PRELIMINARY

Family Size and the IQ of Young Men*

by

Sandra E. Black
Department of Economics
UCLA, IZA and NBER
sblack@econ.ucla.edu

Paul J. Devereux
School of Economics and Geary Institute, University College Dublin, CEPR and IZA
devereux@ucd.ie

Kjell G. Salvanes
Department of Economics
Norwegian School of Economics, Statistics Norway, Center for the Economics of Education (CEP) and IZA
kjell.salvanes@nhh.no

April 2007

Abstract

Cognitive skills have been shown to be a strong predictor of educational attainment and future labor market success; as a result, understanding the determinants of cognitive skills can lead to a better understanding of children's long run outcomes. But how do families influence the ability of children? This paper uses a large dataset on the male population of Norway and focuses on one family characteristic: the effect of family size on IQ. Because of the endogeneity of family size, we instrument for family size using twin births and sex composition. IV estimates using sex composition as an instrument show no negative effect of family size; however, IV estimates using twins imply that family size has a negative effect on IQ. Further investigation of the specific features of the twins instrument – the addition of extra children at zero spacing and typically with much lower birth weight than singletons – suggests that the twins estimates are unlikely to generalize to other types of increase in family size. We conclude that there are no important negative effects of *planned* increases in family size on IQ.

* Black and Devereux gratefully acknowledge financial support from the National Science Foundation and the California Center for Population Research. Salvanes thanks the Research Council of Norway for financial support. We are grateful to the Medical Birth Registry for Norway for providing the birth registry data.

Cognitive skills have been shown to be a strong predictor of educational attainment and future labor market success; as a result, understanding the determinants of cognitive skills can lead to a better understanding of children's long run outcomes. But how do families influence the cognitive ability of their children? Researchers across a variety of fields have attempted to understand the determinants of IQ as a measure of cognitive ability with only limited success. This paper uses a large dataset on the population of Norway and focuses on one family characteristic: the effect of family size on IQ.

Family size has long been of interest to researchers. There is an extensive theoretical literature on the tradeoff between child quantity and quality within a family that dates back to Becker (1960) and Becker and Lewis (1973). The theory is often cited and is used as the basis for many macro growth models (see Becker and Barro (1988) and Doepke (2003)). A key element of the quantity-quality model is an interaction between quantity and quality in the budget constraint that leads to rising marginal costs of quality with respect to family size; this generates a tradeoff between quality and quantity.[2] But is this tradeoff real? In particular, is it true that having a larger family has a causal effect on the "quality" (in our case IQ) of the children? Or is it the case that families who choose to have more children are (inherently) different, and the children would have lower IQs regardless of family size?

This paper uses a dataset on the male population of Norway to examine the role of family size on children's IQ, an outcome not previously available in datasets of this size. Importantly, we also address the issue of the endogeneity of family size.[3] Until recent years, the empirical

[2] Rosenzweig and Wolpin (1980) explicitly derive the assumptions under which an exogenous increase in family size should have a negative effect on child quality.
[3] Our data provide information on the IQ scores of men aged between 18 and 20. IQ at these ages is particularly interesting as it is about the time of entry to the labor market or to higher education.

literature on the effects of family size on child outcomes generally relied on OLS estimation and found a negative relationship between family size and child "quality" (usually education), even after controlling for socio-economic factors.[4] However, few of these findings can be interpreted as causal; family size is endogenously chosen by parents and hence may be related to other, unobservable parental characteristics that affect child outcomes.[5] Also, the absence of information on birth order often means that birth order effects are confounded with family size effects. We are unaware of any studies of IQ that have attempted to deal with both these issues.

There is, however, a literature on the causal effects of family size on child educational attainment starting with Rosenzweig and Wolpin (1980). Recent papers in this literature use the sex composition and/or the birth of twins as instruments to provide exogenous variation in family size and have come up with very little evidence for any quantity-quality tradeoff (Conley and Glauber 2006; Caceres-Delpiano 2006; Angrist et al. 2006). In recent work (Black, Devereux and Salvanes 2005a), we use Norwegian data and the twins instrument to examine the effects of family size on education and earnings. We found little evidence for any family size effect either in the OLS or IV estimates when we study cohorts born between 1912 and 1975. In this paper, we can study a more recent set of cohorts -- all individuals in our sample were born between 1967 and 1987 -- and are able to introduce previously unavailable information on IQ and on birth endowments including birth weight.

Given the two types of interventions the instruments represent—one is a planned increase in family size based on parental preferences for variety in the sex composition of their children and the other is an unplanned shock to family size resulting in two generally lower-birth weight children with zero spacing--it is somewhat surprising that researchers have found similar effects

[4] See Blake (1989) and the numerous studies cited therein.

of family size on child outcomes using the twins and sex composition instruments (for example, Angrist et al. 2006). In contrast, our estimates differ depending on estimation method. OLS estimates suggest that there is no strong relationship between family size and IQ. Likewise, IV estimates using sex composition as an instrument show no negative effect of family size. However, our IV estimates using twins imply that family size has a negative effect on IQ; unexpected shocks to family size resulting from twin births have negative effects on the IQ of existing children. Further investigation of the specific features of the twins instrument – the addition of extra children at zero spacing and typically with much lower birth weight than singletons – suggests that the twins estimates are unlikely to generalize to other types of increase in family size. We conclude that there are no important negative effects of *planned* increases in family size on IQ.

Unlike the papers discussed above, we also have information on birth weight for all individuals in our sample. We use this information to address the recent critique of twin studies by Rosenzweig and Zhang (2006), who suggest that much of this literature is biased away from finding negative effects of larger families on child outcomes due to endowment-driven differential parental investments in children. They suggest conditioning on birth weight in a particular way that can be thought of as an attempt to create a new experiment in which twins and singletons have the same average birth weight. We show that their method of conditioning on birth weight potentially invalidates the assumption of exogeneity of twin births. Thus, it can be seen as an attempt to increase external validity that comes at the cost of reduced internal validity i.e. potential bias. Interestingly, while they argue that endowment differences bias estimates in the literature away from finding negative effects of larger families when twin births are used as

---

[5] There is also a literature examining the effect of family size on parental outcomes. See, for example, Bronars and Grogger (1994) and Angrist and Evans (1998).

an instrument, we find that taking account of endowments actually eliminates the negative effects of family size that we find using our basic 2SLS specification.

