Separated at Girth:
U.S. Twin Estimates of the Effects
of Birthweight[+]

Heather Royer
Case Western Reserve University
heather.royer@case.edu

September 19, 2007

Abstract

The fetal origins hypothesis asserts that nutrient deprivation *in utero* can raise an individual's chronic disease risk. Within economics, this hypothesis has gained acceptance as a leading explanation for the cross-sectional correlations between birthweight, a proxy for fetal nutrient intake, and adult outcomes such as educational attainment, earnings and health. However, tests of this hypothesis using cross-sectional data may not adequately account for the effect of omitted variables such as family background and genetics on these outcomes. To estimate the effects of birthweight while controlling for these factors, I exploit differences in birthweight between twins. I use two datasets of twins: a newly-created dataset, consisting of over 3,000 twin pairs, coming from the universe of 1960-1982 California birth records and the Early Childhood Longitudinal Study-Birth Cohort. Using these data, I find that birthweight is related to educational attainment, later pregnancy complications, and the birthweight of the next generation. With the exception of pregnancy complications, the effects of birthweight are small, especially in relation to recent findings for other countries. However, I find that the protective effects of birthweight vary across the birthweight distribution. For instance, while the effect of higher birthweight on infant mortality is most protective for very low birthweight infants, the effect of birthweight on education is practically zero for babies weighing less than 2500 grams and is largest for births exceeding 2500 grams.

JEL codes: I1, I2
Keywords: birthweight, education, twins

[+] I would like to thank Martha Bailey, David Card, Tom Chang, Ken Chay, John DiNardo, Erica Greulich, Mireille Jacobson, Paco Martorell, Justin McCrary, Doug Miller, and seminar participants at the University of Michigan and Clemson University and attendees at the annual Robert Wood Johnson Scholars in Health Policy conference for their comments and suggestions. I owe all credit for the clever title to Neal Caren. I am grateful to the Robert Wood Johnson Foundation for generous support, and for the valuable research assistance of Meghan Cameron and Feng Pan.

I. Introduction

Studying the geographic distribution of chronic heart disease in England and Wales in the 1980's, David Barker, an English physician, encountered a striking pattern. The rates of such disease were strongly correlated with infant mortality rates 70 years prior (Barker et al., 1989). This observation led to his widely-cited fetal origins hypothesis which claims that:

> "fetal growth restriction - due to nutritional deprivation in early life - is an important cause of some of the most common, costly and disabling medical disorders of adult life including coronary heart disease and the related disorders hypertension, stroke and type 2 diabetes." (Barker, 2006).

Barker argues that fetal nutrient deprivation affects physiological development *in utero* and thereby susceptibility to chronic conditions as an adult. Randomized-controlled trials using animals support this hypothesis. For example, relative to fully-nourished rats, rats starved *in utero* have equal-sized brains but less developed non-neurological organ systems (Ozanne and Hales, 2002). Economists have interpreted Barker's fetal origins hypothesis broadly and have used it to explain the relationship between early health conditions, such as birthweight, and long-run socioeconomic and health status, such as wages, human capital acquisition, and disability (e.g., Almond, 2006; Case et al., 2005; Maccini and Yang, 2006; Meng and Qian, 2005).

The fetal origins hypothesis has many important economic implications. It would suggest that policies that aim to improve the prenatal environment (e.g., Medicaid expansions for pregnant women and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)) could reap large long-run health, human capital, and wage returns for both the current and future generations. Such effects are often ignored in cost-benefit calculations, potentially leading to an underinvestment in these programs. Moreover, tests of the fetal origins hypothesis shed light on research examining the life-cycle evolution of the socioeconomic gradient in health (e.g., Case et al. (2002) and Currie and Stabile (2003)).

Empirical tests of the fetal origins hypothesis in humans are difficult because fetal nutrients are not randomly-assigned.[1] Simple cross-sectional regressions between birthweight and heart disease mortality

[1] Empirical studies of the relationship between fetal conditions and long-run outcomes include Almond, 2006; Barker, 1995; Behrman and Rosenzweig, 2004; Black et al., 2007; Case et al., 2005; Conley and Bennett, 2000; Currie and Hyson, 1999; Currie and Moretti, 2007; Johnson and Schoeni, 2005; Lumey et al., 1997; and Oreopoulos et al., 2006.

provide much of the early evidence on this hypothesis (Barker et al., 1989). But this approach is unlikely to isolate the causal effect of fetal nutrients, as birthweight is strongly associated with socioeconomic status.

Researchers have used two principal empirical strategies to address these empirical limitations. The first involves exploiting historical events (e.g., the Dutch famine (Lumey et al., 1997) and the 1918 influenza epidemic (Almond, 2006)) which altered the *in utero* environment through starvation, stress, and/or sickness. The second approach uses twin comparisons, relating within-twin-pair differences in birthweight to differences in the twins' long-run outcomes (e.g., Behrman and Rosenzweig, 2004; Black et al., 2005; Oreopoulos et al., 2006).[1] Because recent research argues that variation in fetal nutrient uptake is the primary source of within-twin-pair variation in birthweight (Almond et al., 2005; Cunningham, 2001), this second approach is arguably a more direct test of the fetal origins hypothesis.[2] Barker's original hypothesis is explicitly about the long-run effects of fetal nutrients rather than, for instance, the long-run effects of maternal sickness during pregnancy. Maternal sickness could affect both fetal nutrients and other factors that contribute to long-run outcomes (e.g., socioeconomic status or birth defects).[3] The twins approach is also appealing since one's identical twin is a near-ideal counterfactual; genetic makeup and family upbringing, two factors that may drive the cross-sectional correlation between birthweight and long-run outcomes, are as comparable as may be possible in a non-experimental setting.[4] Overall, these twins studies have found large, positive effects of being the heavier twin.

Despite the conceptual appeal of the "twins" approach, the existing studies that use this strategy have some limitations. First, these estimates can be quite unstable, even within a study. For instance, Black et al. (2006) estimate a negligible effect of birthweight on high school completion for the 1967-1976 birth cohort, but for individuals born between 1977 and 1986, the estimate is nearly six times as large.[5,6] Second,

---

[2] In the case that the birthweight discrepancies between twins are due to other factors affecting long-run outcomes besides fetal nutrients, it would be incorrect to interpret the twins studies as a test of the fetal origins hypothesis. However, such estimates are still instructive about the effects of birthweight.

[3] There is some evidence that the influenza epidemic affected the incidence of birth defects (Reid, 2005). Also, given the high rate of maternal mortality during the epidemic, many of the infants affected *in utero* may have grown up motherless.

[4] While twins are alike along many dimensions, they consistently differ in birthweight. The absolute value of the difference in twins' birthweight in my sample is 282 grams, which is larger than the effect of smoking on birthweight (Almond et al., 2005).

[5] These estimates are statistically distinguishable from one another.

measurement error may bias some previous estimates. In particular, Behrman and Rosenzweig (2004) use fetal growth (birthweight/gestational length) as their measure of infant health, but because gestational length is measured with considerable error, such estimates can be inconsistent and with an indeterminable direction of bias.[7] Third, the sample of twins used in other studies are quite small, particularly for the United States (e.g., Behrman and Rosenzweig's (2004) sample consists of 402 twin pairs).

This study uses within-twin-pair comparisons to estimate the short- and long-run effects of birthweight. I use data from two sources: (1) a large new sample of twins, which I construct from the universe of 1960-1982 California birth records and (2) the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), which includes an oversampling of twins and follows them from birth. The birth record data provide information on long-run outcomes. The ECLS-B furnishes data on short-run outcomes including neonatal intensive care unit use and developmental outcomes.

This study makes a number of contributions to the existing literature. First, the datasets include relatively large samples of twins. In the birth record data, there are nearly 3,400 female twin pairs for whom I am able to observe adult outcomes, which is to my knowledge the largest U.S. dataset on twins with information on long-run outcomes. The large sample used in this study adds considerable power to the analysis, facilitating clearer inference about the contribution of birthweight to adult well-being, and overcoming the publication bias criticisms of this literature (Huxley et al., 2002).[8]

Second, due to the larger sample size, I am able to test whether the birthweight effects are non-linear.[9] Understanding not only whether but how increases in birthweight at different points in the birthweight distribution impact later life and intergenerational outcomes is crucial for designing targeted, cost-effective policies.

[6] In contrast, for all outcomes except diabetes, the effects of birthweight estimated in this study are consistent across cohorts.
[7] I prove this in Appendix C.
[8] Due to publication bias in this literature, Huxley et al. (2002) find that across studies, the estimated effect sizes are a decreasing function of sample size.
[9] The studies of Almond et al. (2005) and Currie and Moretti (2007) suggest that there are important non-linearities in the impact of birthweight on short- and long-run outcomes.

Third, this study focuses on recent cohorts in the United States. It is not clear whether the findings from earlier work pertain to the current U.S. context. Due to changes in immigration and economic conditions along with other factors, estimates may differ from those obtained using the previously-studied Minnesota twins (Behrman and Rosenzweig (2004)). Similarly, the relevance of recent estimates based on Norwegian and Canadian data (Black et al., 2005 and Oreopoulos et al. (2006)) for the U.S. are unclear. Intergenerational correlations in mobility are considerably smaller in the United States than in Canada and Norway (Blanden and Machin, 2005), and both countries have universally-provided health insurance.

Fourth, the data used in this study provide information on some adult chronic conditions, the exact data needed to test the fetal origins hypothesis. Such data are unavailable in the studies of Behrman and Rosenzweig (2004), Black et al. (2007), and Oreopoulos et al. (2006). Finally, this is the first study of these recent twin studies to examine both the short- and long-run effects of birthweight and whether investments made by parents, health care providers, and others are related to birthweight. Doing so helps to give a broader view of the mechanisms by which birthweight affects long-run outcomes.

Consistent with previous studies, I estimate a statistically significant relationship between birthweight and long-run and intergenerational outcomes. In particular, the heavier twin obtains more education, gives birth to heavier children, and has fewer pregnancy complications. In sharp contrast to earlier research, however, these effects tend to be quite small with the exception of pregnancy complications. For a 200 gram increase in birthweight, which is likely an achievable policy manipulation, education would be projected to rise by roughly 0.04 of one year. These negative effects of birthweight do not appear to be persistent across generations, as the estimated intergenerational correlation in birthweight is only 0.07. In contrast to other studies (e.g. Black et al., 2007), I find that the effects of birthweight on long-run outcomes are non-linear and for educational attainment, in particular, are largest above 2500 grams, the cutoff for defining low birthweight. These findings suggest that birthweight outside of the lower tail of the distribution (i.e., outside the range of low birthweight) should receive more attention.

Although estimating the long-run effects of birthweight via twin comparisons is theoretically appealing, there are several potential threats to validity. First, external investment by parents and health care providers may vary systematically with birthweight.[10] For instance, parents may seek to neutralize the effects of birthweight by investing more in the lighter twin, which could explain the small effects on long-run outcomes. Using the ECLS-B, I find no within-twin-pair relationship between birthweight and outcomes such as hospital days and neonatal intensive care unit use, suggesting little differential investment correlated with birthweight on the part of health care providers.

Second, as in any test of the fetal origins hypothesis, there is a concern about non-random sample selection. This is a direct implication of the hypothesis itself since a lack of fetal nutrients may increase early-age mortality rates. In addition, the construction of the dataset (i.e., the intergenerational match of birth records) may also lead to sample selection. However, nearly all of the sample selection appears to be due to birthweight-related infant mortality. Sample selection would likely lead to downward-biased estimates of the effects of birthweight since the low birthweight children who survive will be relatively robust. To ascertain the magnitude of bias, I perform a series of tests, which essentially amount to using a twin's birth cohort as an instrument for selection into the sample. The results suggest that sample selection does not explain the small estimated birthweight effects.

Third, like other twin studies (e.g., Almond et al., 2005 and Oreopoulos et al.., 2006), my twin sample includes both monozygotic (i.e., "identical") and dizygotic (i.e., fraternal) twins.[11] Within-twin-pair birthweight differences amongst both monozygotic and dizygotic twins may not be due exclusively to disparities in *in utero* nutrition but also due to differences in genetic makeup. As genetic advantage is positively correlated with birthweight, as suggested by the data, estimates of the effect of birthweight based on all twins will be biased upward. Moreover, in Black et al (2007), the estimates are similar for monozygotic

---

[10] If the estimated birthweight effect is intended to capture the biological effect of birthweight, then one would view parental investment as a confounder. But if one is interested in the reduced-form effect of birthweight inclusive of such parental behaviors, systematic parental investment based on birthweight differences is not problematic. As the second effect may be less policy-amenable, my goal is to estimate the first effect. In either case, the ability to which parents are able to neutralize or exacerbate the harmful effects of "low" birthweight is an independent outcome of interest and thus, deserves further attention.

[11] Monozygotic twins, commonly referred to as identical twins, are sometimes not genetically identical (Gringras and Chen, 2001). Monozygotic twins are defined as twins arising from the split of one fertilized egg. However, these twins can differ genetically, both at the level of chromosomes and DNA. It is unknown what fraction of monozygotic twins are genetically identical.

and dizygotic twins. Thus, this suggests that data on zygosity is not critical for estimating the pure effect of birthweight.

This paper proceeds as follows. Section II describes how I estimate the effects of birthweight using within-twin-pair comparisons. I follow with a description of the constructed panel data set of twins in Section III and a presentation of estimates of the effects of birthweight in Section IV. In Section V, I describe how one might interpret the estimated effects in light of postnatal investment and the inability to distinguish between monozygotic and dizygotic twins in the estimation sample. I compare my results to those of Behrman and Rosenzweig (2004), Black et al. (2007), and Oreopoulos et al. (2006) in Section VI and I conclude in Section VII.

II. Identifying the Effect of Birthweight

To test directly the fetal origins hypothesis, a researcher would need data on the intake of fetal nutrients. Such information is usually unavailable. As a proxy for fetal nutrient intake, I rely on birthweight, arguably the best measure of fetal nutrients. There is a strong cross-sectional relationship between the amount of weight a mother gains during her pregnancy and her infant's birthweight; an extra pound of maternal weight gain results in 0.02 of a pound increase in birthweight, based on calculations from the 1995 Detailed Natality File.[12] Although birthweight may be an imperfect proxy for fetal nutrition, within-twin birthweight differences still provide some signal about the returns to fetal food intake. This is true under the presumption that twin birthweight disparities result from differences in fetal nutrition uptake. In the case that the birthweight discrepancies between twins are due to other factors affecting long-run outcomes besides fetal nutrients, it would be incorrect to interpret the twins studies as a test of the fetal origins hypothesis. If this is true, one should interpret the estimates that follow more broadly as estimates of the effects of birthweight. However, the assumption that differences in fetal nutrient uptake drive twin birthweight differences is quite plausible, which I discuss later in this section.

[12] The effect size is the same whether or not I include controls for maternal age, maternal education, maternal race/ethnicity, state of residence, and birth order in the regression.

To describe the empirical approach, I begin with a simple linear relationship between birthweight and long-run outcomes:

$$(1) \qquad y_{ij} = \alpha + bw_{ij}\beta + \boldsymbol{x}_{ij}\delta + \varepsilon_{ij},$$

where y_{ij} is an adult or intergenerational outcome for individual j born to mother i, bw_{ij} is birthweight, $\boldsymbol{x}_{ij}$ is a vector of observable characteristics, and ε_{ij} is an error term.[13] Cross-sectional estimates of equation (1) likely lead to biased estimates of β, the parameter of interest, because of the correlation between immeasurable and unobservable determinants of y_{ij}, represented by ε_{ij}, and birthweight. Family upbringing and genetics are two examples of such confounding influences.

As noted by Almond et al. (2005), the interest in birthweight as a policy target is not due to its correlation with other factors but due to its direct effect. Indeed, if the correlation between birthweight and y_{ij} is entirely due to family background, policies aimed at increasing birthweight, which are unlikely to change family background, would be ineffective. Therefore, a desirable estimate of β captures the effect of birthweight holding constant omitted and hard-to-measure variables like socioeconomic status.

Suppose that the only confounders leading to an inconsistent estimate of β are family background and genetics. As long as there are no interactive effects of family background and genetic factors with birthweight and other observable characteristics, one can rewrite the error term as follows:

$$(2) \qquad \varepsilon_{ij} = h(f_i, g_{ij}) + u_{ij}$$

where h is a flexible function of family background f_i, and genetics g_{ij}, and u_{ij} is an error term assumed to be uncorrelated with the included variables of equation (1).

[13] I could estimate a more general model that allows the effects of birthweight to vary by birth order, effectively indexing β by j (see Royer (2004) for more information). However, when I estimate this less restrictive model for the sample of twins, I find that the twins are exchangable (i.e., I cannot reject the hypothesis that β is the same for the first- and second-born twin). Thus, I assume that β does not vary by parity.

If identical twins share the same genetic composition and family background, f_i is the same for both twins and g_{j2} equals g_{j1}. Under these assumptions, by taking twin differences of equation (1), one can consistently estimate β.[14] That is,

$$(3) \qquad y_{i2} - y_{i1} = (bw_{i2} - bw_{i1})\beta + (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})\delta + u_{i2} - u_{i1}.$$

In this setup, the effect of birthweight on y_{ij} is identified off of differences in birthweight within each twin pair holding fixed factors that are shared by the twin pair. Such factors include gestational length and maternal prenatal behavior. Birthweight and gestational length are highly correlated and policies that aim to increase birthweight (e.g., WIC) may also raise gestational length. However, birthweight appears to affect outcomes independent of gestation; OLS estimates of the effect of birthweight holding constant gestational length are larger in magnitude than OLS estimates not controlling for gestational length.

