N^{th} child (*N* being the total number of *i*'s younger siblings captured in the study period for each family). For example, for a first-born child, equation (4) will estimate the effects of birth of the second child up to the N^{th} sibling on the first-born child's health outcome. Each $1.\{after\}_{ijt}$ is a 0-1 binary indicator that equals 1 from the time of

arrival of a younger sibling till the next sibling is born. μ_j's are the estimates of the effects of birth of younger siblings. Since equation (4) is a modified version of equation (3), variables Y_{ijt}, H_{it}, and a_i have similar interpretations. We report our estimated linear fixed effects regression coefficients in Table 8 below. Given our sample size, we are able to capture effects of an increase in family size for up to birth of the sixth younger sibling.

Table 8

Effects of birth of additional siblings on older children's health outcomes

	BMI	Overweight	Obese	Underweight
Arrival of younger sibling at	(1)	(2)	(3)	(4)
First parity	-0.326	-0.044*	-0.016	0.001
	(0.308)	(0.024)	(0.019)	(0.020)
Second parity	-0.750	-0.043	-0.043*	0.021
	(0.527)	(0.034)	(0.027)	(0.029)
Third parity	-0.517	-0.008	-0.014	0.029
	(0.512)	(0.050)	(0.036)	(0.043)
Fourth parity	-0.915	-0.038	-0.051	0.009
	(0.769)	(0.070)	(0.075)	(0.054)
Fifth parity	-3.251***	-0.325***	-0.294**	0.070*
	(0.915)	(0.141)	(0.124)	(0.041)
Sixth parity	-0.971	-0.229	-0.175**	0.137***
	(1.441)	(0.318)	(0.089)	(0.042)

*p=0.10 **p=0.05 *** p=0.01

Note: Linear child fixed effect regression coefficients are reported in the above table. Standard errors are corrected for clustering at the child level and reported in parentheses. The models control for time-varying characteristics including child age, mother's age, mother's BMI, and family's poverty status. The results are consistent when we perform regressions using sample of first-born children only. The above regression analyses include 10611 observations.

Regression coefficients in Table 8 indicate that arrival of a younger sibling at higher parities (fifth and sixth parity) leads to a significant decline in the probability of obesity (column 3). Due to limited sample size for siblings born at later birth parities, we observe wide confidence intervals for younger siblings at the fifth and the sixth parity. Arrival of younger siblings at higher birth parities (fifth and sixth parity) is also positively related to the likelihood of being underweight (column 4). The effects are statistically significant at the 10 percent level for younger sibling at the fifth parity and at the 1 percent level at the sixth birth

parity. Our results indicate that in general, increase in family size is negatively related to child's body weight (the BMI results are reported in column 1). We do not find strong evidence of a quantity-quality trade-off when children have younger siblings at lower parities. However, the significant family size effects observed at later birth parities suggest that changes in allocation of parental resources may become more profound in households with large sibship size.

Estimation of regression models by controlling for unobserved child fixed effects allows us to use an alternative empirical approach to check for consistency of our findings in cross-sectional analysis. Further, we gain some deeper understanding of family size and child health relationship by capturing health outcomes across births of younger siblings. A possible threat to identification of effects of increase in family size on child health is that child health outcomes may affect parental decisions to have additional childbirth. In Appendix A.2, we address this empirical concern by performing an event study similar to Juhn et al.'s (2012) study to test the presence of reverse causality. We do not find any evidence with regard to endogenous changes in family size.

7. Conclusion

In general, our analysis suggests that increase in family size does not have any negative relationship with child health quality as predicted by the quantity-quality trade-off model. Our results, with respect to overweight and obesity indicators, are largely consistent with previous findings in the related literature.

Further, we provide additional evidence in this paper to gain understanding of potential mechanisms that can explain a non-negative relationship between child quantity and health quality. Although our cross-sectional regression results differ from panel regression estimates with regard to additional child health outcomes, we do find some evidence that are

indicative of potential health benefits that a child can gain from having additional young members in the family. However, it is important to note that our analysis does provide little evidence of negative health effects of having an additional child in large families. In particular, our fixed effects regression analysis shows that increase in shadow price of child quality induced by arrival of younger sibling at higher parities may be a dominating factor in determining child health.