The paper unfolds as follows. Section 2 describes the data we use and Section 3 discusses our empirical strategy. Section 4 presents our results and Section 5 discusses the external validity of our estimates. Section 6 concludes.

2. Data

Our primary data source on births is the birth records for all Norwegian births over the period 1967 to 1998 obtained from the Medical Birth Registry of Norway. All births, including those born outside of a hospital, are included as long as the gestation period was at least 16 weeks.[6] The birth records contain information on year and month of birth, birth weight, gestational length, age of mother, and a range of variables describing infant health at birth. In these data, we are also able to identify twin births. We can measure family size by counting the number of births to each woman in these data. Our sample is composed of families in which the first birth took place during or after 1967. Given that we observe year of birth, we are able to construct indicators for the birth order of each child directly from the births register.

Using unique personal identifiers, we match these files to the Norwegian Registry Data, a linked administrative dataset that covers the population of Norwegians aged 16-74 in the 1986-2002 period and is a collection of different administrative registers such as the education register, family register, and the tax and earnings register. These data are maintained by Statistics Norway

---

[6] The data also include stillbirths, which constitute approximately 15 per 1000 births. We exclude these from the sample.

[8] Educational attainment is reported by the educational establishment directly to Statistics Norway, thereby minimizing any measurement error due to misreporting. The education register started in 1970; we use information from the 1970 Census for individuals who completed their education before then. Census data are self reported but the information is considered to be very accurate; there are no spikes or changes in the education data from the early to the later cohorts. See Møen, Salvanes and Sørensen [2003] for a description of these data.

and provide information about educational attainment, labor market status, earnings, and a set of demographic variables (age, gender) as well as information on families.[8]

The IQ data are taken from the Norwegian military records from 1984 to 2005. In Norway, military service is compulsory for every able young man. Before entering the service, their medical and psychological suitability is assessed; this occurs for the great majority between their eighteenth and twentieth birthday. The IQ measure is a composite score from three speeded IQ tests -- arithmetic, word similarities, and figures (see Sundet et al. [2004, 2005] and Thrane [1977] for details). The arithmetic test is quite similar to the arithmetic test in the Wechsler Adult Intelligence Scale (WAIS) [Sundet et al. 2005; Cronbach 1964], the word test is similar to the vocabulary test in WAIS, and the figures test is similar to the Raven Progressive Matrix test [Cronbach 1964]. The composite IQ test score is an unweighted mean of the three subtests. The IQ score is reported in stanine (Standard Nine) units, a method of standardizing raw scores into a nine point standard scale with a normal distribution, a mean of 5, and a standard deviation of 2.[9]

Of the men in the 1967-1987 cohorts, 1.2 percent died before 1 year and 0.9 percent died between 1 year of age and registering with the military at about age 18. About 1 percent of the sample of eligible men had emigrated before age 18, and 1.4 percent of the men were exempted because they were permanently disabled. An additional 6.2 percent are missing for a variety of reasons including foreign citizenship and missing observations.[10] There are also some missing IQ scores for individuals who showed up to the military. In total, we have IQ scores on about 84% of the relevant population. One nice feature of our data is that we can construct family size and both instrumental variables for all persons, including those with missing IQ scores. We have regressed a dummy variable for whether ability is missing on family size using both OLS and

---

[9] The correlation between this IQ measure and the WAIS IQ has been found to be .73 (Sundet et al., 1988).
[10] See Eide et al. [2005] for more details.

2SLS. Our finding is that individuals from larger families are more likely to be missing scores (for example, first-borns from 2-child families are 4% less likely to have missing scores than first-borns from 5-child families), but that this relationship disappears when family size is instrumented with the twin or sex composition instruments.[11] Thus it is unlikely that missing IQ data is causing serious bias to our later estimates.

We exclude families in which no child is born before 1987 as no data on military outcomes of children is likely to be available for such families. We use twins to construct our instrumental variables but, as is standard in the literature, we drop twins from our estimating samples. Table 1 presents summary statistics for our sample and Appendix Table 1 shows the distribution of family sizes in our sample.

3. Empirical Strategy

Using Twins as an Instrument for Family Size

Rosenzweig and Wolpin (1980) first discuss the idea of using twin births as unplanned and therefore exogenous variation in family size. In their model, parents have an optimal number of children N*. The birth of twins can vary the actual N from the desired N* and it is this arguably exogenous variation that is used to estimate the effects of family size on child outcomes. Our general estimation strategy is as follows:

$$IQ = \beta_0 + \beta_1 FAMSIZE + X\beta_2 + \varepsilon \qquad (1)$$

$$FAMSIZE = \alpha_0 + \alpha_1 TWIN + X\alpha_2 + \upsilon \qquad (2)$$

In this case, IQ is the IQ of the child and FAMSIZE is the total number of children in the family. Equation (2) represents the first stage of the two stage least squares estimation, where equation

[11] The 2SLS estimates are all small and statistically insignificant and vary in sign.

(1) is the second stage. The controls in X include indicator variables for cohort, age at test, age of mother at birth, age of mother at first birth, years of education of mother, years of education of father, and log birth weight of the child. We also include controls for the IQ of the father for cases in which this is available.[12]

The TWIN indicator is equal to 1 if the n^{th} birth is a multiple birth, and equal to 0 if the n^{th} birth is a singleton. We restrict the sample to families with at least n births and study the outcomes of children born before the nth birth. In practice, we estimate the specification for values of n between 2 and 4. By restricting the sample to families with at least n births, we make sure that, on average, preferences over family size are the same in the families with twins at the n^{th} birth and those with singleton births. Also, by restricting the sample to children born before birth n, we avoid selection problems that arise because families who choose to have another child after a twin birth may differ from families who choose to have another child after a singleton birth.

Our methodology avoids two possible pitfalls that can arise when twins are used as instruments. First, by focusing on whether a twin birth occurs for the n^{th} birth, we avoid the problem that families that have more births are more likely to have at least one twin birth. Second, by only using children born before a possible twin birth, we are able to avoid the problem that a twin birth both increases family size and shifts downwards the birth order of children born after the twins. For example, if a twin birth occurs at the second birth, the next child born is now the fourth born-child rather than the third. Thus, any estimates using children born subsequent to a twin birth will confound family size effects with birth order effects.