Equation (3) highlights several important points. First, β is a reduced-form parameter; it represents the life-course effect of birthweight. For instance, if birthweight differences between twins result in differences in IQ (James, 1982), the observed twin differences in education may be a direct result of these differences in IQ. To the extent that the determination of birthweight occurs before the determination of these other outcomes (e.g., education, IQ, adult health), one should not control for these outcomes when estimating equation (3). Thus, an appropriate interpretation of β is the long-run effect of birthweight through many possible pathways. By looking at a plethora of outcomes, I attempt to distinguish the mechanisms through which birthweight differences translate into differences in adult outcomes.

Second, twin comparisons improve upon simple sibling comparisons. The choice to have a second child may be endogenous to the first birth outcome (Rosenzweig and Wolpin, 1995). Royer (2004) estimates that the probability of having a second child is strongly correlated with whether the first birth was premature.

[14] This, of course, assumes that there are no interactive effects of birthweight across twins. For instance, there could be a psychological impact of twin size differentials. If, as others have shown, birthweight is related to later height and weight, the lighter twin may always be smaller and thus, may feel inferior to the heavier twin. This inferiority complex may affect outcomes such as educational attainment. This suggests that the degree of birthweight discordance should also be included as a regressor as a bigger discordance, if it leads to a larger adult size discordance, may be more traumatic. However, when stratifying based on birthweight discordance, I find, if anything, that the larger birthweight discordance is associated with smaller birthweight effects. But the statistical relationship between the estimated birthweight effect and birthweight discordance is not statistically significant.

Also another potential trouble with the sibling estimator is that several factors kept constant in the twins setting vary in the sibling setting. For example, non-twin siblings will develop at differing points of their parents' lifecycle. As wages vary considerably with age, at each age, these siblings will be subject to different parental resources.[15]

Third, β is still identifiable from within-twin-pair birthweight differences if the function *h* in equation (2), which describes the role of immeasurable and unobservable factors, includes other arguments besides genetics and family background. The crucial assumption is that these confounders are twin-invariant. Note that any twin-varying unobservables that are correlated with birthweight differences and can explain differences in adult outcomes will bias estimates of β. One such factor is a congenital anomaly. However, since the California twin data include information on congenital anomalies, I can control for these differences.[16]

Fourth, if there are interactions between unobservable factors and birthweight, estimates of β will no longer be consistent. For instance, suppose that parental investment is a function of birthweight. A mathematical characterization of parental investment could have the following form:

$$(4) \qquad y_{ij} = \alpha + bw_{ij}\beta + bw_{ij} f_i \gamma + \boldsymbol{x}_{ij}\delta + \varepsilon_{ij}$$

where I have augmented equation (1) with the $bw_{ij} f_i$ term. Then, taking the within twin difference,

$$(5) \qquad y_{i2} - y_{i1} = (bw_{i2} - bw_{i1})\beta + (bw_{i2} - bw_{i1}) f_i \gamma + (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})\delta + u_{i2} - u_{i1}.$$

In this scenario, without appropriate controls for family background, f_i, estimates of β will be inconsistent. As mentioned earlier, I later address the possibility that parental involvement is a systematic function of birthweight differences between twins. To do so, I investigate whether the effects of birthweight vary across different types of families. For instance, given limited resources, one might expect that differential parental investment across twins is more feasible in a small family.

[15] However, some may prefer the sibling comparison because of external validity concerns about twin comparisons.
[16] As a robustness check, I drop all twin pairs in which one or both twins have a congenital anomaly, rather than controlling for these anomalies.

Finally, if the effects of birthweight are non-linear, equation (3) will be misspecified since its explicit assumption is one of linear birthweight effects. To allow for non-linear effects, I estimate piecewise linear spline regressions, which have the general form:

$$(6) \qquad y_{i2} - y_{i1} = \sum_{m=1}^{k-1} (D_{i2}^{m} bw_{i2}^{*} - D_{i1}^{m} bw_{i1}^{*})\beta^{m} + (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})\delta + u_{i2} - u_{i1},$$

where k is the number of knot points, D_{ij}^{m} is a dummy variable equal to 1 if the birthweight of twin j born to mother i is greater than the knot point m, and bw* is (birthweight - knot point for the spline segment (e.g., 1000 grams)).[17]

The Causes of Birthweight Differences Among Twins

The identification strategy outlined above requires within-twin pair variation in birthweight. But if twins are so alike, why do they have different birthweights? First note that birthweight is generally thought to be a function of both gestational length and fetal growth for a fixed gestational length. Hence, an infant can be low birthweight either because of a short gestational period or because of slow fetal growth (otherwise known as intrauterine growth retardation (IUGR)). Twin gestational lengths are identical, so all variation in birthweight amongst twins is attributable to differences in fetal growth.[18]

The fetal growth rate of twins within the womb is governed by different factors, depending on the twin type. For monozygotic twins who are monochorionic (i.e., share the same placenta), the vascular arrangement of the placenta influences nutrient and blood flow, and thereby affects birthweight (Bajoria et al., 2001). Also, structural anomalies resulting from the splitting of the embryo may lead to one twin to receive more nutrients and oxygen than another.

Among all other – dizygotic and dichorionic – twins, the causes of birthweight differences are more disputed. While the exact mechanisms for birthweight differences are highly debated, the disparity in nutrient

[17] Note the coefficients from this regression specification are not directly interpretable as slopes of the relationship between birthweight and later outcomes. However, the marginal effects are linear functions of the estimated coefficients. For example, for a spline with 3 segments, the slope for the first segment is β^1, the slope for the second segment is $\beta^1 + \beta^2$, and the slope for the last segment is $\beta^1 + \beta^2 + \beta^3$.
[18] For some small subset of twins, gestational lengths differ.

uptake resulting from the structural arrangement of the fetuses (e.g., placenta placement) is the leading explanation for within-twin-pair birthweight differences (Almond et al., 2005; Cunningham, 2001).[19]

III. Data

Birth Record Data

I create the primary twins dataset using confidential individual 1960-2002 California birth records. These data, a census of California births, are compiled from forms completed at birth. These forms include questions on maternal and paternal demographics (e.g., age), infant health (e.g., birthweight and gestational length), birth order, and plurality (i.e., whether the birth was a multiple birth). The confidential version of these data also includes the mother's and child's names, which are used for matching.[20] I further describe the creation and matching processes in the Appendix A.

The final dataset consists of same sex female twins born between 1960 and 1982. As outlined in the Appendix A, I observe adult outcomes measured at the time of motherhood, only for those twins who have a birth observed in California between 1989 and 2002.[21] Thus, adult outcomes will be missing for the following potentially overlapping groups: women who have died, women who have moved from California, women who did not give birth between 1989 and 2002, and women for whom birth information (i.e., name and birthdate) are reported incorrectly on the birth certificate.[22]

---

[19] Arguably, an ideal twin study of the effects of birthweight on long-run outcomes would focus on monochorionic twins because their birthweight differences are more likely due to differences in nutrient uptake rather than genetic differences. Identification of such twins is rarely possible. Moreover, monochorionic twins may suffer from twin-twin transfusion syndrome in which blood flow is unevenly distributed amongst the two twins (Victoria et al., 2001). This condition is rare. Such circulatory problems can result in large birthweight discordance between twins. As this condition affects only monochorionic twins, estimates of the long-run effects of birthweight using such twins may not be externally valid and thus, not desirable. However, since the incidence of this syndrome is low, the monochorionic twins estimates may be generalizable to the singleton population.

[20] A mother's maiden name, not her married name, is reported in these data.

[21] As adult outcomes are only observed from the birth certificate, the analysis only focuses on female twins who are mothers.

[22] I estimate that twenty percent of missed matches are due to bad data. This is the percentage of 1989-2002 California births to mothers who were themselves reportedly born in California between 1960 and 1982 but are unmatchable to the mother's own birth.

This selection is only problematic to the extent that within-twin pair differences in birthweight are highly correlated with within-twin pair differences in the probability of later observation. Adult twins usually live near one another (thirty percent of twins observed as adults in the CA data live in the same zip code as their twin), so concerns about differential mobility within twin pairs may be moot. Additionally, as twins often have similar names, it may be reasonable to assume that child-mother matching differences within twin pairs are non-systematic. Sample selection due to death or lack of childbearing, however, may be more disconcerting. Later, I will directly assess the importance of these potential selection biases by testing whether birthweight differences amongst twins predict differences in the probability of later observation. Although I do find differences in the probability of having an observed birth that are related to birthweight differences, the differential probabilities are small. I assess the degree of sample selection bias and show that this bias tends to be small.

On a related note, one might question the generalizability of results using the sample of female same-sex twins born in CA between 1960 and 1982 and who gave birth in CA between 1989 and 2002. In particular, one may wonder (a) how CA-born twins compare to non-CA-born twins and (b) how CA-born twins giving birth in CA between 1989 and 2002 compare to other CA-born twins.[23] Unfortunately, there exists no ideal data to address this because adult twins are usually impossible to identify in standard datasets. As an alternative, I look at women born between 1960 and 1982, regardless of whether twin status using 2003 American Community Survey (ACS), the survey conducted in non-Census years intended to cover Census-type questions.[24] In the top panel of Appendix Table 2, I categorize women into three groups: (1) born in CA, gave birth between 1989 and 2002, and currently live in CA (the group labeled "Born in CA between 1960 and 1982 and meets twin sample criteria"), (2) born in CA but either did not give birth between 1989 and 2002 or does not currently live in CA (the group labeled "Born in CA between 1960 and 1982 and does not meet twin sample criteria"), and (3) born in a state besides CA (the group labeled "Born outside of CA

[23] There are other generalizability issues involved such as (a) how twins compare to singletons (an issue I discuss later) and (b) how twins born in 1960 and 1982 compare to twins born outside this period, which is probably a secondary issue.

[24] In the 2003 ACS, I am best able to identify women who have given birth between 1989 and 2002 (one of the selection criteria for inclusion in the twins sample).

between 1960 and 1982 and meets twin sample criteria"). The sample comparable to that used in this paper is sample (1).

Importantly, differences between the sample of women born in CA and born outside of CA are few. The women meeting the twin criteria (i.e., born in CA, still living in CA, and gave birth between 1989 and 2003) are slightly less educated than the women born outside of CA. Similarly, comparing the two populations born in CA (the first and second columns), in terms of educational attainment, the women meeting the sample criteria for the twins sample are less likely to have progressed beyond a high school degree. There are also some differences in the marital status of these two populations, as expected given that selection into the twins sample is based on childbearing behaviors. Extrapolating from these statistics, it appears that the socioeconomic status of the twins sample is lower than that of the overall population of women. Birthweight interventions such as WIC often target such groups, making the twins sample particularly relevant.

Early Childhood Longitudinal Study, Birth Cohort

To understand further the effects of birthweight, I also use the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), which follows a nationally-representive sample of infants born in 2001 from birth. Fortunately for my purposes, the study oversamples twins; there are 856 twins in the estimation sample. The data provide very detailed information on neonatal intensive care use, parental investment, and measure of development.

Descriptive Statistics

Table 1 presents summary statistics, measured at the time of birth, for singleton females, all twins, and same sex female twins from the birth record data. The first three columns do not condition on whether the twin is observed ever giving birth, while the last four columns only include those females observed giving birth. As expected, the chief characteristic that distinguishes singletons and twins is birthweight. Nearly 50 percent of all same sex female twins are low birthweight (i.e., weigh less than 2500 grams) while only roughly

5 percent of singleton females are low birthweight. Figure 1, which plots the singleton and twin birthweight distributions, further highlights these differences. Shifting the twin distribution to the right by 700 grams, the two distributions would nearly overlap.[25]

The differences between the singleton and twin birthweight distributions naturally question the external validity of the twin estimates. In particular, if twins are naturally small, one might worry that the results are not generalizable to the larger non-twin population, arguably the main population of interest. However, as discussed later, the cross-sectional relationships between birthweight and adult outcomes such as education tend to be similar for singletons and twins (see Figures 3-5). As such, these results may be applicable to singletons.

The final two columns of Table 1 provide at-birth summary statistics for the estimation population – same sex female twins for whom I observe a first or second birth. As a matter of comparison, I also provide the analogous statistics for singleton female births that have been matched to either a first birth (1st birth observed column) or a second birth (2nd birth observed column). The subset of the same sex female twins who are observed later is similar to the overall same sex female twins sample except in terms of birth outcome characteristics. In particular, twins in the estimation sample are less likely to have birthweights in the lower tail of the birthweight distribution than the unconditional population of same sex female twins is. This is not particularly surprising given the findings of Almond et al. (2005), which suggest that, at least at the lower extreme tail of the birthweight distribution, birthweight is a strong predictor of infant mortality. Hence, some non-trivial share of extremely low birthweight infants likely dies before reaching adulthood.

For those twins observed giving birth, I also observe adult outcomes such as their education at motherhood and the birthweight of their offspring. In Table 2, I compare adult outcomes for singleton female mothers with outcomes for same sex female twin mothers. The first three rows display the means and standard deviations for three measures of education – maximum level reported across births, mean level reported across births, and education at birth. Since a mother may obtain additional schooling after her first

[25] The maximum reported birthweight changed over the sample period, so to create a consistent birthweight measure, I topcoded birthweight to 4517 grams (9 lbs 15 ounces). This topcoding accounts for the small spike in the upper tail of the singleton birthweight distribution. Since the topcode is high, the regression results should be only minimally affected.

(and subsequent) births, her education as measured when giving birth may be a noisy measure of her completed education. However, within twin pair differences in education at any point in time, even before the completion of schooling, are informative. These differences may reflect differences in grade progression in addition to differences in eventual educational attainment.[26] In terms of adult outcomes, the twin mothers are very comparable to the singleton mothers.[27] Despite the fact that the twins themselves were likely to be of low birthweight, twin mothers give birth to infants of roughly the same weight as singleton mothers.

As a final useful set of summary statistics, Figure 2 plots the distribution of birthweight differences within twin pairs, which will later be exploited as a means of identifying the effect of birthweight. The gap in birthweight between twins is non-trivial; for over half of the twin sample, this difference exceeds 200 grams. The distribution of birthweight differences is nearly identical for same sex female twins as for all twins.

IV. Results

<u>*Birth Record Data*</u>

Plots of Relationships between Birthweight and Long-Run Outcomes

Before turning to the regression results, Figures 3-5 present plots of educational attainment, birthweight of offspring, and the number of pregnancy complications by the mother's birthweight (as measured in 100-gram increments) to give a sense of the relationships between birthweight and long-run outcomes for the birth record data.[28] The solid lines represent twin mothers and the dashed lines represent singleton mothers. The sample consists of females whose first or second birth is observed. If both the first and second births are observed, then the second birth value is assigned. While this sample will be the main

[26] This is true as long as the twins are observed at the same age. It is conceivable that birthweight affects fertility timing, so observed differences in education could be due to differences in age at childbirth. As discussed later, there are no statistically significant differences of twins' age at childbirth, thus mitigating that potential source of misinterpretation of the within-twin educational difference.
[27] Two outlier observations account for the large standard deviation in fetal growth for same sex female twins giving birth for the first time. For these observations, it appears that gestational length is severely misreported; for instance, a gestational length of 7 days is reported. These outliers, however, do not affect the later regression results.
[28] These outcomes were selected based on the fixed effects regressions in Table 3.

estimation sample, the results are robust to the use of other samples (e.g., twins with a first birth observed) as shown later.

In Figure 3, there is a clear relationship between birthweight and education.[29] The heavier a female is at birth, the more education she obtains. This is true for both singletons and twins. Moreover, the response functions for each subpopulation are nearly identical. The cross-sectional relationship between the twin mother's birthweight and the birthweight of her offspring in Figure 4 also reveals a strong positive correlation. Note that conditional on birthweight, the birthweight of a twin's offspring exceeds the birthweight of a singleton's offspring. However, most importantly, the response functions for the two populations have the same slopes. This pattern is likely reflective of differences in twin and singleton gestational lengths. For a given birthweight, singletons spend less time in the womb than do twins. The final graph, Figure 5 plots the number of pregnancy complications by birthweight.[30] A heavier birthweight is associated with a lower risk of pregnancy complications. Across all three outcomes, these response functions are strikingly alike for both singletons and twins, suggesting that twin-based estimates of the effect of birthweight may be externally valid.

Regression Estimates

Table 3 presents the main fixed effect estimation results relating within-twin-pair birthweight differences to differences in adult outcomes, as measured at the time of childbearing for the birth record data. I also present pooled OLS estimates, which do not control for twin fixed effects. The pooled OLS regressions control for the twins' birth order, the twins' year of birth, and the twins' race. In these models, the effect of birthweight is independent from the length of gestation. The data suggest that this is accurate

[29] Figures 3-5 provide raw, unadjusted means by birthweight. I present such figures as a means of assessing the external validity of the twin estimates. Since the cross-sectional patterns are sensitive to the controls included (results not shown) and furthermore, divergent from the fixed-effects relationships (see Table 3), it is unclear whether the estimates in Figures 3-5 should be adjusted for covariates and if so, which covariates. However, as external validity is ultimately untestable, Figures 3-5 should be interpreted as suggestive and not conclusive evidence of external validity.