Based on previous literature related to family size effects, our analysis shows that children's cognitive outcomes and health outcomes are likely to respond differently to changes in family size. Further, parental investments in child-rearing process are not only affected by changes in family size, but it may also vary by parental priorities assigned to different aspects of child development. It is plausible that child health receives higher degrees of parental importance than academic development during the early stages of childhood. Hence, an exogenous increase in family size is likely to have smaller and insignificant impact on child health compared to cognitive outcomes.

Compliance with ethical standards:
We hereby declare that this project was not funded by any public or private entity. We also declare that this study does not involve any conflict of interest.

References

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review* 88 (3): 450–77.

Angrist, Joshua, Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28 (4): 773–824.

Becker, Gary S. 1960. "An Economic Analysis of Fertility." In *Demographic and Economic Change in Developed Countries*, 209–40. Columbia University Press.

Becker, Gary S., and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81 (2, Part 2): S279–88. doi:10.1086/260166.

Becker, Gary S, and Nigel Tomes. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84 (4, Part 2): S143–62.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics* 120 (2): 669–700.

Blake, Judith. 1981. "Family Size and the Quality of Children." *Demography* 18 (4): 421–42.

Booth, Alison L, and Hiau Joo Kee. 2009. "Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment." *Journal of Population Economics* 22 (2): 367–97.

Cáceres-Delpiano, Julio. 2006. "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources* 41 (4): 738–54.

Cawley, John, John Moran, and Kosali Simon. 2010 "The impact of income on the weight of elderly Americans." *Health Economics* 19, no. 8: 979-993.

Cawley, John, David Frisvold, and Chad Meyerhoefer. 2013. "The Impact of Physical Education on Obesity among Elementary School Children." *Journal of Health Economics* 32 (4): 743–55. doi:10.1016/j.jhealeco.2013.04.006.

Cawley, John, Johanna Catherine Maclean, Mette Hammer, and Neil Wintfeld. 2015. "Reporting Error in Weight and Its Implications for Bias in Economic Models." *Economics & Human Biology* 19: 27–44.

De Haan, Monique. 2010. "Birth Order, Family Size and Educational Attainment." *Economics of Education Review* 29 (4): 576–88.

De Onis, Mercedes, Monika Blössner, Elaine Borghi, Edward A. Frongillo, and Richard Morris. 2004 "Estimates of global prevalence of childhood underweight in 1990 and 2015." Jama 291, no. 21: 2600-2606.

Drewnowski, Adam, and SE Specter. 2004. "Poverty and Obesity: The Role of Energy Density and Energy Costs." *The American Journal of Clinical Nutrition* 79 (1): 6–16.

Edwards, Linda N, and Michael Grossman. 1979. "Adolescent Health, Family Background, and Preventive Medical Care."

Fryar, Cheryl D., and Cynthia L. Ogden. 2014 "Prevalence of Underweight Among Children and Adolescents Aged 2–19 Years: United States, 1963–1965 Through 2007–2010 2012.".

Hanushek, Eric A. 1992. "The Trade-off between Child Quantity and Quality." *Journal of Political Economy* 100 (1): 84–117.

Hesketh, T, J D Qu, and A Tomkins. 2003. “Health Effects of Family Size: Cross Sectional Survey in Chinese Adolescents.” *Archives of Disease in Childhood* 88 (6): 467–71. doi:10.1136/adc.88.6.467.

Hossain, Shaikh I. 1989. “Effect of Public Programs on Family Size, Child Education and Health.” *Journal of Development Economics* 30 (1): 145–58.

Jensen, Eric R, and Dennis A Ahlburg. 2002. “Family Size, Unwantedness, and Child Health and Health Care Utilisation in Indonesia.” *Bulletin of Indonesian Economic Studies* 38 (1): 43–59.

Juhn, Chinhui, Yona Rubinstein, and C. Andrew Zuppann. 2015. “The Quantity-Quality Trade-off and the Formation of Cognitive and Non-Cognitive Skills.” *National Bureau of Economic Research Working Paper Series* 21824 (December).