[12] We do this by setting father's IQ to zero for cases in which it is missing and then interacting father's IQ with a dummy variable which is one if we have information on father's IQ and zero otherwise. Both these variables are then included in the regression. Father's IQ is typically unavailable for men whose father was born before 1950.

The first stage coefficients using the twin instruments are presented in Table 4. The first stage is very strong and suggests that a twin birth increases completed family size by about 0.7 to 0.9. These first stage effects are a little larger than we found in our 2005 paper, presumably because family sizes are smaller for these recent cohorts.[13] As expected, twins at higher parity have a larger effect on family size, presumably because they are more likely to push families above N* (the optimal number of children). The t statistics from the first stage are typically around 40, indicating that there are no concerns about weak instruments in this application.

Sex Composition

It has been documented that parents have preferences for variety and so are more likely to have a child at parity n if the previous n-1 children are all of the same sex (Angrist and Evans, 1998). As a result, if the sex composition is random, families that randomly had two same-sex children are more likely to have a third, all other things equal, making sex composition a valid instrument. We utilize this instrument for n=3 and n=4, and restrict the sample to families with at least n-1 children.

<u>Sample of families with at least 2 children (n=3)</u>

When n=3, we are studying families with at least two children and the instrument is whether or not the first two children are the same sex. Given our sample is restricted to boys, the relevant possibilities for the first two children are BB, BG, and GB (where G is for girl and B is for boy). One obvious comparison is to compare the IQ of the first born children in a sample composed of BB and BG families – here the instrument can be interpreted simply as whether the second child is a boy. A second comparison is to compare the IQ of second born children in a

[13] The first stage coefficients are much bigger than those of Angrist et al. (2006) using Israeli data. For twins at second birth, they found a first stage coefficient of only 0.44. This presumably reflects the much larger proportion of

sample composed of BB and GB families – here the instrument can be interpreted as whether the first child is a boy. This approach is invalid if there are preferences for boys or girls in addition to preferences for variety, as then the probability of a second child (and hence selection into the at least 2 children sample) will depend on the sex of the first child. The evidence suggests no such sex preferences in Norway so we improve the precision of our estimates by using both comparisons in estimation.[14] To do this, we estimate the following equation using the first two children:

$$FAMSIZE = \alpha_0 + \alpha_1 ALLBOYS + \alpha_2 2nd + X\alpha_3 + \upsilon \qquad (3)$$

Here $ALLBOYS$ is an indicator for the first two children being the same-sex, and $2nd$ is an indicator for whether the child is the 2nd born. The controls included in X are the same ones included when the twins instruments are used.

The first stage estimate on $ALLBOYS$, reported in Table 4, is 0.081 with a standard error of 0.003. Similarly, Angrist et al. (2006) find a coefficient of .073 (.010) using Israeli data and Conley and Glauber find that having two same-sex children increases the probability of having a third child by about .07 in the U.S.

Sample of families with at least 3 children (n=4)

Here the only relevant sequences are BBB, BBG, and GGB; no other comparisons are useful because sequences starting with BG or GB are not comparable to those starting with GG or BB. This is because the probability of having a third child differs systematically depending on whether or not the first two children are the same sex (as is exploited in the n=3 analysis).

---

large families in the Israeli data.

[14] There is no statistically significant relationship between the probability of having a second child and the sex of the first born child despite the enormous sample size of 481,731 families.

[16] Twin probabilities increase with parity and maternal age at birth (Jacobsen, Pearce, and Rosenbloom, 1999; Bronars and Grogger, 1994) but this is not a problem as we condition on these variables in the 2SLS estimation.

Analogously to the n=3 case, a simple approach is to compare the outcomes of the first two children in BBB families to those in BBG families. Also, given there are no preferences for boys (or girls), an additional comparison is between the outcomes of third-born children in BBB and GGB families. In practice we implement both using families with at least 3 children whose first two children are the same-sex. The following regression is estimated for the first three children:

$$FAMSIZE = \alpha_0 + \alpha_1 ALLBOYS + \alpha_2 2nd + \alpha_3 3rd + X\alpha_4 + \upsilon \qquad (4)$$

Here $ALLBOYS$ is an indicator for the first three children being the same-sex, $2nd$ is an indicator for whether the child is the 2nd born, and $3rd$ is an indicator for whether the child is third born.

The first stage estimate on $ALLBOYS$, reported in Table 4, is 0.070 with a standard error of 0.005. While not directly comparable, this is smaller than the Angrist et al. (2006) finding of .120 (.015) using the sample of families with 3 or more children in Israeli data.

Validity of Instruments

Randomness

In order for our IV estimates to be consistent, it must be that the instrument is uncorrelated with the error term in equation (1). One concern is that the occurrence of a twin birth may not be random and may be related to unobservable family background characteristics. By definition, this is untestable, but we do examine whether the probability of twins is related to observed characteristics such as mother's and father's education and father's IQ. To do so, we estimate a linear probability model of the probability of a twin birth at the Nth birth (for N=2, 3, and 4), conditional on having at least N births, using the full set of control variables. Over these three regressions, of the nine coefficients on mother's education, father's education, and father's

IQ, none is statistically significant.[16] Given the enormous sample sizes, these results strongly suggest that twinning probabilities are not related to parental characteristics.[17]

We use a similar approach to examine the randomness of the sex composition instruments. Once again, we find no statistically significant relationship between any of the family background characteristics and the instrument for either the sample of families with at least 2 children, or the sample of families with at least three children where the first two children are the same-sex.

Other Threats to Instrument Validity

The large first stage coefficients for the twins instrument (.7 - .9) are reassuring, as they imply that this can be thought of as an experiment with almost perfect compliance -- the birth of twins leads to family size increasing by one in almost all cases. This is important, as the major possible biases using twins is that we may be capturing the effects of having closely-spaced children, or of having low birth weight children, rather than the effects of having more children. Given that twin births add extra children for almost all families, these possibilities are best seen as threats to the external validity of the twins approach rather than as sources of bias. Later in the paper, we analyze the external validity issues in some depth.