[30] Pregnancy complications include hypertension, eclampsia (seizures during pregnancy), renal disease, kidney infection, cardiac disease, sexually-transmitted diseases, diabetes, hepatitis, rubella, Rh sensitization (the compatibility of Rh factor of the mother and the infant), hemoglobinopathy (presence of abnormal hemoglobins), uterine bleeding before labor, lung disease, polyhydramnios/oligohydramnios (excess/deficient amount of amniotic fluid), incompetent cervix, cervical circlage (tying the cervix closed in response to an incompetent cervix), and premature labor.

representation; for all of the outcomes of Table 3, I am able to reject the hypothesis that the effect of birthweight depends on gestational length.[31] I present the estimates in Table 3 such that the outcomes located at the top of the table are outcomes that birthweight may affect more directly (e.g., infant mortality, education, birth outcomes of the next generation, and health status). Outcomes for which the effect of birthweight may be less direct (e.g., age at motherhood when the twin gives birth and characteristics of the twins' mate) are at the bottom of the table.

As means of comparison to other studies (e.g., Almond et al., 2005), I also present the effects of birthweight on infant mortality within the first year of life. As in Almond et al. (2005), the estimates suggest that there is a strong cross-sectional bias in the birthweight-infant mortality relation. The twin fixed-effect estimate is one-tenth the size of the cross-sectional estimate and is nearly identical to the Almond et al. (2005) estimate of -0.0222.[32]

In this same row of Table 3, the fixed effects estimates for education imply that a one-kilogram increase in birthweight leads to a 0.12 to 0.16 of a year increase in educational attainment. These coefficients are 15 to 30 percent smaller than the cross-sectional coefficients.[33] The direction of bias in the cross-sectional estimates is as predicted. However, the size of these coefficients is misleading because any reasonable policy that manipulates birthweight is unlikely to alter birthweight by one kilogram. A foreseeable manipulation ranges from 200 to 250 grams. Thus, the fixed effects estimates indicate that while birthweight affects years of schooling, a realistic policy would only lead to a 0.03 to 0.04 increase in years of schooling. Assuming no other benefits to increasing birthweight, this hardly seems like a cost-effective investment.

It should be noted that education at motherhood is not necessarily completed education. For women 24 years old or older, education at motherhood is likely completed education.[34] The estimated effects

[31] This is accomplished by including a birthweight and gestational length interaction variable in the regressions.

[32] The data used to estimate the effect of birthweight on infant mortality are the 1960 and 1965-1980 California Birth Cohort files as opposed to the 1960-1982 California Natality data. This results in a slightly different sample than the sample used in the remainder of the table. However, if I use 1960 and 1965-1980 California Natality data for the non-infant mortality outcomes, the results are similar to those presented in Table 3.

[33] Currie and Moretti (2007) use the California birth records to look at the intergenerational transmission of birthweight. For mothers born between 1970 and 1974, they estimate that a one-kilogram increase in birthweight results in a 0.1548 increase in educational attainment. After they include grandmother fixed effects (i.e., compare mothers who are siblings), this point estimates drops in size to 0.0836, or about half the size of the twin fixed effect estimates in Table 3.

[34] In the 2000 Census, female school enrollments by age flatten at age 24.

of birthweight for this older sample are essentially identical to those in Table 3. Hence, the effects of birthweight on educational attainment in Table 3 are reflective of within-twin differences in completed education and not simply differences in educational progression.[35]

The second panel of estimates in Table 3 investigates whether birthweight affects one's own health and the health of one's offspring. Of the outcomes presented in Table 3, these outcomes test most directly the fetal origins hypothesis (i.e., the effect of birthweight on chronic conditions). It may be too early in the lifecycle to observe the impact of birthweight on chronic conditions; Barker's studies usually look at individuals in their 60's and 70's (Barker, 2006). Although disputed in the epidemiological literature, some studies have found that birthweight predicts adult outcomes such as hypertension (Poulter et al., 1999), coronary heart disease (Eriksson et al., 1999), and diabetes (Hales et al., 1991). The cross-sectional estimates imply that a 100-gram increase in a mother's birthweight leads to an 18-gram rise in her child's birthweight. Meanwhile, the fixed effects estimates of Table 3 suggest that this intergenerational transmission is much smaller – only one-third the size of the cross-sectional OLS estimate.[36,37] Gestational length is unaffected by a mother's birthweight.

In terms of one's own health, none of the estimated effects of birthweight on adult health outcomes (e.g., hypertension, diabetes, and anemia) in the third panel of Table 3 are statistically significant, but the magnitude of the hypertension estimate is sizable. A birthweight increase of 250 grams decreases the probability of hypertension by about 0.4 of a percentage point, a decline of 14 percent. The effects on diabetes and anemia are much smaller. The most notable estimate within this set of estimates is that for pregnancy complications. A birthweight increase of 250 grams implies an 11 percent fall in pregnancy complications. These estimates may be downward-biased due to misreporting of these conditions on the

---

[35] This interpretation is appropriate as long as the twins give birth at the same age, which is true.

[36] In Currie and Moretti's (2007) comparison of California-born siblings, the intergenerational transmission of birthweight is estimated as 0.2 (or an increase of 200 grams in the child's birthweight for every 1 kilogram increase in the mother's birthweight). This estimate is statistically indistinguishable from the OLS pooled twins estimate.

[37] If the variance of maternal birthweight equals the variance of the birthweight of the mothers' offspring, or in other words, birthweight across generations follows a stationary process, the coefficient from a regression of child's birthweight on mother's birthweight is directly interpretable as the intergenerational correlation in birthweight. In the data, there are slight differences in these variances. However, after taking into account these differences, the estimated intergenerational correlation in birthweight is larger only by a factor of 1.12 relative to the coefficient reported in Table 3.

birth certificate. The accurate reporting of clinical measures on the birth certificate can be poor (Buescher et al. (1993) and DiGiuseppe et al. (2002)). The general findings of the literature suggest that the measurement error in variables related to obstetric history, birthweight, and delivery type is small but for other outcomes, such as maternal risk factors and comorbidities, measurement error may be more problematic.[38]

Since these pregnancy complications include a heterogeneous group of conditions, I have disaggregated these complications as those reflective of long-term health complications (e.g., anemia and diabetes) versus those indicative of predominately pregnancy-related health (e.g., premature labor and eclampsia).[39] Birthweight appears to have a larger impact on pregnancy-related conditions as opposed to long-term health.

The third set of estimates in Table 3 examine whether there are any birthweight-induced differences in income-related outcomes. Unfortunately, California birth records have no direct income information. As a next best alternative, I look at four different outcomes that are indirectly related to a mother's income. Looking first at the C-section results and considering the high prevalence of this type of delivery among more affluent mothers, one might expect to observe a positive relationship between birthweight and C-section rates. But lighter twins have higher risks of pregnancy complications, which would likely result in a negative association between birthweight and C-section rates. However, the influence of birthweight on C-section delivery rates is weak at best, possibly reflecting the interaction of these two countervailing mechanisms. None of the other income-related outcomes – public payment for delivery (e.g., Medicaid-financed birth), the income of the zipcode of residence, or the poverty rate of the zipcode of residence – is strongly related to birthweight. Thus, with the exception of pregnancy complications, birthweight is only weakly associated with the outcomes in Table 3.

[38] The health-related estimates may be prone to measurement error due to the dichotomous nature of the dependent variable. In the standard measurement error model, measurement error in these discrete clinical measures (e.g., diabetes) will lead to attentuation bias where the attentuation factor is 1-probability of a false positive-probability of a false negative (Hausman et al., 1998), under the assumption that the missclassification rates are uncorrelated with birthweight. The estimates from DiGiuseppe et al. (2002) imply that ws should inflate the estimates in Table 3 by about a factor of 2-3 for most health-related outcomes except for anemia which we should inflate by 10.

[39] Eclampsia is a pregnancy condition that is characterized by convulsions.

Suppose as theories of fertility and mating predict and as empirical studies show, that a rise in a mother's education level leads to fertility delays and higher "quality" mates.[40,41] Then one might expect that the improvements in education associated with increases in birthweight would lead to delayed childbearing and maternal selection of older and more educated mates. The final panel of estimates in Table 3 tests this conjecture. The lighter and heavier twin give birth at the same age. While this estimate is informative about the effects of birthweight on fertility timing, it is also instructive about selection bias. If the twins have children at different ages, then age could potentially confound the estimated returns to birthweight because I only observe women at their chosen time of motherhood. None of the effects of birthweight on mate "quality" are significant or large, but the effect of birthweight on paternal education parallels the analogous effect on maternal education, suggesting a large mating market effect of education.

While the regressions in Table 3 are based on the sample of female twins with an observed first or second birth, as a robustness check, I replicate Table 3 using the sample of female twins with an observed first birth and not necessarily an observed second birth in Appendix Table 3. The estimates are consistent with those in Table 3, and as such, I use the larger sample of twins.

Another concern, besides the estimation sample, is that within-twin pair differences in the incidence of congenital anomalies (i.e., birth defects) could potentially explain the persistence of birthweight. Suppose one twin is born with a congenital anomaly and the other is not. This congenital anomaly discordance could cause a within-twin birthweight difference. Then, it would inappropriate to attribute the within-twin differences in long-run outcomes to their differences in birthweight via fetal nutrition. As such, it may be more appropriate to exclude twins with a congenital anomaly from the estimations. For twins born between 1960 and 1967 or between 1978 and 1982, I can identify whether a twin had a congenital anomaly. Appendix Table 4 replicates Table 3, restricting the estimation sample to twins born in years in which congenital anomalies were recorded on birth certificates. Appendix Table 5 also duplicates Table 3 but further excludes any twins with a congenital anomaly. Overall, the estimates in Appendix Table 4 are somewhat larger although statistically indistinguishable from the estimates in Appendix Table 5. Both sets of estimates are

[40] For these theories of fertility and mating, see Becker (1960), Becker and Lewis (1973), and Mincer (1963).
[41] For empirical studies of fertility and mating, see Currie and Moretti (2003) and McCrary and Royer (2006).

roughly of the same magnitude as the estimates for the overall sample in Table 3. Thus, the within-twin-pair variation in birthweight exploited in Table 3 does not appear to be due to within-twin-pair variation in congenital anomalies.[42]

Non-Linear Effects of Birthweight

Prior studies suggest that birthweight has non-linear effects on later outcomes. In fact, some studies (e.g., Johnson and Schoeni (2005)) focus exclusively on the lower tail of the birthweight distribution implicitly arguing that birthweight only matters when it falls below a certain threshold. If the effects of birthweight are a function of the level of birthweight, the regression estimates in Table 3 may not be representative of the effects throughout the distribution.

While recent economic studies (Almond et al. (2005), Behrman and Rosenzweig (2004), and Currie and Moretti (2007)) agree about the existence of non-linear birthweight effects, they find disparate locations of these non-linearities. Almond et al. (2005) argue that for infant mortality, birthweights at the bottom end of the distribution matter. In their study, the effect of birthweight on infant mortality is only sizable for birthweights below 1500 grams. For birthweight of one's offspring, Currie and Moretti (2007) find that the marginal return to birthweight is largest for mid-range birthweights. These seemingly contradictory results may reflect the different mechanisms through which birthweight affects different outcomes.[43]

To allow for the possibility of non-linear birthweight effects, I estimate a piecewise linear spline with a knot at 2500 grams in Table 4.[44] The reported F-statistics test whether the two segments of the linear spline have equal slopes.[45] The first set of estimates, shown in the top panel, indicates that the effect of birthweight on education and infant mortality is highly non-linear. Consistent with earlier work (Almond et al., 2005), the relationship between birthweight and infant mortality is strongest for the lower birthweight births. While

[42] This bias may also be small simply because the number of twins with congenital anomalies is small. The percent of twins with a congenital anomaly is 1.3 percent for the overall twins sample. In comparison, for the Almond et al. (2005) sample of infants born in the United States in 1989, this percent is 2.7.
[43] Other possible explanations for the disparate findings are varying samples and identification strategies.
[44] The use of 2500 grams as the knot point was chosen as 2500 grams is approximately the median birthweight; the choice is also based on the patterns found in Figures 3-5. I have experimented with other knot points and the substantive conclusions remain unchanged.
[45] For these F-statistics, the numerator degrees of freedom is 1 and the denominator degrees of freedom is the number of twin pairs minus two.

there is some indication of non-linearity in the effect of birthweight on infant death in both the OLS and fixed effects specifications, the inclusion of twin fixed effects dampens the non-linearity of the relationship. Meanwhile for education, the marginal benefit of birthweight on education is strongest in the 2500+ gram range according to the cross-sectional OLS estimates. The fixed effect estimates confirm these cross-sectional relationships, but the suggested degree of non-linearity is magnified. In particular, with education at birth as the dependent variable, the two segments of the linear spline have statistically distinct slopes. The estimated effects on education in the <2500 gram range are negative but insignificant. But at the upper end of the birthweight distribution, the effects on education are nearly twice as large as those reported in Table 3. An increase in birthweight of 200-250 grams in this part of the distribution is associated with an economically-insignificant increase in educational attainment on the order of 0.08-0.10 of a year.

The second set of estimates in Table 4 shows that the effects of birthweight on adult health are largest for mothers whose birthweight was low; hypertension, diabetes, and pregnancy complications all are declining functions of birthweight amongst low birthweight female twins. There is little, if any, adult health effects for mothers whose birthweights exceeded 2500 grams.

Looking at the intergenerational effects of birthweight, in the cross-section, the effect of a mother's birthweight on her child's birthweight is twice as large if the mother's birthweight was greater than 2500 grams than if it was below this threshold. These cross-sectional estimates are consistent with the findings of Currie and Moretti (2007). This non-linearity disappears after controlling for twin fixed effects.

The third panel of Table 4 shows the effects of birthweight on indirect measures of income. Except for the effect of birthweight on median household income of the mother's residential zipcode, the birthweight effects appear to be independent of birthweight levels. Given the positive wage returns to education and the positive estimated effects of birthweight on education, one would expect that the effect of birthweight on income would be largest among mothers in the upper half of the twin birthweight distribution. Instead, one sees the same pattern as observed for the health effects; the returns to birthweight as measured by residential median income are positive and statistically significant for mothers with birthweights below 2500 grams but negative and statistically insignificant for birthweights above this threshold. However, the effects of birthweight for low birthweight mothers are economically small; an increase of one standard deviation in birthweight leads to an increase of approximately $1,300 in the median income of a mother's residential area. This is equivalent, for example, to moving from Santa Cruz County to San Francisco County.[46]

The effects of birthweight on fertility and mating market opportunities in the final panel of Table 4 are consistent with the effects on education; they too are largest amongst mothers who weighed over 2500 grams at birth. For these "high" birthweight mothers, being heavier is correlated with delays in fertility and maternal partnering with an older and more-educated mate. The only effect that is statistically significant within this set of estimates is that of paternal age.

To further pinpoint the location of these non-linearities, I add two additional knot points at 1500 and 3000 grams to the linear spline specification used in Table 4. Appendix Table 6 presents these additional regression results. The main insight provided by this closer look is that the pregnancy complication risks of birthweight are only present for mothers with birthweights falling between 1500 and 2500 grams. Meanwhile, the impact of birthweight on the birthweight of one's offspring and educational attainment are roughly in agreement with the earlier spline estimates in Table 4.[47]

[46] This example is based on a cross-county move, although the household median income data is measured at the zipcode level.

[47] The linear spline specification is one of several ways to model the non-linear effects of birthweight. Studies such as Case et al. (2005), Conley and Bennett (2000), and Johnson and Schoeni (2005) have focused on the long-run effects of low birthweight. Implicit in such a specification is that the effects of birthweight are negligible for birthweights exceeding the low birthweight threshold of 2500 grams. The results in Table 4 suggest that this assumption is too strong. To better compare my estimates to those in this other literature, Appendix Table 7 reports the pooled OLS and twin fixed effect estimates of low birthweight. The estimates suggest that low birthweight has a detrimental effect on educational attainment but little effect on other outcomes. The effects of low birthweight on the likelihood of being a high school dropout are about one-fifth of the magnitude of the effects found by Johnson and Schoeni (2005) using sibling comparisons (results not shown).

Sample Selection

A credible empirical test of the fetal origins hypothesis inherently is difficult because of sample selection. In particular, this hypothesis predicts that individuals experiencing unfavorable *in utero* conditions may not survive into adulthood and thus, would not be observed in the data. Additionally, given the construction of the data, there are three other reasons why long-run outcomes may be missing: 1) the twin moved away from California; 2) she did not have a child between 1989 and 2002; and 3) there were data errors in her birth records. If a woman's birthweight affects her probability of later observation, the fixed effect estimates could be subject to sample selection bias. Estimates from the 2000 Census suggest that migration out of California is not strongly related to educational attainment; hence, it is unlikely that birthweight has an impact on migration out of California. Because I am using within-twin-pair variation, sample selection bias due to mortality and fertility are probably the most disconcerting. In this section, I assess the degree to which sample selection bias affects the estimates. While there is a correlation between birthweight and the probability of being observed, sample selection bias is minimal.

Table 5 presents estimates of the effect of birthweight on the probability of later observation.[48] *A priori* one would predict that selection into the sample would be an increasing function of birthweight. This is exactly what is found. The baseline fixed effects estimates, shown in panel A, imply that a birthweight increase of 200 grams increases the probability of a later observed birth by 0.5 percentage points. These estimates seem small and inconsequential. In Panel B, selection appears to be strongest among fairly normal-sized twins (i.e., those with birthweights between 2500 and 3000 grams). To further gauge the size of these effects, one can compare these estimates to the effect of birthweight on infant mortality. The effect of birthweight on infant mortality on the probability of later observation is about two-thirds of the size of the effect of birthweight on the probability of selection into the sample. As such, most of the sample selection appears to be the result of low birthweight infants dying in the first year of life.