Kruger, R, HS Kruger, and UE Macintyre. 2006. “The Determinants of Overweight and Obesity among 10-to 15-Year-Old Schoolchildren in the North West Province, South Africa–the THUSA BANA (Transition and Health during Urbanisation of South Africans; BANA, Children) Study.” *Public Health Nutrition* 9 (03): 351–58.

Lee, Hedwig, Kathleen Mullan Harris, and Penny Gordon-Larsen. 2009. “Life Course Perspectives on the Links between Poverty and Obesity during the Transition to Young Adulthood.” *Population Research and Policy Review* 28 (4): 505.

Lee, Jungmin. 2008. “Sibling Size and Investment in Children’s Education: An Asian Instrument.” *Journal of Population Economics* 21 (4): 855–75.

Leonard, J, JK Heimbach, M Malinchoc, K Watt, and M Charlton. 2008. “The Impact of Obesity on Long-term Outcomes in Liver Transplant Recipients—Results of the NIDDK Liver Transplant Database.” *American Journal of Transplantation* 8 (3): 667–72.

Li, Hongbin, Junsen Zhang, and Yi Zhu. 2008. “The Quantity-Quality Trade-off of Children in a Developing Country: Identification Using Chinese Twins.” *Demography* 45 (1): 223–43.

Lundborg, Petter, Hilda Ralsmark, and Dan-Olof Rooth. 2015. “The More the Healthier? Health and Family Size.” Mimeo., Lund University.

McHale, Susan M., Kimberly A. Updegraff, and Shawn D. Whiteman. 2012 "Sibling relationships and influences in childhood and adolescence." *Journal of Marriage and Family* 74, no. 5: 913-930.

Meyers, Alan F, Robert J Karp, and John G Kral. 2006. “Poverty, Food Insecurity, and Obesity in Children.” *Pediatrics* 118 (5): 2265a–2266.

Millimet, Daniel L., and Le Wang. 2011. “Is the Quantity-Quality Trade-Off a Trade-Off for All, None, or Some?” *Economic Development and Cultural Change* 60 (1): 155–95. doi:10.1086/661216.

Must, Aviva, Paul F Jacques, Gerard E Dallal, Carl J Bajema, and William H Dietz. 1992. “Long-Term Morbidity and Mortality of Overweight Adolescents: A Follow-up of the Harvard Growth Study of 1922 to 1935.” *New England Journal of Medicine* 327 (19): 1350–55.

Nandy, Shailen, Michelle Irving, David Gordon, SV Subramanian, and George Davey Smith. 2005. “Poverty, Child Undernutrition and Morbidity: New Evidence from India.” *Bulletin of the World Health Organization* 83 (3): 210–16.

Newman, J. 1996 "The more the merrier? Effects of family size and sibling spacing on sibling relationships." *Child: Care, Health and Development* 22, no. 5: 285-302.

Ogden, Cynthia L., Margaret D. Carroll, Hannah G. Lawman, Cheryl D. Fryar, Deanna Kruszon-Moran, Brian K. Kit, and Katherine M. Flegal. 2016 "Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014." Jama 315, no. 21: 2292-2299.

Peters, Christina, Daniel I Rees, and Rey Hernández-Julián. 2014. “The Trade-off between Family Size and Child Health in Rural Bangladesh.” *Eastern Economic Journal* 40 (1): 71–95. doi:10.1057/eej.2012.36.

Ponczek, Vladimir, and Andre Portela Souza. 2012. “New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country.” *Journal of Human Resources* 47 (1): 64–106.

Ponsonby, Anne-Louise, David Couper, Terence Dwyer, and Allan Carmichael. 1998. “Cross Sectional Study of the Relation between Sibling Number and Asthma, Hay Fever, and Eczema.” *Archives of Disease in Childhood* 79 (4): 328–33.

Power, Chris, JK Lake, and Tim J Cole. 1997. “Review: Measurement and Long-Term Health Risks of Child and Adolescent Fatness.” *International Journal of Obesity & Related Metabolic Disorders* 21 (7).

Ravallion, Martin, Monika Huppi, and Banco Mundial. 1989. *Poverty and Undernutrition in Indonesia during the 1980s*. Agriculture and Rural Development Department, World Bank.