Because the first stage coefficients are much lower for the sex composition instruments, the possibility of bias due to direct effects of sex composition on child IQ is greater. Rosenzweig and Wolpin (2000) conjecture that there may be beneficial effects of having a sibling of the same sex and this would tend to bias away from finding family size effects using the instrument. The literature is mixed on whether sex composition has direct effects on child outcomes (Dahl and Moretti, 2004, Butcher and Case, 1994; Conley, 2000; Deschenes, 2002 all find some evidence of

[17] In recent years, fertility treatments have led to an increase in the incidence of multiple births. This phenomenon did not take off in Norway until the late 1980s so the probability of a multiple birth is fairly constant over time for

sex-composition effects. Kaestner (1997) and Hauser and Kuo (1998) find no evidence of such effects). Such effects would imply that sex composition may not be a valid instrument for family size. In our data, we find no evidence for sex composition effects on IQ when we study situations in which the sex composition has no correlation with family size (for example whether the 3rd child is a boy in families in which the first child is a boy and the second is a girl).

While it may be reasonable to view both type of instruments as being valid, they constitute very different interventions to increase family size. In the next section, we compare the estimates from the two approaches with each other and with OLS and discuss the external validity of the IV approaches.

4. Family Size Results for IQ

In Table 2 and Figure 1, we show the mean IQ and the distribution of IQ scores by family size. There are two clear findings. First, only children have lower scores than the average child in 2 child families. Second, from family sizes of 2 to 5, we see a monotonic relationship that greater family size accompanies lower average scores.

The unconditional relationship between family size and IQ is only suggestive; for example, it could simply represent cohort effects or birth order effects.[18] To better understand the relationship, we regress IQ of children on family size dummies and carry out separate regressions by birth order. We show estimates from regressions with and without additional controls. As described in section 3, these controls include indicator variables for birth order, cohort, age at test, age of mother at birth, age of mother at first birth, years of education of mother, years of education of father, father's IQ, and log birth weight of the child.

---

the cohorts we are studying (persons born 1967 – 1987). See Figure 1 in Black, Devereux, Salvanes (2005b).

The results are presented in Table 3. The omitted category in each case is Family with 5 or more children. These estimates suggest that the highest achieving children come from 2 and 3 child families, with lower achievement for only children and children from families with four or more. However, the results are not robust to introducing the control variables. With the control variables, almost nothing is statistically significant and estimates are much smaller than before. There still appears to be a penalty to being an only child, but there is little indication that family size matters otherwise. The fact that adding family background variables and log(birth weight) as controls largely eliminates the family size effect suggests that, if unobserved factors impacting IQ have similar correlations with family size as the observed factors, the true causal effect of family size is not very negative. The IV estimates below take a more direct approach to the issue of causality.

Results Using Twins as an Instrument for Family Size

The 2SLS estimates are presented in Table 4. We also include two sets of OLS coefficients estimated on exactly the same samples as the 2SLS estimates; the first set includes controls only for birth order, the second set includes all control variables. As was the case in Table 3, the OLS estimates without controls suggest negative effects of increased family size, but those with controls never provide any evidence for negative family size effects. Instead they provide very precisely estimated family size coefficients that are very close to zero.

The 2SLS estimate of the effect on the first child of changes in family size induced by the second birth being a twin birth is -0.15 (0.05). This implies an adverse effect of increased family size on IQ – each additional child reduces IQ of the first child by about 15% of a stanine or 8% of

---

[18] Birth order is highly correlated with family size and has been shown to be strongly related to child outcomes in Norway (Black, Devereux, Salvanes 2005a).

a standard deviation. The equivalent estimate for families that have at least 3 births is -0.17 (0.05), and for families that have at least 4 births it is -0.12 (0.08). Taken together, these three estimates are all more negative than the OLS estimates (the first two are statistically different from OLS) and suggest fairly sizeable negative effects of family size on IQ of about 15 percent of a stanine for each additional child. This translates into approximately a 1% difference in annual earnings as an adult.[19] Unexpected random shocks to family size, induced by the birth of twins, lead to poorer IQ outcomes for existing children.[20]

Results Using Same-Sex as an Instrument for Family Size

In sharp contrast to the estimates using twins as an instrument for family size, the 2SLS estimates using the sex composition as instruments are both positive. For the sample of families with 2 or more births the estimate is .07 (.07), and for the sample of families with 3 or more births the estimate is .08 (.17). While these family size coefficients are not very precisely estimated, the first one does rule out large negative effects of family size on IQ and is statistically significantly different from the family size estimates using twins. The contrast with the twins estimates may result from the fact that, in this case, induced increases in family size are expected and result in normally spaced singletons.

Overall, our OLS estimates provide no evidence of negative effects of increasing family size on IQ, twin IV estimates suggest negative effect of family size on IQ for early-born children,

---

[19] We arrive at this number by regressing log earnings in 2002 on IQ scores for individuals aged at least 30 in 2002. For both cross-sectional and family fixed effects specifications we obtain precisely estimated coefficients on IQ of about .06, suggesting an extra stanine in IQ (about half a standard deviation) leads to about 6% higher earnings.

[20] These results are rather different from the tiny statistically insignificant IV family size effects we found for education in our 2005 paper. The differences appear largely due to our use of different cohorts. When we do the education analysis on a sample as close as possible to our current sample (individuals aged at least 21 in 2002) and use twins as our instrument, we find sizeable negative IV effects of family size on the probability of high school graduation (see Appendix Table 2). The differences between cohorts may relate to the very different environment concerning contraception, abortion, and childcare for later cohorts relative to earlier ones.

and IV estimates using the same-sex instrument suggest no negative effect of larger families on IQ. We next explore why these different estimation methods provide different answers, with particular emphasis on the external validity of the results using twins.

5. External Validity of the Twin IV Estimates

Angrist et al. (2006) discuss the many differences between the twins and sex composition instruments, in particular focusing on how the two instruments affect different parts of the family size distribution and how the ethnic composition of individuals whose fertility responds to the twins instrument differs from that of individuals who respond to the same-sex instrument. In our case these are not issues, as there is no significant ethnic variability in our sample, and the two instruments impact the family size distribution in similar ways. Therefore, we concentrate on two different aspects of the twins instrument that do differ from the same-sex one – birth spacing and birth endowments.

The twins experiment increases family size in a particular way – adding an extra birth unexpectedly, with zero spacing between the children, and adding children that are lighter at birth than average (the difference in birth weight between twins and singletons averages about 1000 grams). For these reasons, the twins IV estimates may not be representative of the effects of other types of increases in family size. We investigate this issue below.