[48] Although the outcomes are dichotomous, I estimate linear probability models for ease of interpretation.

To measure the extent to which this sample selection potentially biases the twin fixed-effects estimates, I perform a series of "non-parametric" tests, which I describe in full detail in Appendix B.[49] Briefly, I first test whether the effect of birthweight on the probability of later observation is the same across birth cohorts. Then, I test whether the effect of birthweight on long-run outcomes is identical for these same cohorts. The intuition is that if I find that the effect of birthweight on the probability of being observed later differs across cohorts, there should be heterogenous effects of birthweight on long-run outcomes across birth cohorts in the presence of sample selection bias. This is assuming that the effect of birthweight on long-run outcomes is the same across cohorts, which may be justified given that the effect of birthweight on infant mortality is similar across cohorts in the sample. If instead the effects of birthweight on long-run outcomes are identical across cohorts but the effects of birthweight on selection into the sample are not, then sample selection bias may not be an issue.

In this case, I am able to strongly reject that the effect of birthweight on the probability of later observation is the same across cohorts. And, for all outcomes excluding diabetes, I am unable to reject the null hypothesis that the long-run effects of birthweight are identical across cohorts. Thus, the results of this non-parametric test suggest that sample selection bias is not problematic.

National Childhood Longitudinal Study, Birth Cohort Data

To understand how birthweight affects long-run outcomes, it is important to examine the effect of birthweight earlier on in the life cycle. Table 6 displays results for the twins sample for the ECLS-B. The first two sets of results examine the relationship between birthweight and neonatal intensive care use (NICU) and days in hospital following birth. The OLS relationships indicate a strong correlation between birthweight and post-birth care. For instance, a typical within-twin-pair difference in birthweight would lead a within-twin-pair difference in the probability of NICU use of 0.1, which is quite large given that the mean NICU use

[49] Alternatively, I could estimate the model with a sample selection correction. To do so, one would have to overcome the difficulty of finding a variable that affects the probability of later observation but not the outcome variable. The difficulty of this task is exacerbated in the context of twins because the requested variable must be measured at birth, must differ within twin pairs, and also must affect the probability of later observation in the 1989-2002 California birth records.

is 0.34. For both NICU use and days in hospital, however, the estimated effect of birthweight falls quite dramatically with the inclusion of twin fixed effects.

The developmental outcomes suggest a similar pattern; that is, the effect of birthweight in the short-run is negligible. These outcomes measure infant's skills such as the ability to recognize the source of a sound and the ability to hold a ball. The mental and motor scores are standardized. The estimates of the effect of birthweight on these outcomes are quite small; a 250 gram increase in birthweight only translates into a 0.03-0.04 of a standard deviation increase in these scores. Overall, consistent with the earlier finding using the birth records, these results suggest that the effects of birthweight on short- and long-run outcomes are negligible. Although not displayed, estimates are similar when the sample is confined to identical twins and female twins.

V. Understanding the Effects of Birthweight

Postnatal Investments

The long-run effects of birthweight presented in Tables 3 and 4 are reduced-form estimates. They represent the effects of birthweight throughout a woman's life course, including postnatal investment by her parents and her health care providers. Such parental investments may obfuscate identification of the biological effect of birthweight. For example, parents may seek to equalize the opportunities of their children, and thus invest more heavily in the lighter, disadvantaged twin.[50] Such behavior would dampen the long-run effects of birthweight. On the other hand, recognizing that there are potentially larger average returns to investing in the heavier twin, parents may favor the heavy twin. This would exacerbate the twin differences. In addition, the lighter twin may receive more medical care because of the risks associated with

[50] To discern whether parents invest differentially in their children, an extensive public finance literature has examined gift-giving and bequests from parents (see Bernheim and Severinov, 2003 for citations). Bequests are usually split equally between children, but gifts before death tend to be unequal. Bernheim and Severinox (2003) develop a model to explain this puzzle. They argue that gifts can be unequal because children cannot directly observe the degree to which their parents love them, so gift-giving acts as a signal. If gift-giving is observable, siblings who do not receive gifts will infer that their parents do not love them as much. However, if gift-giving is secret, parents make unequal gifts to their children without their children's knowledge. In the context of this study, parental investment during childhood may be most important for long-run outcomes. Differential parental investment is unlikely to be secret if the twins are living together.

low birthweight. Under this scenario, the estimated effects would be a downward-biased estimate of the biological effects of birthweight. Without knowing whether compensatory or reinforcing investment is more common, it is impossible to know the direction of bias due to postnatal investments. However, independent of this potential bias, the degree to which postnatal interactions offset the long-term effects of birthweight is of interest to both parents and policymakers.

Using the ECLS-B (results in Table 6), I find no systematic relationship between birthweight and early medical care, which may be a postnatal investment decision made by health care professionals rather than parents. These results hold along other dimensions such as breastfeeding, which is not probably surprising. However, parents and health care providers can participate in compensatory or equalizing behaviors that may be difficult to measure via a survey. For example, the quality and length of time spent with each child may not be accurately reported or remembered. But these estimates suggest that along observable dimensions, there is little evidence of either compensatory or reinforcing behavior.

Measuring postnatal investments is more difficult in the birth records. For this reason, rather than measuring whether investments are responsive to birthweight, I use the birth records to examine whether the effects of birthweight differ across different families that may have varying abilities to invest in one twin versus another. For instance, it is quite plausible that the potential for parents to treat each differently varies by family size. A large family with limited resources may be unable to treat each twin differently, and thus, estimates based on large families may be closer to the true biological effect of birthweight. In results not shown, the birthweight effects on education tend to be smaller but non-negligible in larger families (i.e., families where the twins have at least two older siblings). The effects of birthweight on pregnancy complications are larger in bigger families. The educational attainment results support the theory that parents offer more resources to the heavier twin; this leads to upward-biased estimates of the effect of birthweight. These results are only suggestive since the imprecision of these estimates does not allow me to differentiate these new estimates from the estimates in Table 3.

While the two data sources suggest that equal resources are devoted to each twin, it is important to put these estimates in context with findings from other studies. Loughran et al. (2004) and Datar et al. (2006)

test whether parental investments vary with birthweight. Both of these studies take advantage of sibling comparisons from the National Longtitudinal Survey of Youth-Child File and correlate differences in birthweight with differences in parental investments (i.e., age at school entry, maternal labor supply, and family size in Loughran et al. (2004) and breastfeeding, well-baby visits, immunizations, preschool attendance, and kindergarten entry age in Datar et al. (2006)).[51] For the relevance of this study, these estimates are probably upward-biased estimates of the effect on the level of parental investment as it is likely easier for parents to invest differentially in non-twin siblings relative to twins.[52] Along all measured dimensions except kindergarten entrance age, the results of Datar et al. (2006) suggest that parents participate in reinforcing behavior. The results of Loughran et al. (2004) are less clear. Using Chinese twins, Rosenzweig and Zhang (2006) also find some supportive evidence that parents participate in reinforcing behaviors in terms of schooling expenditures. Given all of these results, we might interpret the twins estimates of the returns to birthweight as upward-biased estimates of the biological effect of birthweight on long-run outcomes although results from the birth records and the ECLS-B suggest that this degree of bias is negligible.

Monozygotic versus Dizygotic Twins

As in many other twin studies (e.g., Almond et al., 2005; Oreopoulos et al., 2006; Conley et al., 2006), in these data, I cannot distinguish between monozygotic and dizygotic twins in these data.[53] Genetic advantage is likely positively correlated with birthweight as the incidence of congenital anomalies, many of which are genetic, are decreasing with birthweight.[54] Therefore, the twin fixed effects estimates of the long-

[51] Some of these outcomes (e.g., age at school entry) are very unlikely to differ among twins.
[52] If resources are fixed, then an increased investment in one sibling mechanically leads to a decrease in the investment in the other sibling. As such, the sibling estimator is likely an upper bound of the effect of birthweight on parental investment.
[53] Roughly sixty to eighty percent of all twins (and a lower percent of same sex female twins) are dizygotic. This percentage has grown recently with the increasing popularity of assisted reproductive technologies such as *in vitro* fertilization (Cunningham et al., 2001). The first successful use of *in vitro* fertilization in the United States occurred in 1981 (Buckles, 2007) and thus would potentially affect only the youngest cohorts in the twins sample. But, as seen in Appendix Figure 2, both the overall twinning rate and the fraction of births that are same sex female twins remain relatively constant over the 1960-1982 period. Moreover, in United States natality data, the rise in multiple births per pregnancy is only evident in the late 1980's (Buckles, 2007). These facts provide assurance that while most of the twins in the sample are likely dizygotic, the fraction that is monozygotic is not changing substantially over the sample period.
[54] By definition, congenital anomalies are defects at time of birth. They can be genetic defects or damage incurred in the uterus or at the time of birth.

run effects of birthweight, calculated using data on both dizygotic and monozygotic twins, will likely provide an upper bound of the effect of birthweight via prenatal nutritional deprivation.[55,56] It should be noted that for Black et al. (2006), the estimates for monozygotic twins are quite similar to those for dizygotic twins, which suggests that the genetic bias is small.

VI. Comparison to Existing Literature

There is a plethora of mainly small-scale epidemiological studies examining the long-run effects of birthweight.[57] However, publication bias may be a concern with such studies; Huxley et al. (2002) document a strong inverse relationship between estimated effect sizes and sample size. Recently economists have estimated such long-run relations, focusing mainly on human capital outcomes, which are usually ignored in epidemiological studies. While this economics literature improves upon the earlier epidemiological studies, particularly by employing large samples, the results can be very inconsistent across and even within studies. For example, Black et al. (2006) estimate substantial and statistical significant differences in the effect of birthweight across different birth cohorts. One potential explanation for such inconsistencies is sample

[55] To determine the extent to which the birthweight differences signal differences in underlying health rather than genetic differences, other studies (Almond et al., 2005; Black et al., 2007; Conley et al., 2006) contrast estimates based on opposite sex twins to those based on same sex twins. The underlying assumption is that sex composition does not have an independent effect on the outcome. For the Almond et al. (2005) and the Conley et al. (2006) studies, which relate birthweight differences to differences in infant mortality, this assumption may be innocuous. However, when looking at adult outcomes, as Black et al. (2007) do, it is not. Moreover, as working behaviors of males and females differ dramatically, it is not surprising that the estimated effects of birthweight on earnings and education in Black et al. (2007) differ by twin type.

[56] To assess the degree of bias due to genetic factors, I compare the effects of birthweight by race. From other studies, it is clear that the rate of dizygosity varies by race. Conditional on a twin birth, blacks are more likely give to birth to dizygotic twins relative to whites (Cunningham et al., 2001). Surprisingly, when comparing black and white twins, the within-twin pair estimates of the effect of birthweight on education are smaller amongst black twins than amongst the full sample. However, the effects on pregnancy complications are larger amongst black twins. Given that the implied direction of bias, at least for the effects on education, is opposite of that predicted, the differences between these estimates and the main set of estimates may be due to heterogeneous birthweight effects across racial groups.

[57] Such studies include Dwyer et al. (1999), IJzerman et al. (2005), Loos et al. (2001), Poulter et al. (1999), and Zhang et al. (2001).

selection bias; the effects of birthweight on education are largest for the cohorts who are less likely to be observed as adults.[58]

The purpose of this section is to directly compare the estimates across these studies. Unfortunately, simple comparisons across studies are nearly impossible due to a lack of a unifying regression framework across these studies (e.g., differing functional form and dependent and independent variables). Overall, my estimates in relation to other economic studies (e.g., Behrman and Rosenzweig (2004), Black et al. (2007), and Oreopoulos et al. (2006)), are much smaller.

Behrman and Rosenzweig (2004) is the only study of this group using US twins. They use the Minnesota Twins Registry consisting of monozygotic twins born in Minnesota between 1936 and 1955, who were re-surveyed as adults. Of the 10,400 surviving twins born within these years, Behrman and Rosenzweig have complete data for 804 female twins.[59,60] In the bottom panel of Table 7, I replicate Behrman and Rosenzweig's estimates; the sample means of the two samples are similar (top panel of Table 7) with the exception of educational attainment. Behrman and Rosenzweig use fetal growth as their measure of healthiness of birth because fetal growth is arguably a better measure than gestation or birthweight, alone. But dividing birthweight by gestational length likely introduces substantial measurement error and thus leads potentially to inconsistent estimates, which may be either upward- or downward-biased (see Appendix C for the proof).[61]

---

[58] Usually we would believe that selection bias would lead to downward-biased estimates based on selection into the sample being an increasing function of birthweight. Black et al. (2006), however, do not present such estimates.

[59] The 10,400 total does not include those twins born and dying during infancy. Only roughly 80 percent of the live-born twin pairs born during this period were intact after one year (i.e., neither of the twins died within the first year of life) (Lykken et al., 1990).

[60] Behrman and Rosenzweig do not explicitly address the potential selection bias due to this response and reporting bias. They recognize that birthweight could potentially affect infant mortality and thus lead to selection bias; they argue that such a worry is unfounded, given the results of Almond et al. (2005). However, the cohorts studied in Almond et al. (2005) were born thirty years later than the cohorts examined by Behrman and Rosenzweig (2004). Between the births of these two cohorts, there were significant improvements in infant mortality for low birthweights (Cutler and Meara, 1999), suggesting that the effects of birthweight on infant mortality are time-variant. Almond et al. (2005) estimate comparable effects of birthweight on infant mortality for twins born in the 1980's and 1990's. But the sharpest reductions in mortality occurred in the 1960's and 1970's.

[61] Gestational length is usually calculated from a woman's reported date of last menses, which may be easily forgotten and misreported and is only an approximation of the date of conception while birthweight is measured with considerably less error (Cunningham et al., 2001). New technologies such as sonograms provide more accurate estimates of gestational age than do imputations based on date of last menses, but the estimates using these advanced technologies are rarely reported in the natality files.

The OLS estimates, particularly that of offspring's birthweight, are of the same magnitude after accounting for the sampling variation. Nevertheless, once I control for twin fixed effects, the effect of fetal growth on education at birth is halved. In contrast, the fixed-effects estimate of Behrman and Rosenzweig greatly exceeds the OLS estimate and is more than six times the size of my fixed effects estimate. Given that their sample contains only 804 twins, this estimate is relatively imprecise. My fixed-effect estimate of the effect of fetal growth on the birthweight of one's offspring is extremely similar to that of Behrman and Rosenzweig.

Black et al. (2007) use Norwegian data created from the merger of several administrative datasets and rely on an estimation sample of 14,882 twins born between 1968 and 1981.[62] In Table 8, one can observe that the sample means for fetal growth and birthweight are nearly identical in the two studies, but the birthweight and fetal growth variances are larger in the Black et al. sample. The birthweight and infant mortality relationships (middle panel of Table 8) are quite similar although the birthweight and infant mortality relation is stronger in the United States.

In terms of long-run outcomes, Black et al. (2007) focus on the returns to birthweight on high school degree completion rather than years of education because of worries of sample size restrictions.[63] My OLS estimates of the educational returns to birthweight, using three different measures of infant health as a function of birthweight, are consistently half the size of the Black et al. OLS estimates. Once controlling for twin fixed effects, this difference is magnified – the Black et al. estimates are roughly three times larger than my own.[64,65]

[62] Of the 14,882 twins, 5,198 are same sex female twins.
[63] Black et al (2007) note that in order to use years of education as the dependent variable, they must restrict the sample to individuals 25 years old and older. This sample restriction apparently results in very imprecise estimates.
[64] To maintain a consistent sample across regression specifications, I drop all observations for which gestational length is missing, since fetal growth is a function of gestational length. This sample restriction removes a non-trivial fraction of the twins sample. The results that do not use gestational length are qualitatively similar with and without the observations that are missing gestational length. The fixed-effect estimate of the effect of birthweight is larger – 0.035 rather than 0.018 – but the two estimates are within sampling error of one another.
[65] To further understand these differing results, I have estimated the effects of birthweight across the educational distribution. For the California twins, the biggest effects are observed along the margin of a high school degree. The non-linear effects of birthweight on educational attainment are driven by the effects of birthweight on college completion.

Although not displayed in Table 8, the estimates of Black et al. (2006) vary considerably across cohorts. For instance, the effect of birthweight on high school completion in Black et al. (2006) is 0.04 for the 1967-1976 cohort (number of observations=9,500) and 0.22 for the 1977-1986 cohort (number of observations=3,622).[66] Sample selection appears to be strongest for the cohort with the largest birthweight effects; the effects of birthweight on infant mortality are bigger for the 1977-1986 cohort.[67] Looking across cohorts in my sample, however, with the exception of diabetes, I find no differential effects of birthweight.

The last study, Oreopoulos et al. (2006) focuses on Canadian twins and siblings from the Manitoba province born between 1979 and 1985. In total, there are approximately 40,000 siblings and 1,300 twins (750 twin pairs) in their sample.[68] In the top panel of Table 9, I contrast estimates from linear models (i.e., models in which an indicator for high school completion is regressed on birthweight). While the Oreopoulos et al. (2006) OLS estimates greatly exceed the analogous California estimates, the twin fixed effects estimates are quite comparable. In the non-linear regressions, the birthweight effects exhibit similar patterns across the birthweight distribution, but the magnitudes of birthweight effects are larger for the sample of Canadian twins. However, given that Oreopoulos et al. (2006) do not have many twin pairs, their estimates are comparatively imprecisely measured.