Rona, Roberto J, Enric Duran-Tauleria, and Susan Chinn. 1997. “Family Size, Atopic Disorders in Parents, Asthma in Children, and Ethnicity.” *Journal of Allergy and Clinical Immunology* 99 (4): 454–60.

Rosenzweig, Mark R, and Kenneth I Wolpin. 1980. “Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment.” *Econometrica: Journal of the Econometric Society*, 227–40.

———. 1986. "Evaluating the effects of optimally distributed public programs: Child health and family planning interventions." *The American Economic Review* 76, no. 3 (1986): 470-482.

Rosenzweig, Mark R, and Junsen Zhang. 2009. “Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China’s ‘one-Child’ Policy.” *The Review of Economic Studies* 76 (3): 1149–74.

Reilly, John J., and Joanna Kelly. 2011 "Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review." *International journal of obesity* 35, no. 7: 891-898.

Strachan, David P. 1989. “Hay Fever, Hygiene, and Household Size.” *BMJ: British Medical Journal* 299 (6710): 1259.

———. 1997. “Allergy and Family Size: A Riddle Worth Solving.” *Clinical & Experimental Allergy* 27 (3): 235–36.

Tanumihardjo, Sherry A, Cheryl Anderson, Martha Kaufer-Horwitz, Lars Bode, Nancy J Emenaker, Andrea M Haqq, Jessie A Satia, Heidi J Silver, and Diane D Stadler. 2007. “Poverty, Obesity, and Malnutrition: An International Perspective Recognizing the Paradox.” *Journal of the American Dietetic Association* 107 (11): 1966–72.

Walker SP, Grantham-McGregor SM, Powell CA, Chang SM. 2000 “Effects of growth restriction in early childhood on growth, IQ, and cognition at age 11 to 12 years and the benefits of nutritional supplementation and psychosocial stimulation.” *J Pediatr.* 137:36-41.

Willis, Robert J. 1973. “A New Approach to the Economic Theory of Fertility Behavior.” *Journal of Political Economy* 81 (2, Part 2): S14–64.

Zhong, Hai. 2016. “The Effect of Sibling Size on Children’s Health and Education: Is There a Quantity-Quality Trade-Off?” *The Journal of Development Studies*, 1–13.

Appendix

A.1 Quantity-Quality trade-off model of child health

Millimet and Wang (2011) presented a simplified extension of Becker and Tomes's (1976) quantity-quality trade-off model to include health inputs and health endowment as inputs of child quality. Households' objective is to maximize their utility U given by the function $U = U(n, q, c)$, where n represents the quantity of children, q represents children's quality, and c is consumption. Further, q is a function of market-based health inputs w and children's health endowment θ. In particular, child quality is given by the production function $q = q(w, \theta)$. Child quality is positively related to both the inputs w and θ ($q_w > 0$ and $q_\theta > 0$). Millimet and Wang (2011) also include children's sex-ratio in their model assuming that more number children of the same sex can be provide certain cost advantages to households (such as sharing rooms, clothes and other resources). However, to provide a basic understanding of the quantity-quality trade-off, we do not require to account for sex-ratio in the main model.

Households maximize their utility subject to a budget constraint given by:

$$c.p_c + n.p_n + wn.p_w = I \qquad (1a)$$

where I denotes household income, p_c is the unit price of consumption, p_n is the cost per child and p_w represents price of market purchased health inputs. The equilibrium condition is:

$$\frac{\partial U}{\partial c} = \lambda p_c = \lambda \pi_c$$

$$\frac{\partial U}{\partial q} = \lambda \frac{p_w}{\partial q / \partial w} n = \lambda \pi_q \qquad (2a)$$

$$\frac{\partial U}{\partial n} = \lambda (w p_w + p_n) = \lambda \pi_n$$

In the above equations π_c, π_q, and π_n are the shadow prices of consumption, child quality, and child quantity respectively. The equilibrium condition suggests that an

exogenous increase in number of children increases the shadow price of child quality and an increase in child quality increases the shadow price of child quantity. In other words, it is costlier for parents to increase or ensure a high child quality if there are a large number of children in the household and it is costlier to have more children if child quality is high. The interaction between quantity and quality yields a trade-off. However, it is important to note that the quantity-quality trade-off model does not account for positive health externalities that are likely to be associated with increase in family size as discussed by the hygiene theory.