Lower Birth Weight of Twins

Recent twins studies (including this one) have estimated the effect of an extra birth at parity N (due to the birth of twins) on the outcomes of children at parities 1 to N-1, for all women who have at least N births. Rosenzweig and Zhang (2006, henceforth RZ) criticize this approach by arguing that the endowments of twins are lower than those of singletons, and parents may allocate resources differentially across their children on the basis of their endowments.[21] Thus, the effect of adding two twins to a family may differ from the effects of adding two singletons.

RZ suggest an alternative methodology that uses information on both twins and singletons. Within this methodology, they use two different approaches – twins at first birth and twins at second birth. The twins-at-first-birth approach involves restricting the sample to first born children, whether singletons or twins, and to use the following reduced form specification:

$$IQ_{ij} = \eta T_j + \delta e_{ij} + \lambda a_{ij} + \varepsilon_{ij} \qquad (5)$$

Here, T is a twins indicator, e is the birth-weight of the child, and a is the age of the mother at birth. If (5) is estimated without a control for birth weight (e), then η is probably biased negatively because twins have poorer endowments and hence are likely to have lower IQs. Adding a control for birth weight should reduce this bias, and they argue this provides an upper bound (in absolute terms) of the negative effects of being a twin (and hence of family size).

The twins-at-second-birth approach is to estimate equation (5) separately for (a) singleton first-borns, and (b) second-borns. The sample is restricted to families with at least 2 births. T now refers to whether the second birth is a twin birth, and e now refers to the average birth weight of second born children. RZ argue that η is biased positively for first-borns because they benefit from re-enforcing parental investments.[22] They claim that controlling for the birth weight of second-borns reduces the size of the positive bias and provides a lower bound (in absolute terms)

[21] RZ argue that parental investment is likely to reinforce endowment differences and thus bias findings in the literature away from finding negative effects of larger families. They provide some evidence of reinforcing parental investments using twins data from China. However, in the U.S., much of the available evidence suggests compensatory investment behavior (Behrman, Pollak, and Taubman 1989; Ashenfelter and Rouse, 1998; Ermish and Francesconi, 2000). Datar et al. [2005] find evidence that is generally suggestive of reinforcing parental investments among singletons using U.S. data.

of the negative family size effect. On the other hand, after conditioning on birth weight, they argue that the estimate for second-borns provides an upper bound of the negative family size effect for the same reasons as in the twins-first case. RZ provide estimates that are generally consistent with these arguments using data from China.

The problem with the RZ methodology is that, while twinning may be a random event that is orthogonal to family characteristics, birth weight is not. Therefore, conditioning on birth weight creates a correlation between having twins and unobserved family characteristics. For example, families that have twins may be very similar to families that have singletons. But a family that has a 2000 gram singleton is likely to be different in some ways than a family that has two 2000 gram twins (for example, the mother may be more likely to have smoked during pregnancy). Thus, conditioning on birth weight will tend to bias the family size effects in a positive direction. Given this bias, one cannot be sure that any estimate provides an upper bound of the family size effect.[23]

Nonetheless, we report estimates using the RZ methodology on our data in Table 5. Note that, for this section, we include twins in our estimating samples. In the twins-at-first-birth analysis, we see that, on average, twins have lower IQ scores than singleton first-borns. When one controls for birth weight, each additional kilogram is associated with quarter of a stanine increase in IQ, and the twins dummy becomes positive and statistically significant. Like RZ, we implement the twins-at-second-birth strategy in one equation by interacting second-born twins with a first-born indicator. Twins at second birth have a negative effect on both first-born and

---

[22] We don't consider these parental investments to be a source of bias. Any mechanism that increases family size may lead to a reallocation of parental resources amongst children. So, we consider this an issue of external rather than internal validity for the twins instrument.

[23] Recent papers have found that the impacts of birth weight on adult outcomes are generally similar in magnitude in the cross-section to when twin or sibling differences are used in estimation [Black et al. 2007, Oreopoulos et al. 2006]. To the extent that twin and sibling differences difference out confounding effects of other factors, this is

second born children. However, once one controls for the mean birth weight of second-borns, the twins effect now becomes positive. Thus, both specifications suggest that adding a child of the same birth weight has a positive effect on IQ of the first child.

The RZ approach can be thought of as an attempt to create a new experiment in which twins and singletons have the same average birth weight. Unfortunately, in the real world no such experiment is available to us and their method of conditioning on birth weight may invalidate the exogeneity of twin births. Thus, it can be seen as an attempt to increase external validity that comes at the cost of reduced internal validity i.e. potential bias. However, given that such bias may be small, the fact that the effect of a twin birth becomes positive when one controls for birth weight provides another reason to doubt the generalizability of the twin IV estimates.[24]

Birth Spacing

In the case in which we use twins at second birth to study the outcomes of the first child, it may be that the first child is worse off if the second birth is a twin birth because he gets less attention if his siblings are more closely spaced. To examine this, we have looked at non-twin families and estimate the effect of the spacing between later children on earlier children's outcomes. In particular, to simulate the timing of twins, we create a variable for whether the gap between the later children is less than or equal to one year. Also included are the standard set of control variables augmented by a control for family size and a quadratic function of the time gap between children born before the two children whose spacing is being investigated. It is

consistent with the notion that birth weight is not highly correlated with these unobserved factors and hence the bias from the RZ approach may not be large.

[24] Interestingly, while RZ argue that endowment differences bias estimates in the literature away from finding negative effects of larger families, we find that taking account of endowments moves our twins-based family size estimates from negative and statistically significant to positive and statistically insignificant.

important to note that spacing is not exogenous, so these results should be interpreted with caution.

The estimates are reported in Appendix Table 3. The reported coefficients are those relating to the distance in years between the two children immediately after the children listed in column 1. For example, in the first box, the coefficient of -.26 suggests that first child IQ is lower by .26 if the gap between the second and third child is less than or equal to one year (in families with at least 3 children).