Overall, the estimates presented in this study suggest a much more muted role of birthweight in predicting long-run and intergenerational outcomes. While I have presented reasons for the disparities in results across studies, it is not entirely clear that one would expect similar estimates across studies, especially across countries. Even if the biological mechanisms by which birthweight affects long-run outcomes are invariant, it is not necessarily true that the social factors that can intensify or weaken this relationship do not have a country-specific or a time-specific component. In particular, we might think of the United States as a less egalitarian society when compared to Canada or Norway. Given this observation, we may expect that parents in the United States participate in less compensatory behavior. This would lead one to believe that the birthweight effects would be bigger in the United States than in these other countries. But this is not the

[66] These estimates are statistically distinguishable from one another.
[67] However, intuitively, it seems that sample selection would bias the estimates downward.
[68] These sample sizes come from Table 8 of their paper.

case. With the available data though, we know little about cross-country differences in these parental behaviors. Hopefully, with a growing interest in this field of research, researchers will collect such data in the future.

VII. Conclusion

This paper uses a new, large sample of California-born twins to estimate the long-run and intergenerational effects of birthweight, a prime important measure of infant health. To do this, I exploit the fact that twins, even monozygotic twins, frequently have unequal birthweights. I measure the extent to which these differences in birthweight translate into differences in adult and intergenerational outcomes. This approach is appealing because it controls for unobserved heterogeneity across individuals, a potential confounder in cross-sectional analyses.

While birthweight does have a statistically significant impact on many long-run outcomes – education, birthweight of one's offspring, and pregnancy complications – the estimated effects are typically small. Increasing birthweight by a conceivable 250 grams only leads to 0.03-0.04 of a year of additional schooling. Additionally, the short-run effects of birthweight are quite small; in terms of development and health care investment, I observe no differences between the lighter and heavier twin. However, I do find large effects for pregnancy complications. Specifically, a 250-gram increase in birthweight is associated with a 1.3 percentage point, or an 11.4 percent, decline in the number of such complications but these complications are reflective of pregnancy-related rather than long-run health problems.

These mean effects mask the effects of birthweight at different points of the birthweight distribution. The positive effect of birthweight on education is largest for births exceeding 2500 grams, a range where outcomes are often assumed to be unaffected by birthweight. This is a new and important finding suggesting that returns to increases in birthweight may be reaped from "normal-weight" births. As such, the concentration on low birthweight may be misplaced. On the other hand, the negative effects of birthweight on pregnancy complications are concentrated among low birthweight women. While it is not surprising that these effects are non-linear, it is unanticipated that the shape of the birthweight response function differs across outcomes. As such, a uniform theory such as the fetal origins hypothesis is unlikely to be a completely satisfactory explanation for the long-run effects of birthweight. Many different mechanisms may be at work. This is an important area for future research.

Establishing the existence and determinants of the non-linear effects of birthweight is important for policy decisions. Policies with goals of increasing birthweight (e.g., Medicaid expansions) often only target women at risk of delivering low birthweight babies. This research suggests that benefits, in the form of increases in educational attainment, may be reaped by raising birthweights amongst other populations. The robustness analyses suggest that, if anything, these estimated birthweight effects are upward-biased, implying an even more muted role of birthweight in the determination of short- and long-run outcomes.

Appendix A: Data Creation

I use two steps to create the California twins dataset. In the first step, I identify the birth of twins in the 1960-1988 California birth files using the plurality indicator variable. I retain all twins for which there are exactly two observations for each unique child birth date and child first and last name combination.[69] Since I focus on differences in adult outcomes for twins later observed giving birth, I discard all twin pairs in which at least one of the twins is male.

In the second step, I match the birth record information of the twins' births to that of the birth of their offspring. That is, for each twin, I determine whether I later observe her giving birth in California. I perform this matching procedure for all 1989 to 2002 California births. The matching is based on the twin's first and last name and her date of birth.[70] This matching algorithm will miss all births to twins occurring outside of California or happening before 1989 or after 2002. The final dataset consists of same sex female twins born between 1960 and 1988; for those twins whom I observe having a California birth between 1989 and 2002, I have longer-run outcomes (e.g., health, education) measured when they give birth. Hence, because birthweight may be a predictor of whether I observe a twin as a mother, the estimation sample used to estimate the effect of birthweight on adult outcomes may be selective, an issue I address directly in the main text.

Appendix Table 1 provides further details about the creation of the estimation sample (e.g., the dropping of observations). Each row in this table represents a different birth year. The columns represent different samples. For instance, in 1979, there were 380,271 births in California. Of those births, there were 7,407 twin births and 2,490 same sex female twin births.[71] My primary estimation sample consists of twins for whom I observe both twins having their first birth or both twins having their second birth (Selection 7). As the median number of children is two, this sample is likely reflective of the typical mother. Moreover, qualitatively, the results are the same when only using twins for whom I observe the first birth. The foreseeable advantage to including the second-birth mothers is the increase in statistical precision.

In addition to imposing the first and second birth restriction, I also limit the twins sample to women born between 1960 and 1982. As seen in Appendix Table 1, some of the 1960-1988 twins were born too recently to have had births yet. Since the mean age at first birth is twenty, the earlier cohorts in the estimation sample are not so selective.

Appendix Figure 1 provides a sense of the performance of the matching algorithm, which links the twin's own birth to that of her children. The solid line displays the percent of same-sex female births born in California having an observed first birth in California in the data. To indicate how large this percentage could be, the long dashed line represents the fraction of all women in California born between 1960 and 1988 who later gave birth in California. Here I am not performing the one-to-one matching I did for the twins. Instead, based on the California birth records for 1989-2002, I calculate the number of women who were born in each year and had births between 1989 and 2002. Then, I divide these counts by the number of females born in those years. The long-dashed line is this ratio by year of birth. Therefore, if the fertility patterns of twins exactly followed that of the overall female population and if matching were perfect, the solid line and the long-dashed line would coincide. These lines differ in most years, but imperfect matching likely explains the difference. The best guess at the match rate is 80 percent, which is consistent with the gap between the two curves. It does not explain, however, why the gap varies across years.

From the national Detailed Natality data, I also estimate the fraction of women who were born in California but gave birth outside of California. While this is not a trivial fraction, women born in California

[69] This selection criterion will delete all twins born on different days and those with different last names. However, the number of twins discarded is small.

[70] Some women born between 1960 and 1988 will have births before 1989. I could match these pre-1989 births to the birth certificates of their twin mothers born between 1960 and 1988. However, I use the twin's exact date of birth and her first and last name to match her to the births of her children, and mother's exact date of birth was first reported in 1989.

[71] In these calculations, each twin pair represents two births. However, the number of twin births is odd. Errors in the plurality variable, which I use to identify twin pairs, cause this discrepancy. When I create the twins' dataset, I confirm that both twins are present in the data. Otherwise, the twin pair (or lack thereof) is dropped.

usually give birth in California. To reiterate, while Appendix Figure 1 effectively summarizes how well the matching works, one should not be so concerned with these measures. The twin estimates will only be biased by selection to the extent that amongst twins, there is a differential probability of giving birth that is correlated with the twins' differences in birthweight. One cannot ascertain this from Appendix Figure 1.

Appendix B: A Test of Sample Selection Bias

There are several methods resting on different assumptions that one could use to test or quantify the effects of sample selection. Here I develop a "non-parametric" test for sample selection bias motivated by a simple model of sample selection. To understand this approach, first consider the traditional selection model applied to the fixed-effects model:

(1') $$s_{ij}^* = z_{ij}\theta + \varepsilon_{ij}$$

and

(2') $$y_{ij} = \begin{cases} bw_{ij}\beta + \gamma_i + u_{ij} & \text{if } s_{ij} = 1(s_{ij}^* > 0) = 1 \\ \text{unobserved} & \text{otherwise} \end{cases}$$

where s_{ij}^* is the selection index, z_{ij} is a vector of characteristics affecting sample selection, γ_i is the twin fixed effect. Therefore, differencing across twins,

(3') $$y_{i2} - y_{i1} = \begin{cases} (bw_{i2} - bw_{i1})\beta + u_{i2} - u_{i1} & \text{if } s_{i1} = 1 \text{ and } s_{i2} = 1 \\ \text{unobserved} & \text{otherwise} \end{cases}$$

Now suppose we assume the standard assumptions of the Heckman (1979) selection model including the joint bivariate normality of ε_{ij} and u_{ij}, then

(4') $$y_{i2} - y_{i1} = \begin{cases} (bw_{i2} - bw_{i1})\beta + \left(\dfrac{\phi(z_{i2}\theta)}{\Phi(z_{i2}\theta)} - \dfrac{\phi(z_{i1}\theta)}{\Phi(z_{i1}\theta)} \right)\rho\sigma + \xi_{i2} - \xi_{i1} & \text{if } s_{i1} = 1 \text{ and } s_{i2} = 1 \\ \text{unobserved} & \text{otherwise} \end{cases}$$

where ϕ is the standard normal pdf, Φ is the standard normal cdf, ρ is the correlation between u and ε, σ is the standard deviation of u, and ξ_{ij} is an idiosyncratic error term. If z_{ij} is twin-invariant (i.e., factors contributing to selection are the same for each twin), then sample selection is not an issue because the selection term is zero. However, it is highly-probable that z_{ij} varies within twin pairs. In particular, it might be a function of birthweight. If so, we might be inclined to include powers of birthweight (e.g., birthweight-squared) in the fixed-effects regressions. However, the simple equation above is not informative about how many powers of birthweight to include and moreover, the joint normality assumption makes the selection bias term additive separable whereas under other distributional assumptions, this would not be necessarily true.

The reasoning behind the non-parametric test becomes clearer if I rewrite equation (4') as follows (DiNardo et al. (2006)):

(5')

$$y_{i2} - y_{i1} = \begin{cases} (bw_{i2} - bw_{i1})\beta + \left(\dfrac{\phi(\Phi^{-1}(p_{i2}))}{p_{i2}} - \dfrac{\phi(\Phi^{-1}(p_{i1}))}{p_{i1}} \right)\rho\sigma + \xi_{i2} - \xi_{i1} & \text{if } s_{i1} = 1 \text{ and } s_{i2} = 1 \\ \text{unobserved} & \text{otherwise} \end{cases}$$

where p_{ij} is the fraction observed beyond birth (i.e., at time of motherhood).

A priori, we might believe that $\hat{\beta}$ is downward-biased. This is because ρ is likely negative and the inverse Mill's ratio, a decreasing function of the degree of sample selection, is probably positively correlated with birthweight. However, this intuition does not give us any sense of the magnitude of the sample selection bias. But equation (5') implies that for samples in which the twin birthweight difference strongly predicts the within twin-pair difference in the probability of later observation, the usual fixed-effects estimate will be the most biased. Therefore, because the degree of bias is related to the correlation between $\frac{\phi(\Phi^{-1}(p_{i2}))}{p_{i2}} - \frac{\phi(\Phi^{-1}(p_{i1}))}{p_{i1}}$ and $bw_{i2} - bw_{i1}$, if selection bias is severe, we should expect the fixed effects estimates to vary across groups classified by the degree of correlation between $\frac{\phi(\Phi^{-1}(p_{i2}))}{p_{i2}} - \frac{\phi(\Phi^{-1}(p_{i1}))}{p_{i1}}$ and $bw_{i2} - bw_{i1}$.

This intuition provides the basis for the "non-parametric" test. First, suppose I arrange the twin pairs into k groups, which I believe *a priori* may differ by the degree of correlation between $\frac{\phi(\Phi^{-1}(p_{i2}))}{p_{i2}} - \frac{\phi(\Phi^{-1}(p_{i1}))}{p_{i1}}$ and $bw_{i2} - bw_{i1}$. Then, I estimate the following regression:

(6') $$s_{i2} - s_{i1} = \sum_k \lambda_k D_i^k (bw_{i2} - bw_{i1}) + \psi_{i2} - \psi_{i1}$$

where D_i^k is an indicator for whether the twin pair i belongs to the kth group and Ψ_{ij} is an idiosyncratic error term. Then, I test whether $\lambda_k = \lambda_l \ \forall k,l$. If I reject this null hypothesis that the effect of birthweight on selection is the same across groups, then when I estimate the main regression equation but allow the effect of birthweight to differ across these k groups (i.e., estimate $y_{i2} - y_{i1} = \sum_k \eta_k D_i^k (bw_{i2} - bw_{i1}) + \phi_{i2} - \phi_{i1}$), I should be able to reject the hypothesis that $\eta_k = \eta_l \ \forall k,l$ in the presence of strong sample selection bias. If I am unable to reject this hypothesis, then I would suspect that although selection into the sample is related to within-twin-pair birthweight differences, such selection does not severely bias the fixed-effects estimates. I perform this test for all outcomes where each group is a birth-year cohort.

Appendix C: The Inconsistency of the Twin Fixed Effect Estimator Using Fetal Growth

While there are several appealing reasons to use fetal growth (birthweight/gestation) rather than birthweight to look at the long-run effects of infant health, measurement error in models using fetal growth as an independent variable is likely quite problematic. In particular, under classical measurement error, one can show that measurement error in fetal growth due to mismeasurement of gestational length can potentially lead to *upward*-biased estimates of the effect of fetal growth.

To demonstrate this, consider a simplified version of equation (3) where I replace the within-twin difference in birthweight by the within-twin difference in fetal growth[72]:

(1") $$y_{i2} - y_{i1} = \left[\frac{bw_{i2} - bw_{i1}}{g_i^*}\right]\theta + \varepsilon_{i2} - \varepsilon_{i1}$$

where g_i^* is the true gestational length of the twins, assumed to be identical for the twins. Equation (4") is the true model. Now suppose that birthweight is measured without error but gestational length is imprecisely measured. That is,

(2") $$g_i = g_i^* + \eta_i$$

where g_i is the mismeasured gestational length and the measurement error η_i is classical (i.e., uncorrelated with the true gestational length and with an expected value of 0). Then, the twin fixed effect estimate of the effect of fetal growth using the mismeasured gestational length will be:

[72] I am ignoring covariates in this equation as the estimating equation excluding covariates more closely resembles the main estimating equation.

(3”) $$\hat{\theta} = \frac{\sum \left[\frac{bw_{i2} - bw_{i1}}{g_i} \right] [y_{i2} - y_{i1}]}{\sum \left[\frac{bw_{i2} - bw_{i1}}{g_i} \right]^2}.$$

or, equivalently,

(4”) $$\hat{\theta} = \frac{\theta \sum \left[\frac{(bw_{i2} - bw_{i1})^2}{(g_i^* + \eta_i) g_i^*} \right] + \sum \left[\frac{bw_{i2} - bw_{i1}}{g_i^* + \eta_i} \right] [\varepsilon_{i2} - \varepsilon_{i1}]}{\sum \left[\frac{bw_{i2} - bw_{i1}}{g_i^* + \eta_i} \right]^2}$$

Taking the plim of both sides,

(5”) $$\operatorname{plim} \hat{\theta} = \theta \operatorname{plim} \frac{\frac{1}{N} \sum \left[\frac{(bw_{i2} - bw_{i1})^2}{(g_i^* + \eta_i) g_i^*} \right]}{\frac{1}{N} \sum \left[\frac{bw_{i2} - bw_{i1}}{g_i^* + \eta_i} \right]^2}$$

Therefore, only in the unusual case that

(6”) $$\operatorname{plim} \frac{1}{N} \sum \left[\frac{(bw_{i2} - bw_{i1})^2}{(g_i^* + \eta_i) g_i^*} \right] = \operatorname{plim} \frac{1}{N} \sum \left[\frac{bw_{i2} - bw_{i1}}{g_i^* + \eta_i} \right]^2$$

will the probability limit of $\hat{\theta}$ equal θ. Unlike the classical error-in-variables model, from the theory, the direction of bias is unclear. In particular, it is unclear whether $\operatorname{plim} \frac{1}{N} \sum \left[\frac{(bw_{i2} - bw_{i1})^2}{(g_i^* + \eta_i) g_i^*} \right]$ is larger or smaller than $\operatorname{plim} \frac{1}{N} \sum \left[\frac{bw_{i2} - bw_{i1}}{g_i^* + \eta_i} \right]^2$ without further assumptions.

References

Almond, Douglas (2006). "Is the 1918 Influenza Pandemic Over? Long-term Effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114(4): 672-712.

Almond, Douglas, Kenneth Chay, and David S. Lee (2005). "The Costs of Low Birth Weight." *Quarterly Journal of Economics* 120(3): 1031-1083.

Bajoria, Rekha, Suren R. Sooranna, Stuart Ward, Stephen D'Souza, and Maggie Hancock (2001). "Placental Transport Rather Than Maternal Concentration of Amino Acids Regulates Fetal Growth in Monochorionic Twins: Implications for Fetal Origin Hypothesis." *American Journal of Obstetrics and Gynecology* 1239-1246.

Barker, D.J.P., P.D. Winter, C. Osmond, B. Margetts, and S.J. Simmonds (1989). "Weight in Infancy and Death from Ischaemic Heart Disease." *Lancet* 2(8663):577-80.

Barker, D.J.P. (1995). "Fetal Origins of Coronary Heart Disease." *British Medical Journal* 311: 171-174.

Barker, David (2006). http://www.barkertheory.com/press.html, Accessed July 7, 2006.

Becker, Gary S.. (1960). "An Economic Analysis of Fertility." *Demographic and Economic Change in Developed Countries* Universities—National Bureau Conference Series 11, Princeton: Princeton University Press, pp. 209–240.