A.2 Event Study – Test for endogenous changes in family Size

Children's health outcomes may affect their parental decision to have additional childbirth. For example, healthy children may provide incentives for parents to have additional children. This may lead to reverse causality concerns that may bias our fixed effects regression coefficients. Hence, using the children's birth date information, we perform an event study to see if there is any evidence of endogenous birth timing (Juhn et al. 2013). We estimate-

$$Y_{ijt} = \theta_0 + \sum_{r=-4}^{-1} \delta_r T_{ijt}^r + \sum_{r=1}^{5} \delta_r T_{ijt}^r + X_{it}\theta_2 + a_i + \mu_{it} \qquad (3a)$$

where T^r_{ijt} is a dummy variable that equals 1 for year t when t is r years relative to the year of birth of the next younger sibling. Given our sample size, for anticipatory effects, we create dummies for up to 4 years prior to a younger sibling's birth year and for post-treatment effects (where childbirth is the treatment), we create dummies for up to 5 years after the younger sibling's birth keeping the birth year as the excluded category. Regression results of our event study are reported in Table A.1. We find no evidence of significant change in older children's health outcomes leading up to a younger sibling's birth. Further, F-values for the overall significance of our leads do not provide any empirical evidence on a reverse causality between child outcomes and parental decisions to have an additional child.

Table A.1
Event study to explore trends in children's body weight before and after birth of an additional sibling birth

	BMI	Underweight	Overweight	Obese
Year relative to younger sibling's birth	(1)	(2)	(3)	(4)
4+ years before	0.755	-0.006	-0.029	0.059
	(0.622)	(0.049)	(0.057)	(0.051)
3 years before	0. 573	0.002	0.010	0.034
	(0.692)	(0.051)	(0.061)	(0.051)
2 years before	-0.255	-0.061*	-0.063	-0.027
	(0.482)	(0.033)	(0.048)	(0.043)
1 year before	-0.070	-0.001	-0.025	-0.020
	(0.611)	(0.045)	(0.052)	(0.041)
1 year after	-0.715	-0.001	-0.043	-0.002
	(0.545)	(0.037)	(0.040)	(0.030)
2 years after	-1.153**	-0.005	-0.071**	-0.017
	(0.566)	(0.031)	(0.032)	(0.024)
3 years after	-1.361**	-0.009	-0.061*	-0.020
	(0.536)	(0.033)	(0.037)	(0.027)
4 years after	-1.617***	-0.014	-0.049	-0.025
	(0.589)	(0.030)	(0.033)	(0.025)
5+ years after	-1.416***	-0.008	-0.071**	-0.022
	(0.534)	(0.029)	(0.035)	(0.026)
F of $\delta_{lead1} + \delta_{lead2} + \delta_{lead3} + \delta_{lead4} = 0$	F=0.24 p=0.63	F= 0.24 p=0.62	F=0.48 p=0.49	F=0.26 p=0.61

*p=0.10 **p=0.05 ***p=0.01

Note: Linear child fixed effect regression coefficients are reported in the above table. Standard errors are corrected for clustering at the child level and reported in parentheses. The models control for time-varying characteristics including child age, mother's age, mother's BMI, and family's poverty status. The sample contains 11131 observations.

Table A.2
Additional NLSY information used in the analysis

NLSY Question	Purpose	Response coding
-Does Father of Child (Living in HH) live in this Household?	Selection of regression sample	1: Yes
-Usual Residence of Child.	Selection of regression sample.	1: 'In household of mother'
-Health: How was weight reported – Scale Measure or Mother Report?	To estimate accurate BMI measures and bodyweight indicators.	1: Scale Measurement
-Health: How was height reported – Tape Measure or Mother Report?	To estimate accurate BMI measures and bodyweight indicators.	1: Tape Measurement
-Health: Has child had any illness that required medical attention? (Mother's report)	For figure 1 and analysis involving additional child outcomes.	0: No 1: Yes
-Mother's rating of child's relationship with siblings.	For analysis involving additional child outcomes.	4: Excellent. 3: Good, 2: Fair, 1: Poor