We find that earlier children have higher IQs when later children are not more closely spaced together -- children do better if the two children immediately after them are spaced further apart. If this result can be extrapolated to the case of twins, in which the space is zero, it implies that the effect of a subsequent case of twins is more adverse than the effects of having two subsequent singletons.[25] The reduced form effect of twins on IQ of earlier born children in Table 4 is about -.1 for both twins at second and at third birth. The analogous magnitudes in Appendix Table 3 are about -.2, suggesting that the 2SLS estimates using twins may not be generalizable to a situation in which one adds two singletons who are spaced apart.[26]

6. Conclusions

Understanding the determinants of a child's cognitive ability is the first step to understanding their longer run outcomes. In this paper, we examine the role of family size on an individual's IQ, attempting to isolate the causal relationship between the two. Our results point

[25] One caveat here is that, unlike with twins, the spacing of singletons can be chosen by parents and families that closely space births are different in terms of observables than families that space out births.
[26] A closely related issue is that, in the sample of families with 2 or more children, generally first-borns are older when a singleton 3rd child is born (7.9 years old on average) than when the 3rd-born child is part of a twin second birth (4.4 years old on average). It is plausible that increased family size has a more negative impact if it occurs earlier in the child's life and this could cause the twins 2SLS estimates to be more negative than OLS or the sex

to an impact of family size that depends on exactly how family size is increased. OLS estimates with controls for family background suggest that there is no strong relationship between family size and IQ. Likewise, IV estimates using the same-sex instrument also suggest no negative effect of larger families on IQ. In contrast, our IV estimates using twins imply that family size has a negative effect on IQ. However, our investigation of the effects of the zero spacing of twins and their low birth weights suggests that these estimates should not be generalized to other types of variation in family size. Therefore, our conclusion is that there is no important negative effect on IQ of *planned* increases in family size that come through normally spaced singleton births; however, unexpected shocks to family size through twinning have negative effects on IQ.

There are many possible reasons why planned increases in family size may not have deleterious consequences on the IQs of existing children.[27] Younger siblings may actually play a beneficial role for children. For example, Zajonc (1976) suggests that older children develop intellectually through having younger children to teach. Also, parents may respond to the extra expense of an extra child by reducing expenditures on items that are not beneficial to child development such as alcohol or nights out. While we cannot test these hypotheses, there are additional possibilities that we can shed some light on using our data. First, families with more children may be less likely to suffer from marital breakdown. We have examined this by estimating the effect of family size on the probability that both parents are living together 20 years after the first child is born (69% are). The only statistically significant IV estimate is -.032

composition 2SLS estimates. However, it seems unlikely that this factor could account for the big differences between our twins and sex composition estimates.

[27] One possible reason is that scores on our IQ test are not very sensitive to parental investments and other experiences after the first few years of life. This seems unlikely as it has been clearly demonstrated that scores on the AFQT test in the U.S. depend to a large extent on years of schooling of the test-taker (for example, by Neal and Johnson, 2006). Also, this explanation is inconsistent with the large negative effects we find using the twins instrument.

(.011) using twins at 3rd birth.[28] This provides some tentative evidence that twins may destabilize families and this may be another reason for the large negative effects of twins on child IQ. Second, parents may be more likely to increase their incomes as a result of having more children. We have found no evidence for this using either instrument when studying earnings of fathers 15 years after the birth of the first child. Finally, parents may be more likely to leave the labor market and stay home when family sizes are larger and this may be beneficial to the children. Consistent with earlier empirical work for the U.S. (Angrist and Evans, 1998), we find negative effects of family size on labor supply of mothers using all methodologies. However, there is no evidence of any effects for fathers.

[28] Estimates for twins at 2nd birth and at 4th birth are both also negative but statistically insignificant. Caceres-Delpiano (2006) similarly finds evidence that twins destabilize marriage using U.S. Census data.

References

Ashenfelter, Orley, and Cecilia Rouse, "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins," *Quarterly Journal of Economics*, 113, 1998, 253-284.

Angrist, Joshua D. and William N. Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review*, Vol. 88, No. 3, June, 1998, 450-477.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser, "Multiple Experiments for the Causal Link between the Quantity and Quality of Children", MIT Working Paper06-26. 2006.

Becker, Gary S., "An Economic Analysis of Fertility" In Becker, ed., *Demographic and Economic Change in Developed Countries*. Princeton, N.J.: Princeton University Press 1960.

Becker, Gary S. and Robert J. Barro. "A Reformulation of the Economic Theory of Fertility." *Quarterly Journal of Economics*, v103, n1, pp 1-25, 1988.

Becker, Gary S. and H. Gregg Lewis "On the Interaction Between the Quantity and Quality of Children" *Journal of Political Economy*, 1973, Vol. 81, No. 2, S279-S288.

Behrman, Jere, Robert Pollak, and Paul Taubman, "Family Resources, Family Size, and Access to Financing for College Education," *Journal of Political Economy*, 97(2), 1989, 398-419.

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

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2005b), "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes" NBER Working Paper #11796 November.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes, "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes" *Quarterly Journal of Economics*, February 2007.

Blake, Judith, *Family Size and Achievement*, 1989, University of California Press, Berkeley and Los Angeles.

Bronars, Stephen G. and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment." *American Economic Review*, Vol 84, No. 5. December, 1994. 1141-1156.

Butcher, Kristin F., and Anne Case (1994), "The Effect of Sibling Sex Composition on Women's Education and Earnings", *Quarterly Journal of Economics*, Vol. CIX, Issue 3, 531-563.

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

Conley, Dalton, and Rebecca Glauber (2006) "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *Journal of Human Resources*, 41(4).

Conley, Dalton, "Sibling Sex Composition: Effects on Educational Attainment" (2000), *Social Science Research*, Vol. 29, 441-457.

Cronbach, Lee J., *Essentials of Psychological Testing, 2nd Edition*, (London, UK: Harper and Row, 1964).

Dahl, Gordon and Enrico Moretti, "The Demand for Sons: Evidence from Divorce, Fertility, and Shotgun Marriage." Working paper, 2004.

Datar, Ashlesha, M. Rebecca Kilburn, and David S. Loughran, "Health Endowments and Parental Investments in Infancy and Early Childhood," RAND Labor and Population Working Paper No. WR-367, 2005.

Deschenes, Olivier, "Estimating the Effects of Family Background on the Return to Schooling" (March 1, 2002). *Department of Economics, UCSB. Departmental Working Papers*. Paper 10-02.

Doepke, Matthias, "Accounting for Fertility Decline During the Transition to Growth." Working Paper, UCLA, 2003.

Eide, Martha G., Nina Øyen, Rolv Skjærven, Stein Tore Nilsen, Tor Bjerkedal and Grethe S. Tell, "Size at Birth and Gestational Age as Predictors of Adult Height and Weight," *Epidemiology*, XVI (2005), 175-181.