Becker, Gary S. and H. Gregg Lewis (1973). "On the Interaction Between the Quantity and Quality of Children." *Journal of Political Economy*, Part 2: New Economic Approaches to Fertility 81 (2): S279–S288.

Behrman, Jere R. and Mark R. Rosenzweig (2004). "Returns to Birthweight." *Review of Economics and Statistics* 86(2): 586-601.

Bernheim, B. Douglas and Sergei Severinov (2003). "Bequests as Signals: An Explanation for the Equal Division Puzzle." *Journal of Political Economy* 111(4):733-764.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2006). "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes" available at http://client.norc.org/jole/SOLEweb/black_devereux_salvanes.pdf , Accessed July 11, 2006.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2007). "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics* 122(1): 409-439.

Blanden, J, P. Gregg and S. Machin (2005). 'Educational Inequality and Intergenerational Mobility' in *What's the Good of Education? The Economics of Education in the United Kingdom,* eds. S. Machin and A. Vignoles, Princeton University Press.

Buckles, Kasey (2007). "Stopping the Biological Clock: Infertility Treatments and the Career-Family Tradeoff." Mimeo, University of Notre Dame.

Buescher PA, Taylor KP, Davis MH, Bowling JM. (2003). "The Quality of the New Birth Certificate Data: A Validation Study in North Carolina." *American Journal of Public Health* 3:1163–5.

Case, Anne, Angela Fertig, Christina Paxson (2005). "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24(2): 365–389.

Case, Anne, Darren Lubotsky, Christina Paxson (2002). "Socioeconomic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92 (5): 1308–1334.

Conley, Dalton and Neil G. Bennett (2000). "Is Biology Destiny? Birth Weight and Life Chances." *American Sociological Review* 65(3): 458-467.

Conley, Dalton and Kate W. Strully, and Neil G. Bennett (2003). *The Starting Gate: Birth Weight and Life Chances* Berkeley: University of California Press.

Conley, Dalton, Kate Strully, and Neil G. Bennett (2006). "Twin Differences in Birth Weight: The Effects of Genotype and Prenatal Environment on Neonatal and Post-Neonatal Mortality." *Economics & Human Biology* 4(2): 151-183.

Cunningham, F. Gary, Norman Gant, Kenneth Leveno, Larry Gilstrap III, John Hauth, and Katharine Wenstrom (2001). *Williams Obstetrics*. New York: McGraw-Hill.

Currie, Janet and Rosemary Hyson (1999). "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight." *American Economic Review* 89(2): 245-250.

Currie, Janet and Enrico Moretti (2003). "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics* 118(4): 1495-1532.

Currie, Janet and Enrico Moretti (2007). "Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight." *Journal of Labor Economics* 25(2): 231-264.

Currie, Janet and Mark Stabile (2003). "Socioeconomic Status and Child Health: Why is the Relationship Stronger for Older Children?" *American Economic Review* 93(5): 1813-1823.

Cutler, David and Ellen Meara (1999). "The Technology of Birth: Is it Worth it?" National Bureau of Economic Research Working Paper No. 7390.

Datar, Ashlesha, M. Rebecca Kilburn, and David S. Loughran (2006). "Health Endowments and Parental Investments in Infancy and Early Childhood." RAND Working Paper WR-367.

DiGiuseppe, David L., David C. Aron, Lorin Ranbom, Dwain L. Harper, and Gary E. Rosenthal (2002). "Reliability of Birth Certificate Data: A Multi-Hospital Comparison to Medical Records Information." *Maternal and Child Health Journal* 6(3): 169-179.

DiNardo, John, Justin McCrary, and Lisa Sanbonmatsu (2006). "Constructive Proposals for Dealing with Attrition: An Empirical Example." Mimeo, University of Michigan.

Dwyer, Terence, Leigh Blizzard, Ruth Morley, and Anne-Louise Ponsonby (1999). "Within Pair Association Between Birth Weight and Blood Pressure at Age 8 in Twins from a Cohort Study." *British Medical Journal* 319: 1325-1329.

Eriksson, J G, T Forsén, J Tuomilehto, P D Winter, C Osmond, and D J P Barker (1999). "Catch-up Growth in Childhood and Death from Coronary Heart Disease: Longitudinal Study." *British Medical Journal* 318: 427-431.

Gringras, Paul and Wai Chen (2001). "Mechanisms for Differences in Monozygous Twins." *Early Human Development* 64(2): 520-525.

Hales CN, DJP Barker, PMS Clark, LJ Cox, C Fall, C Osmond, and PD Winter (1991). "Fetal and Infant Growth and Impaired Glucose Tolerance at Age 64." *British Medical Journal* 303: 1019 - 1022 .

Hausman, Jerry, Jason Abrevaya, and F.M. Scott Morton (1998). "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87: 239-269.

Heckman, James (1979). "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153-161.

Huxley, Rachel, Andrew Neil, and Rory Collins (2002). "Unravelling the Fetal Origins Hypothesis: Is There Really an Inverse Association Between Birthweight and Subsequent Blood Pressure?" *Lancet* 360: 659-665.

IJzerman, Richard G., Dorret I. Boomsma, and Coen D. A. Stehouwer (2005). "Intrauterine Environmental and Genetic Influences on the Association Between Birthweight and Cardiovascular Risk Factors: Studies in Twins as a Means of Testing the Fetal Origins Hypothesis." *Paediatric and Perinatal Epidemiology* 19(Suppl. 1): 10-14.

James, Williams H. (1982). "The IQ Advantage of the Heavier Twin." *British Journal of Psychology* 73(Pt 4): 513-517.

Johnson, Rucker C. and Robert F. Schoeni (2005). "Early-Life Events and Health and Labor Market Outcomes in Adulthood." Mimeo, University of California-Berkeley and University of Michigan.

Loughran, David S., Ashlesha Datar, and M. Rebecca Kilburn (2004). "The Interactive Effect of Birth Weight and Parental Investment on Child Test Scores." RAND Labor and Population Working Paper #168.

Lumey, LH and Aryeh D Stein (1997). "In Utero Exposure to Famine and Subsequent Fertility: The Dutch Famine Birth Cohort Study." *American Journal of Public Health* 87(12): 1962-1966.

Loos, Ruth J.F., Robert Fagard, Gaston Beunen, Catherine Derom, and Robert Vlietinck (2001). "Birth Weight and Blood Pressure in Young Adults: A Prospective Twin Study." *Circulation* 104:1633-1638.

Lykken DT, TJ Bouchard Jr, M McGue, and A. Tellegen (1990). "The Minnesota Twin Family Registry: Some Initial Findings." *Acta Genetica Medicae et Gemellologiae (Roma)* 39(1): 35–70.

Maccini, Sharon and Dean Yang (2006). "Under the Weather: Health, Schooling, and Socioeconomic Consequences of Early-Life Rainfall," Mimeo, University of Michigan.

McCrary, Justin and Heather Royer (2006). "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth." National Bureau of Economic Research Working Paper No. 12329.

Meng, Xin and Nancy Qian (2006). "The Long Run Health and Economic Consequences of Famine on Survivors: Evidence from China's Great Famine." Mimeo, Australian National University and Brown University.

Mincer, Jacob (1963). "Market Prices, Opportunity Costs, and Income Effects." in C. Christ., ed. *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehund Grunfeld* Stanford: Stanford University Press.

Oreopoulos, Phil, Mark Stabile, Randy Walld, and Leslie Roos (2006). "Short, Medium, and Long Term Consequences of Poor Infant Health: An Analysis using Siblings and Twins." National Bureau of Economic Research Working Paper No. 11998.

Ozanne, Susan E. and C. Nicholas Hales (2002). "Early Programming of Glucose-Insulin Metabolism." *Trends in Endocrinology & Metabolism* 13(9): 368-373.

Poulter, N R, C L Chang, A J Mac Gregor, H Snieder, and T D Spector (1999). "Association Between Birth Weight and Adult Blood Pressure in Twins: Historical Cohort Study." *British Medical Journal* 319: 1330-1333.

Reid, Alice (2005). "The Effects of the 1918–1919 Influenza Pandemic on Infant and Child Health in Derbyshire." *Medical History* 49: 29–54.

Rosenzweig, Mark R. and Wolpin, Kenneth I. (1995). "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes." *Econometrica* 63(2): 303-326.

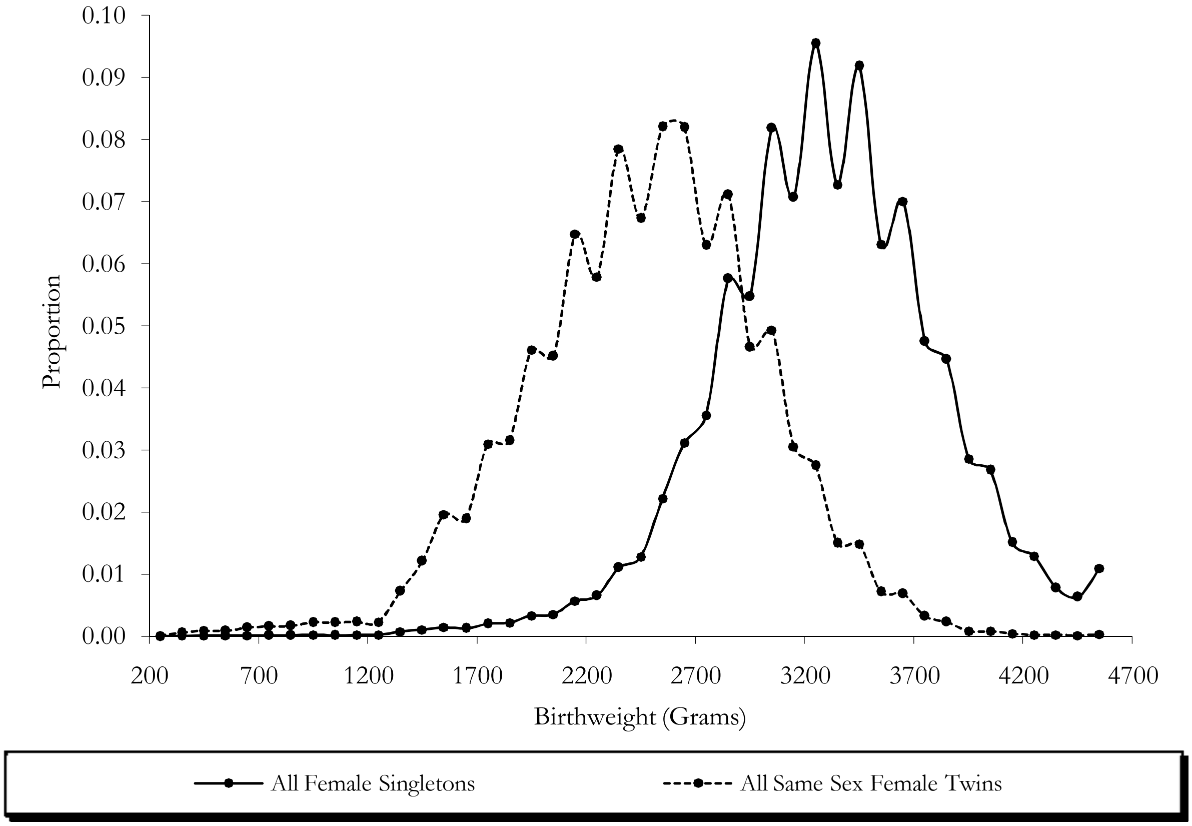
Rosenzweig, Mark R. and Zhang, Junsen (2006). "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One Child' Policy." Mimeo, Yale University and Chinese University of Hong Kong.

Royer, Heather (2004). "What All Women (and Some Men) Want to Know: Does Maternal Age Affect Infant Health?" University of California-Berkeley Center for Labor Economics Working Paper No. 68.

Victoria, Alejandro, Gerardo Mora, and Fernando Arias (2001). "Perinatal Outcome, Placental Pathology, and Severity of Discordance in Monochorionic and Dichorionic Twins." *Obstetrics and Gynecology* 97(2): 310-315.

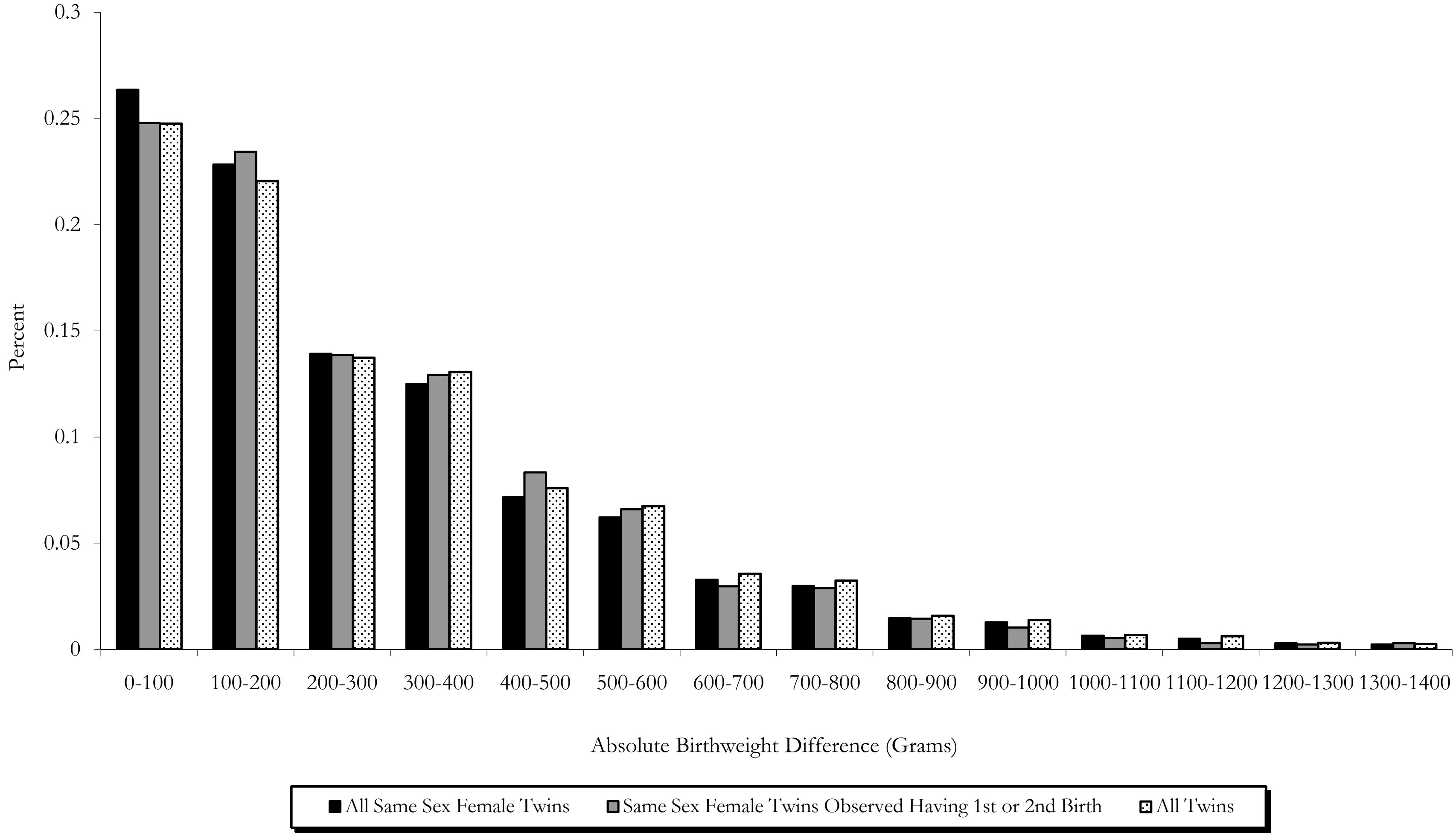
Zhang, Jun, Ruth A. Brenner, and Mark A. Klebanoff (2001). "Differences in Birth Weight and Blood Pressure at Age 7 among Twins." *American Journal of Epidemiology* 153(8): 779-782.

Figure 1 - Distribution of Birthweight - Female Singletons and Female Twins



Notes: The sample includes females born between 1960 and 1982 in California unconditional on whether they had a later observed birth.

Figure 2 - Distribution of Absolute Birthweight Differences Between Twins

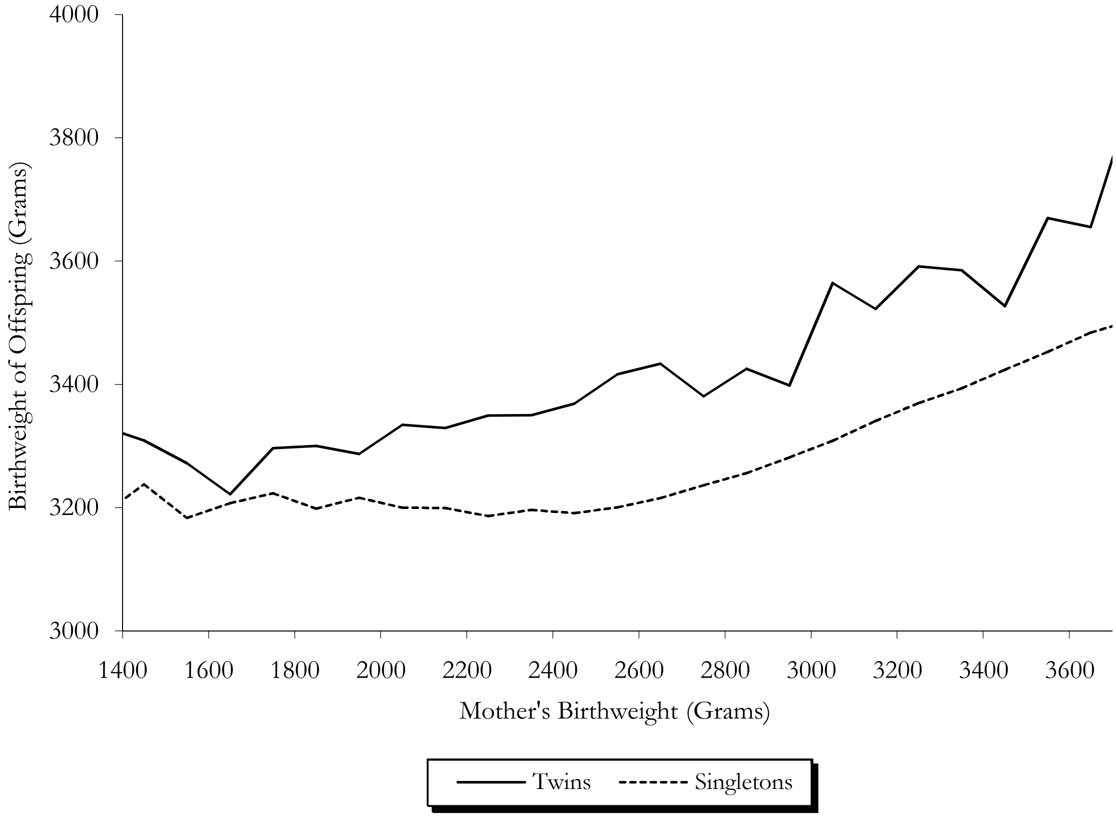


Notes: The sample includes twins born between 1960 and 1982 in California. The all twins sample includes twins of both genders, including mixed-sex twin pairs.

Figure 3 - Cross-Sectional Relationship Between Birthweight and Education

Notes: This figure is a plot of education averages by birthweight where birthweight is categorized into 100 grams intervals starting with the interval 1400-1500 grams. The sample used in this figure includes those mothers observed having a first or second birth whose birthweight was between 1400 and 3700 grams. Within the twins sample, there are only 42 observations with birthweights falling below 1400 grams and 59 observations with birthweights exceeding 3700 grams. In the case of the twins sample, both twin mothers must be observed for their first or second births. If they are observed for both births, then only their second birth is used in the calculations. The education measure is education reported at time of motherhood.

Figure 4 - Cross-Sectional Relationship Between Own Birthweight and Offspring's Birthweight



Notes: This figure is a plot of birthweight of offspring averages by own birthweight where own birthweight is categorized into 100 grams intervals starting with the interval 1400-1500 grams. The sample used in this figure includes those mothers observed having a first or second birth whose birthweight was between 1400 and 3700 grams. Within the twins sample, there are only 42 observations with birthweights falling below 1400 grams and 59 observations with birthweights exceeding 3700 grams. In the case of the twins sample, both twin mothers must be observed for their first or second births. If they are observed for both births, then only their second birth is used in the calculations.

Figure 5 - Cross-Sectional Relationship Between Birthweight and Pregnancy Complications

Notes: This figure is a plot of the average number of pregnancy complications by own birthweight where own birthweight is categorized into 100 grams intervals starting with the interval 1400-1500 grams. The sample used in this figure includes those mothers observed having a first or second birth whose birthweight was between 1400 and 3700 grams. Within the twins sample, there are only 42 observations with birthweights falling below 1400 grams and 59 observations with birthweights exceeding 3700 grams. In the case of the twins sample, both twin mothers must be observed for their first or second births. If they are observed for both births, then only their second birth is used in the calculations.

Table 1: Descriptive Statistics as Measured at Birth (1960-1982 California Births)

	Unconditional on Later Observation			*Conditional on Later Observation*			
				Singleton Female Births:		Same Sex Female Twin Births:	
	Singleton Female Births	Twin Births	Same Sex Female Twin Births	1st Birth Observed	2nd Birth Observed	1st Birth Observed	2nd Birth Observed
Proportion white	0.85	0.84	0.84	0.86	0.87	0.84	0.87
Proportion black	0.09	0.11	0.11	0.10	0.10	0.12	0.10
Year of birth	1971.2	1971.2	1971.5	1971.5	1969.4	1971.8	1969.4
Parity	2.30	3.10	2.98	2.34	2.49	3.07	3.32
Proportion with birthweight < 1000 grams	0.0008	0.0086	0.0093	0.0002	0.0001	0.0005	0.0010
Proportion with birthweight < 1500 grams	0.003	0.033	0.035	0.002	0.001	0.014	0.015
Proportion with birthweight < 2500 grams	0.053	0.453	0.496	0.047	0.048	0.478	0.473
Birthweight (grams)	3298.1	2533.2	2475.1	3304.6	3296.9	2524.6	2533.2
	(508.0)	(556.8)	(538.5)	(492.1)	(489.6)	(485.9)	(482.5)
Proportion premature	0.07	0.33	0.33	0.07	0.07	0.30	0.30
Gestation (weeks)	39.6	37.3	37.4	39.7	39.7	37.6	37.7
	(2.8)	(3.4)	(3.5)	(2.7)	(2.6)	(3.0)	(3.0)
Fetal growth (grams/week)	83.7	68.0	66.3	83.7	83.6	67.2	67.2
	(30.0)	(21.3)	(18.6)	(27.6)	(28.2)	(11.6)	(11.5)
Proportion with congenital defects	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Observations	3,823,491	139,822	49,592	945,309	678,804	5,670	3,028

Notes: Table reports means and standard deviations (in parentheses). The twin births include twins of both genders, including mixed sex twins.

Table 2: Descriptive Statistics as Measured When Giving Birth (1960-1982 California Births)

	Conditional on Later Observation			
	Singleton Female Births:		Same Sex Female Twin Births:	
	1st Birth Observed	2nd Birth Observed	1st Birth Observed	2nd Birth Observed
Mother's maximum education	13.2	13.2	13.1	13.2
	(2.1)	(2.1)	(2.1)	(2.0)
Mother's mean education	13.0	12.9	12.9	12.9
	(2.1)	(2.1)	(2.1)	(2.0)
Mother's education at birth	12.9	13.0	12.8	12.9
	(2.3)	(2.1)	(2.3)	(2.1)
Mother's age	23.9	26.6	23.6	26.4
	(5.4)	(5.0)	(5.3)	(5.0)
Father's age	26.8	29.3	26.6	29.2
	(6.3)	(6.0)	(6.2)	(5.9)
Proportion with father present	0.95	0.97	0.95	0.97
Father's education at birth	12.7	12.9	12.7	13.0
	(2.8)	(2.5)	(2.8)	(2.5)
Birthweight of child (grams)	3338.2	3435.9	3363.3	3460.8
	(563.6)	(541.0)	(550.5)	(533.5)
Gestational length of child (weeks)	39.7	39.6	39.7	39.6
	(2.7)	(2.6)	(2.7)	(2.5)
Fetal growth of child (grams/week)	85.3	88.1	86.5	88.7
	(29.6)	(29.1)	(45.9)	(15.0)
Number of infant abnormalities	0.09	0.07	0.08	0.07
Proportion diabetic	0.01	0.02	0.01	0.02
Proportion hypertensive	0.03	0.01	0.03	0.02
Proportion anemic	0.01	0.01	0.01	0.01
Number of pregnancy complications	0.12	0.11	0.12	0.10
Number of labor complications	0.56	0.38	0.58	0.38
Proportion receiving neonatal intensive care	0.03	0.02	0.03	0.02
Poverty rate of residential zipcode	0.15	0.15	0.15	0.15
Median household income of residential zipcode in 1999	47844.2	47782.5	47420.0	47681.6
	(17283.3)	(17421.5)	(17664.6)	-17653.3
Observations	945,309	678,804	5,670	3,028

Notes: Table reports means and standard deviations (in parentheses). Father presence is ascertained from the presence of both his date of birth and educational attainment on the birth certificate data.

Table 3: Pooled OLS and Fixed Effect Estimates: Effects of Birthweight in Kilograms
Female Twins with First or Second Birth Observed

Infant Mortality		**Education**					
Death within First Year		Maximum Education		Mean Education		Education at Birth	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.18	-0.02	0.19	0.13	0.19	0.16	0.19	0.16
(0.004)	(0.004)	(0.05)	(0.08)	(0.05)	(0.07)	(0.06)	(0.08)
Mean: 0.0584		Mean: 13.09		Mean: 12.87		Mean: 12.87	

Birth and Adult Health Outcomes							
Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
177.87	70.42	2.42	0.87	-0.004	-0.02	-0.005	-0.002
(14.72)	(30.67)	(0.49)	(1.18)	(0.004)	(0.01)	(0.003)	(0.007)
Mean: 3399.89		Mean: 277.19		Mean: 0.02		Mean: 0.01	
Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
0.0004	-0.003	-0.01	-0.05	0.01	0.01	-0.004	0.00002
(0.0023)	(0.006)	(0.01)	(0.02)	(0.02)	(0.05)	(0.004)	(0.01)
Mean: 0.01		Mean: 0.11		Mean: 0.49		Mean: 0.02	

Birth Delivery				**Residential Location**			
C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.02	-0.03	-0.03	-0.04	1042.47	219.55	-0.003	-0.003
(0.01)	(0.02)	(0.01)	(0.02)	(470.53)	(770.99)	(0.002)	(0.004)
Mean: 0.24		Mean: 0.37		Mean: 47418.04		Mean: 0.15	

Maternal and Paternal Characteristics							
Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-5.83	75.56	0.01	-0.003	24.73	69.34	0.27	0.19
(33.92)	(66.28)	(0.01)	(0.012)	(53.79)	(110.39)	(0.07)	(0.14)
Mean: 9125.28		Mean: 0.96		Mean: 10283.04		Mean: 12.89	

Notes: Robust standard errors adjusted for within twin pair correlation are shown in parentheses. All regressions are based on the sample of twins observed having a first or second birth in California between 1989 and 2002 (3,396 twin pairs) with the exception of the infant mortality results. The infant mortality results are based on all same-sex female twins in the 1960, 1965-1980 California birth cohort files (18,628 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions. The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for year of birth, birth order, and race.

Table 4: Linear Spline Estimates of the Effect of Birthweight in Kilograms
Female Twins with First or Second Birth Observed

	Infant Mortality		**Education**					
	Death within First Year		Maximum Education		Mean Education		Education at Birth	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
<2500 g	-0.33	-0.044	0.03	-0.07	0.03	-0.02	-0.01	-0.10
	(0.01)	(0.006)	(0.11)	(0.13)	(0.11)	(0.13)	(0.11)	(0.14)
2500 g+	0.11	0.004	0.32	0.29	0.33	0.31	0.38	0.38
	(0.003)	(0.006)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)	(0.12)
F-stat of equal slopes	2591.02	23.15	2.68	3.25	2.86	2.97	4.37	5.13

	Birth and Adult Health Outcomes							
	Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
<2500 g	108.78	75.71	1.78	-0.17	-0.011	-0.035	-0.012	-0.015
	(29.48)	(54.04)	(1.00)	(2.09)	(0.009)	(0.017)	(0.007)	(0.013)
2500g+	239.51	66.10	2.98	1.71	0.002	-0.003	0.001	0.009
	(26.36)	(47.51)	(0.91)	(1.82)	(0.007)	(0.015)	(0.006)	(0.011)
F-stat of equal slopes	7.70	0.01	0.39	0.36	0.84	1.61	1.40	1.59

	Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
<2500 g	0.002	0.003	-0.05	-0.107	0.01	-0.02	-0.010	-0.002
	(0.004)	(0.010)	(0.02)	(0.040)	(0.04)	(0.09)	(0.009)	(0.017)
2500g+	-0.001	-0.008	0.02	-0.004	0.02	0.03	0.002	0.002
	(0.004)	(0.009)	(0.02)	(0.035)	(0.04)	(0.07)	(0.008)	(0.015)
F-stat of equal slopes	0.31	0.48	5.14	3.07	0.00	0.16	0.94	0.03

	Birth Delivery				**Residential Location**			
	C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
<2500 g	-0.04	-0.03	-0.02	-0.05	1985.79	2733.55	-0.004	-0.010
	(0.02)	(0.04)	(0.02)	(0.04)	(908.37)	(1353.61)	(0.005)	(0.007)
2500g+	-0.01	-0.03	-0.05	-0.03	192.31	-1833.57	-0.002	0.003
	(0.02)	(0.04)	(0.02)	(0.04)	(855.91)	(1191.54)	(0.004)	(0.006)
F-stat of equal slopes	0.84	0.00	0.43	0.17	1.42	5.10	0.05	1.52

	Maternal and Paternal Characteristics							
	Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
<2500 g	-32.63	-28.52	0.04	0.03	-117.06	-354.70	0.13	0.22
	(66.53)	(116.77)	(0.01)	(0.02)	(111.53)	(196.46)	(0.16)	(0.24)
2500g+	18.09	160.42	-0.01	-0.03	145.87	408.32	0.39	0.17
	(60.49)	(102.65)	(0.01)	(0.02)	(96.31)	(170.45)	(0.13)	(0.21)
F-stat of equal slopes	0.24	1.17	5.85	3.81	2.39	6.80	1.10	0.01

Notes: The point estimates represent the estimated slope within the relevant birthweight interval (e.g., 0-2500g and 2500g+). Robust standard errors adjusted for within twin pair correlation are reported in parentheses. The reported F-stat tests whether the two segments of the linear spline have equal slopes. All regressions are based on the sample of twins having first or second births in California between 1989 and 2002 (3,396 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions. The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for year of birth, birth order, and race.

Table 5: Probability of Observation of a Later Birth as a Function of Birthweight

Dependent Variable	At Least One Birth Observed		First Birth Observed		Second Birth Observed		First or Second Birth Observed	
	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
A. Linear Model								
Birthweight	0.025	0.025	0.022	0.025	0.017	0.013	0.025	0.028
	(0.004)	(0.008)	(0.004)	(0.008)	(0.003)	(0.007)	(0.004)	(0.008)
B. Linear Spline Model								
Birthweight segment:								
<1500 g	0.12	-0.10	0.13	-0.08	0.01	-0.04	0.12	-0.09
	(0.02)	(0.10)	(0.01)	(0.09)	(0.01)	(0.08)	(0.02)	(0.10)
1500-2500 g	0.05	0.03	0.04	0.02	0.04	0.02	0.05	0.02
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
2500-3000 g	-0.02	0.04	-0.01	0.05	-0.01	0.02	-0.02	0.04
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
3000g+	-0.02	0.002	-0.02	0.01	-0.01	-0.02	-0.01	0.02
	(0.02)	(0.026)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)

Notes: The coefficient estimates presented in this table represent the effect of a one kilogram increase in birthweight on the probability of observation. The probability of observation is the probability that the twin is observed giving birth in California between 1989 and 2002. Robust standard errors, adjusted for within twin pair correlation, are in parentheses. The estimation sample includes all same sex female twin pairs born in California between 1960 and 1982. All regressions are based on 49,592 twin observations. The pooled OLS regressions include controls for year of birth, birth order, and race.

Table 6: Pooled OLS and Fixed Effect Estimates: Effects of Birthweight in Kilograms
ECLS-B data (twins only)

Post-Birth Care				Developmental Outcomes			
NICU use		Days in Hospital		Standardized Mental Score		Standardized Motor Score	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.45	-0.04	-24.06	-3.89	0.18	0.12	0.35	0.17
(0.02)	(0.04)	(1.51)	(2.31)	(0.05)	(0.07)	(0.05)	(0.09)
Mean: 0.34		Mean: 11.48		Mean: -0.30		Mean: -0.29	

Notes: Robust standard errors adjusted for within twin pair correlation are shown in parentheses. All regressions are based on the sample of twins in the Early Childhood Longitudinal Study, Birth Cohort (856 twins). The motor and mental scores have been standardized using the mean and standard deviation for the entire sample (i.e., the sample that includes singletons). The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for birth order, race, and age at assessment.

Table 7: Comparison with Behrman and Rosenzweig (2004) Estimates

Descriptive Statistics

	This Study's Sample	B&R Sample	B&R Sample of First Born
Fetal Growth (oz per week)	2.36	2.34	2.34
Birthweight (oz)	89.08	90.2	90.1
Schooling	12.92	13.8	--------
Birthweight of 1st Child (oz)	118.87	--------	118.1
Observations	5,754	1,418	1,207

Regression Estimates

This Study

	Education at Birth			Offspring's Birthweight (oz)		
Variable	OLS	OLS	FE	OLS	OLS	FE
Fetal Growth (oz/week)	0.30	0.21	0.10	8.01	7.89	2.30
	(0.08)	(0.06)	(0.09)	(0.68)	(0.67)	(1.30)
Age		0.25			0.37	
		(0.01)			(0.05)	
Observations	5,604	5,604	5,604	5,604	5,604	5,604

Behrman & Rosenzweig (2004)

	Education		Offspring's Birthweight (oz)	
Variable	OLS	FE	OLS	FE
Fetal Growth (oz/week)	0.385	0.657	7.48	1.87
	(0.139)	(0.211)	(1.496)	(3.667)
Age	-0.0429	NA	-0.198	NA
	(0.012)		(0.105)	
Observations	1,418	804	1,207	608

Notes: Standard errors adjusted for within twin pair correlation are in parentheses. My estimation sample includes all mothers observed having first or second births between 1989 and 2002, who were themselves born in California between 1960 and 1982. The reported number of observations is the number of females: each twin pair constitutes two observations. For those twins who are observed giving first and second births, only the second birth is included in the regressions. For the birthweight regressions, the age variable in the Behrman and Rosenzweig regressions is mother's current age at survey. My measure of education is education at time of motherhood. Other measures of education - mean education across births and maximum education across births - produce similar results.