Ermish, John and Marco Francesconi. "Educational Choice, Families, and Young People's Earnings," *Journal of Human Resources*, 35(1), 2000: 146-176.

Hauser, Robert M., and Hsiang-Hui Daphne Kuo (1998), "Does the Gender Composition of Sibships affect Women's Educational Attainment?", *Journal of Human Resources*, Vol. 33, No. 3, 644-657.

Jacobsen, Joyce P., James Wishart Pearce III, and Joshua L. Rosenbloom, "The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment." *Journal of Human Resources*, Summer 1999.

Kaestner, Robert "Are Brothers really Better? Sibling Sex Composition and Educational Achievement Revisited" *The Journal of Human Resources*, Vol. 32, No. 2. (Spring, 1997), pp. 250-284.

Møen, Jarle, Kjell G. Salvanes and Erik Ø. Sørensen, “Documentation of the Linked Empoyer-Employee Data Base at the Norwegian School of Economics,” Mimeo, The Norwegian School of Economics and Business Administration, 2004.

Neal, Derek A., and William R. Johnson. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104, no. 5 (1996): 869-895.

Oreopoulos, Philip, Mark Stabile, Randy Walld, and Leslie Roos, “Short, Medium, and Long Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins,” NBER Working Paper No 11998, January 2006.

Rosenzweig Mark R., and Kenneth I. Wolpin, "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment”, *Econometrica*, Vol. 48, No. 1, 227-240 (Jan, 1980).

Rosenzweig Mark R., and Kenneth I. Wolpin, “Natural “Natural Experiments” in Economics”, *Journal of Economic Literature*, December 2000.

Rosenzweig Mark R., and Junsen Zhang, “Do Population Control Policies Induce more Human Capital Investment? Twins, Birthweight, and China’s ‘One Child’ Policy” IZA Working Paper 2082. April 2006.

J.M. Sundet, K. Tambs, P. Magnus and K. Berg, “On the question of secular trends in the heritability of IQ test scores: A study of Norwegian twins,” *Intelligence* **12** (1988), pp. 47–59.

Sundet, Martin Jon, Dag G. Barlaug, and Tore M. Torjussen, "The End of the Flynn Effect? A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts During Half a Century", *Intelligence* , XXXII (2004), 349-362.

Sundet, Jon Martin, Kristian Tambs, Jennifer R. Harris, Per Magnus, and Tore M. Torjussen (2005), “Resolving the Genetic and Environmental Sources of the Correlation Between Height and Intelligence: A Study of Nearly 2600 Norwegian Male Twin Pairs,” *Twin Research and Human Genetics*, VII (2005), 1-5.

Thrane, Vidkunn Coucheron, “Evneprøving av Utskrivingspliktige i Norge 1950-53,” Arbeidsrapport nr. 26, INAS 1977.

Zajonc R. B., “Family Configuration and Intelligence” *Science,* Vol. 192, 227-236, 1976.

Figure 2:
Ability Distribution by Family Size

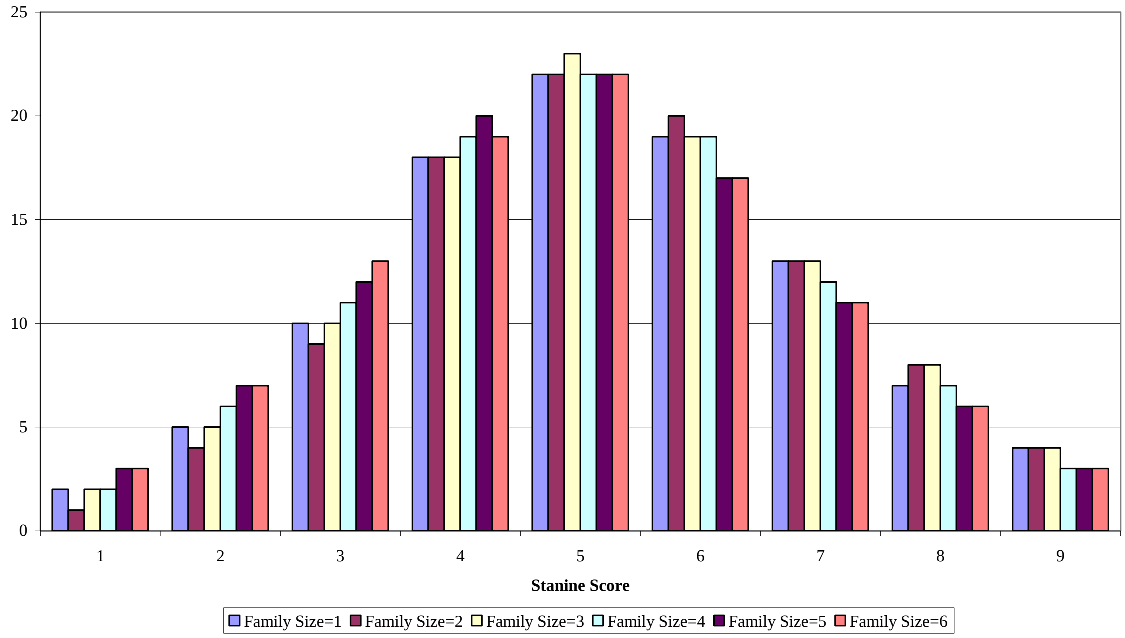

Table 1
Summary Statistics

Variable	Mean (Standard Deviation)
Age in 2002	25 (6)
Education	12.3 (1.9)
IQ (stanines)	5.2 (1.8)
Mother's Education	11.3 (2.5)
Mother's Age in 2002	50.5 (6.2)
Father's Education	11.8 (2.9)
Father's Age in 2002	53.6 (6.7)

Sample Size: N=388,405 male singletons.
Education is calculated for a subsample of 282,855 men aged at least 21 in 2002.

Table 2
Family Size and IQ

Family Size	Average IQ	Mother's Education	Father's Education	N
1	5.2	11.1	11.4	28,083
	(1.8)	(2.5)	(2.8)	
2	5.3	11.3	11.8	170,628
	(1.8)	(2.5)	(2.8)	
3	5.2	11.4	11.9	132,929
	(1.8)	(2.5)	(2.9)	
4	5.1	11.2	11.7	41,805
	(1.8)	(2.6)	(2.9)	
5	4.9	11.0	11.5	10,214
	(1.8)	(2.5)	(3.0)	
6 or more	4.9	10.6	11.2	4,746
	(1.8)	(2.4)	(2.8)	

Table 3: Effect of Family Size on Children's IQ
OLS

	No Controls				Controls			
	First Child	Second Child	Third Child	Fourth Child	First Child	Second Child	Third Child	Fourth Child
One Child Family	.099* (.031)				-.037 (.028)			
Two Child Family	.274* (.029)	.222* (.031)			.034 (.026)	-.015 (.028)		
Three Child Family	.256* (.030)	.236* (.031)	.184* (.035)		.052* (.026)	.037 (.028)	-.055 (.032)	
Four Child Family	.163* (.032)	.144* (.034)	.093* (.039)	.063 (.046)	.051 (.029)	.037 (.031)	-.028 (.035)	-.045 (.044)
N	201789	134692	42042	7879	201789	134692	42042	7879

* indicates statistical significance at the 5% level. Standard errors (in parentheses) allow for correlation of errors within family. Each column represents a separate regression. The regressions with controls include indicators for age, test year, mother's age, mother's age at first birth, mother's education, father's education, father's IQ, and the log of birth weight of the children studied.
Omitted Category is 5 or more children.

Table 4
Effect of Family Size on IQ of Children
Instrumental Variable Estimation

	OLS (no controls)	OLS (controls)	First Stage	Second Stage	N
Instrument: Twin at Second Birth (Sample: First child in families with 2 or more births)			.717* (.019)		173706
Number of Children in Family	-.052* (.005)	.006 (.005)		-.149* (.052)	
Instrument: Twin at Third Birth (Sample: First and second children in families with 3 or more births)			.839* (.018)		135970
Number of Children in Family	-.082* (.008)	-.002 (.007)		-.170* (.052)	
Instrument: Twin at Fourth Birth (Sample: First, second and third children in families with 4 or more births)			.903* (.036)		44695
Number of Children in Family	-.046* (.012)	.009 (.010)		-.115 (.080)	
Instrument: First Two Children Same Sex (Sample: First and second children in families with 2 or more births)			.082* (.003)		304755
Number of Children in Family	-.045* (.004)	.014* (.004)		.065 (.074)	
Instrument: First Three Children Same Sex (Sample: First, second, and third children in families with 3 or more births and first two births same-sex)			.070* (.005)		94578
Number of Children in Family	-.077* (.010)	.008* (.009)		.076 (.171)	

* indicates statistical significance at the 5% level. Standard errors (in parentheses) allow for correlation of errors within family. N represents number of individuals. The regressions also include indicators for age, test year, mother's age, mother's age at first birth, mother's education, father's education, father's IQ, and log birth weight of the children studied.

Table 5: Effect of Twin Births on Children's IQ

Dependent Variable: Child's IQ	Twins-at-first-Birth		Twins-at-Second-Birth	
First-Birth Twins	-.152* (.039)	.091* (.040)		
Birth Weight		.255* (.007)		
Second-Birth Twins			-.182* (.041)	.047 (.042)
Second-Birth twins x First Birth (non-twin)			.065 (.053)	.004 (.054)
First-Birth (non-twin)			.577* (.007)	.848* (.040)
Mean Birth Weight of second birth				.254* (.009)
Mean Birth Weight of second birth x first-birth (non-twin)				-.071* (.011)
N	204646	204646	308616	308616

* indicates statistical significance at the 5% level. Standard errors (in parentheses) allow for correlation of errors within family. All regressions include indicators for age, mother's age, and year of test. N represents number of individuals.

Appendix Table 1
Distribution of Family Size

Family Size	Percentage
1	10
2	47
3	31
4	9
5	2
6 or more	1

Appendix Table 2
Effect of Family Size on Whether Children Finished 12 Years of Schooling
Instrumental Variable Estimation

	OLS (No Controls)	OLS (Controls)	First Stage	Second Stage	N
Instrument: Twin at Second Birth (Sample: First child in families with 2 or more births)			.726* (.021)		150706
Number of Children in Family	-.030* (.001)	-.013* (.001)		-.029 (.015)	
Instrument: Twin at Third Birth (Sample: First and second children in families with 3 or more births)			.847* (.020)		113475
Number of Children in Family	-.034* (.002)	-.016* (.002)		-.039* (.015)	
Instrument: Twin at Fourth Birth (Sample: First, second and third children in families with 4 or more births)			.900* (.040)		36927
Number of Children in Family	-.016* (.003)	-.007 (.003)		-.040 (.025)	
Instrument: First Two Children Same Sex (Sample: First child in families with 2 or more births)			.081* (.003)		260236
Number of Children in Family	-.028* (.001)	-.012* (.001)		-.019 (.021)	
Instrument: First Three Children Same Sex (Sample: First, second, and third children in families with 3 or more births and first two births same-sex)			.069* (.005)		77410
Number of Children in Family	-.028* (.001)	-.014* (.002)		.023 (.049)	

* indicates statistical significance at the 5% level. Standard errors (in parentheses) allow for correlation of errors within family. N represents number of individuals. The regressions also include indicators for age, test year, mother's age, mother's age at first birth, mother's education, father's education, father's IQ, and the log birth weight of the children studied. Children are aged at least 21 in 2002.

Appendix Table 3: The Effect of Child Spacing on Previous Children's IQ

	Effect of Gap less than or equal to 1 Year	Effect of Gap less than or equal to 18 months
First child in families of three or more N = 74996	-.26* (.06)	-.12* (.02)
First and Second Child in Families of four or more N = 32772	-.18 (.09)	-.08* (.03)
First three children in families of five or more N = 9651	-.55* (.17)	-.19* (.06)

* indicates statistical significance at the 5% level. Standard errors (in parentheses) allow for correlation of errors within family. Each box represents a separate regression. The regressions include indicators for age, test year, mother's age, mother's age at first birth, mother's education, father's education, father's IQ, birth order, and the log birth weight of the child studied. Also included is a control for family size and a quadratic function of the time gap between children born before the two children whose spacing is being investigated.

The reported coefficients are those relating to the distance in years between the two children immediately after the children listed in column 1. For example, in the first box, the coefficient of -.26 suggests that first child IQ is lower by .26 if the gap between the second and third child is less than or equal to one year (in families with at least 3 children).