Table 8: Comparison with Black, Devereux, and Salvanes (2005) Estimates

Descriptive Statistics

	This Study's Sample	Black et al. Sample
Fetal Growth (g/week)	66.95	67.18
	(11.55)	(13.38)
Birthweight (g)	2526	2531
	(488)	(602)
Percent Completing High School	0.82	0.75
Observations	5,604	5,198

Regression Estimates

Dependent Variable: Death in First Year of Life

	This Study's Estimates		Black et al. Estimates	
Variable	OLS	FE	OLS*	FE*
Birthweight (kg)	-173.87	-15.88	-146.45	-1.99
	(4.6461)	(4.3222)	(6.35)	(5.31)
Ln(Birthweight)	-403.35	-51.27	-390.13	-9.54
	(7.7004)	(9.8151)	(13.42)	(12.98)
Fetal Growth (g/week)	-3.98	-0.68	-5.96	-0.06
	(1.34)	(0.16)	(0.28)	(0.20)
Observations	29,272	29,272	14,882	14,882

Dependent Variable: Completion of High School

	This Study's Estimates		Black et al. Estimates	
Variable	OLS	FE	OLS**	FE
Birthweight (kg)	0.02	0.02	0.03	0.05
	(0.01)	(0.02)	(0.01)	(0.02)
Ln(Birthweight)	0.04	0.05	0.08	0.13
	(0.03)	(0.05)	(0.02)	(0.05)
Fetal Growth (g/week)	0.001	0.001	0.002	0.002
	(0.0005)	(0.0007)	(0.0003)	(0.0008)
Observations	5,604	5,604	13,472	4,762

Notes: In the top descriptive statistics panel, standard deviations are in parentheses. In the bottom regression estimates panel, standard errors adjusted for within twin pair correlation are in parentheses. * denotes the sample of 1967-1981 twins and ** denotes that the estimation sample includes all twins, not just same sex female twins. With the exception of the infant mortality regressions, the estimates from this study are based on the sample of twins whose first or second births are observed in California between 1989 and 2002 (2,802 twin pairs). The infant mortality estimates come from the sample of same sex female twins born in California in 1960 or between 1965 and 1980. For those twins whose first and second births are both observed, only the second birth is included in the regressions. In these regressions, the education measure is education as measured at time of motherhood. A mother is considered to have completed high school if she reports 12 or more years of education.

Table 9: Comparison with Oreopoulos, Stabile, Walld, and Roos (2006) Estimates

Dependent Variable: Completion of High School

Linear Model

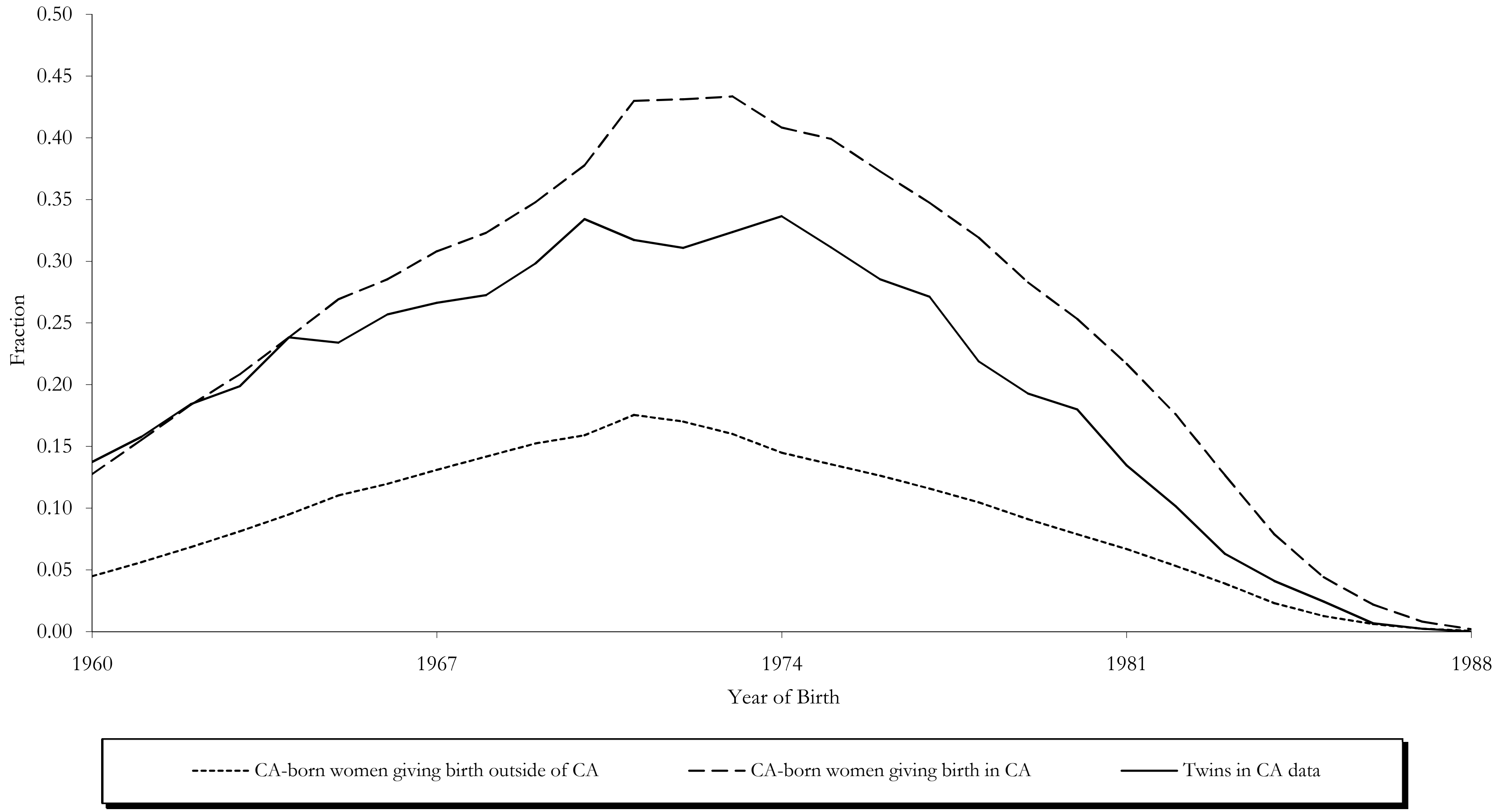
	This Study's Estimates		Oreopoulos et al. Estimates	
Variable	OLS	FE	OLS	FE
Birthweight (kg)	0.02	0.03	0.0741	0.0271
	(0.01)	(0.02)	(0.0214)	(0.0319)

Nonlinear Model

	This Study's Estimates		Oreopoulos et al. Estimates	
Variable	OLS	FE	OLS	FE
Birthweight < 1000g	0.11	0.02	-0.3346	-0.0837
	(0.21)	(0.41)	(0.1968)	(0.3434)
Birthweight 1000-1500g	-0.10	-0.10	-0.2708	-0.2315
	(0.05)	(0.07)	(0.0874)	(0.0954)
Birthweight 1500-2500g	-0.01	0.02	-0.1284	-0.0685
	(0.03)	(0.04)	(0.0621)	(0.0607)
Birthweight 2500-3000g	0.005	0.03	-0.1094	-0.0921
	(0.034)	(0.04)	(0.0624)	(0.0570)
Birthweight 3000-3500g	0.002	0.02	-0.0639	-0.0488
	(0.035)	(0.04)	(0.0656)	(0.0533)
Observations	6642	6642	1354	1354

Notes: This study's estimates are based on the sample of twins whose first or second births are observed in California between 1989 and 2002. For those twins whose first and second births are both observed, only the second birth is included in the regressions. In these regressions, the education measure is education as measured at time of motherhood. A mother is considered to have completed high school if she reports 12 or more years of education. The Oreopoulos et al. regressions are based on a sample of Canadian twins, which includes male/male twins, male/female twins, and female/female twins. The dependent variable is an indicator for whether the twin reached grade 12 by age 17.

Appendix Figure 1: Fraction of Women Having a First Birth Between 1989 and 2002 by Their Own Year of Birth



Notes: The fraction of CA-born women having a 1st birth between 1989 and 2001 outside of CA is calculated using the national Natality Detail Files. The other fractions are computed using the California Birth Statistical Master Files.

Appendix Figure 2: Fraction of Births that are Twins By Year

Notes: The sample includes all California births.

Appendix Table 1: Selection of the Twins Sample

Year of Birth	Births	All Twins Births	Selection 1	Selection 2	Selection 3	Selection 4	Selection 5	Selection 6	Selection 7
				Number of Observations					
1960	372,994	7,616	6,761	6,430	2,184	970	296	104	158
1961	382,420	7,722	6,731	6,442	2,130	996	298	122	186
1962	379,562	7,615	6,768	6,470	2,172	1,120	368	140	210
1963	381,756	7,745	6,781	6,454	2,162	1,094	418	186	270
1964	375,762	7,335	6,506	6,214	2,098	1,192	472	214	312
1965	356,268	7,339	6,469	6,060	2,038	1,110	402	180	256
1966	338,915	7,046	6,246	5,658	2,004	1,082	464	216	332
1967	337,952	7,397	6,624	5,776	1,944	1,044	458	242	330
1968	340,507	6,946	6,224	5,856	2,048	1,092	440	240	320
1969	354,069	7,060	6,351	6,042	2,076	1,120	474	282	386
1970	363,799	6,998	6,329	6,042	2,122	1,154	542	402	468
1971	330,796	6,084	5,511	5,262	1,834	926	414	326	374
1972	307,400	5,617	5,137	4,906	1,798	882	378	304	342
1973	298,874	5,526	5,048	4,806	1,736	840	396	326	366
1974	312,802	5,964	5,456	5,180	1,952	980	424	370	400
1975	318,390	6,051	5,501	5,242	1,924	880	370	340	354
1976	333,130	6,364	5,812	5,516	2,068	848	392	352	370
1977	348,467	6,441	5,936	5,652	2,012	838	300	268	276
1978	357,135	6,919	6,722	6,414	2,416	802	306	276	288
1979	380,271	7,407	7,231	6,914	2,490	724	288	264	270
1980	403,980	7,668	7,497	7,194	2,712	762	242	230	236
1981	421,771	7,925	7,754	7,462	2,704	560	188	178	178
1982	430,905	8,211	8,126	7,830	2,968	498	112	108	110
1983	436,984	8,532	8,440	8,082	3,072	324	68	62	64
1984	448,694	8,766	8,630	8,300	2,832	200	36	36	36
1985	472,190	9,338	9,214	8,868	3,300	146	18	16	16
1986	483,381	9,893	9,780	9,436	3,360	42	4	4	4
1987	504,853	10,406	10,289	9,900	3,590	16	0	0	0
1988	534,174	10,908	10,806	10,398	3,782	0	0	0	0
Total	11,108,201	218,839	204,680	194,806	69,528	22,242	8,568	5,788	6,912

Selection criteria:
Selection 1: Non-missing name, non-missing date of birth, non-missing birthweight
Selection 2: Selection 1 + Twins with only two births per unique date of birth and last name combination + twins with each twin with a unique date of birth, first and last name
Selection 3: Selection 2 + Same sex female
Selection 4: Selection 3 + Observe at least one twin giving birth
Selection 5: Selection 3 + Observe both twins giving birth
Selection 6: Selection 3 + Observe first birth for both twins
Selection 7: Selection 3 + Observe first or second birth for both twins

Appendix Table 2: 2003 American Community Survey Comparison

Descriptive Statistics - Fraction of Female Population with Certain Characteristics

	Born in CA between 1960 and 1982		*Born outside of CA between 1960 and 1982*
	Meets twin sample criteria	Does not meet twin sample	
Less than high school degree	0.11	0.07	0.09
High school degree	0.29	0.19	0.28
Some college	0.40	0.38	0.35
College	0.15	0.26	0.21
More than college	0.05	0.10	0.07
Married	0.68	0.38	0.54
Separated, divorced, or widowed	0.16	0.13	0.15
Never married	0.16	0.50	0.31
Employed	0.92	0.92	0.92
Not in labor force	0.69	0.78	0.77
Disability	0.04	0.05	0.04
Mobility limitation	0.02	0.03	0.02
Personal care limitation	0.01	0.02	0.01
Physical difficulty	0.05	0.05	0.04
White	0.72	0.74	0.79
Number of observations	4,020	7,691	144,377

Notes: To meet the twin criteria, a woman must (a) have given birth between 1989 and 2003 and still be living with that child and (2) have been living in her state of birth.

Appendix Table 3: Pooled OLS and Fixed Effect Estimates: Effects of Birthweight in Kilograms
Female Twins with First Birth Observed

Education

Maximum Education		Mean Education		Education at First Birth	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
0.10	0.17	0.10	0.21	0.11	0.21
(0.06)	(0.08)	(0.06)	(0.08)	(0.07)	(0.09)
Mean: 13.12		Mean: 12.89		Mean: 12.77	

Birth and Adult Health Outcomes

Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
174.25	60.48	2.44	0.99	-0.002	-0.02	-0.003	-0.01
(17.50)	(33.43)	(0.57)	(1.34)	(0.005)	(0.01)	(0.003)	(0.01)
Mean: 3363.34		Mean: 277.71		Mean: 0.03		Mean: 0.01	

Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.002	-0.01	-0.02	-0.08	-0.01	-0.04	-0.01	-0.01
(0.003)	(0.01)	(0.01)	(0.03)	(0.03)	(0.06)	(0.01)	(0.01)
Mean: 0.01		Mean: 0.12		Mean: 0.58		0.0259	

Birth Delivery / **Residential Location**

C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.03	-0.05	-0.04	0.0003	1077.21	-542.85	-0.003	0.001
(0.01)	(0.03)	(0.01)	(0.0245)	(579.88)	(844.84)	(0.003)	(0.004)
Mean: 0.24		Mean: 0.39		Mean: 47363.1		Mean: 0.15	

Maternal and Paternal Characteristics

Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-47.09	75.29	0.02	0.003	-70.23	72.78	0.14	0.24
(40.75)	(68.71)	(0.01)	(0.014)	(63.94)	(120.07)	(0.09)	(0.16)
Mean: 8621.99		Mean: 0.95		Mean: 9785.37		Mean: 12.80	

Notes: Robust standard errors adjusted fro within twin pair correlation are shown in parentheses. All regressions are based on the sample of twins observed having a first birth in California between 1989 and 2002 (2,835 twin pairs). The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for year of birth, birth order, and race.

Appendix Table 4: Pooled OLS and Fixed Effect Estimates: Effects of Birthweight in Kilograms
Female Twins Born Between with First or Second Birth Observed
Sample with Congenital Anomaly Information

Education

	Maximum Education		Mean Education		Education at Birth	
	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
	0.14	0.13	0.15	0.16	0.18	0.22
	(0.09)	(0.11)	(0.08)	(0.11)	(0.09)	(0.11)
	Mean: 13.4		Mean: 13.19		Mean: 13.2	

Birth and Adult Health Outcomes

Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
188.88	50.75	3.58	0.48	-0.005	-0.03	-0.01	-0.01
(22.73)	(44.32)	(0.73)	(1.60)	(0.007)	(0.01)	(0.01)	(0.01)
Mean: 3411.03		Mean: 276.87		Mean: 0.03		Mean: 0.02	

Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.002	0.01	-0.02	-0.05	0.02	0.02	-0.01	0.01
(0.002)	(0.01)	(0.02)	(0.03)	(0.03)	(0.07)	(0.01)	(0.01)
Mean: 0.01		Mean: 0.12		Mean: 0.5		Mean: 0.03	

Birth Delivery / **Residential Location**

C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.03	0.02	-0.04	-0.01	1757.61	1131.53	-0.005	-0.01
(0.02)	(0.04)	(0.02)	(0.03)	(800.87)	(1176.89)	(0.004)	(0.01)
Mean: 0.26		Mean: 0.31		Mean: 49682.62		Mean: 0.14	

Maternal and Paternal Characteristics

Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-15.83	63.66	0.03	0.01	-65.11	98.26	0.18	0.46
(48.99)	(94.10)	(0.01)	(0.02)	(78.05)	(157.75)	(0.10)	(0.18)
Mean: 9710.46		Mean: 0.95		Mean: 10918.62		Mean: 13.37	

Notes: Robust standard errors adjusted for within twin pair correlations are shown in parentheses. All regressions are based on the sample of twins born between either 1960 and 1967 or 1978 and 1982 and observed having a first or second birth in California between 1989 and 2002 (1,568 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions. The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for year of birth, birth order, and race.

Appendix Table 5: Pooled OLS and Fixed Effect Estimates: Effects of Birthweight in Kilograms
Female Twins with First or Second Birth Observed
Sample with No Congenital Anomalies

Education

Maximum Education		Mean Education		Education at Birth	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
0.13	0.15	0.14	0.19	0.17	0.25
(0.09)	(0.11)	(0.08)	(0.11)	(0.09)	(0.12)
Mean: 13.41		Mean: 13.2		Mean: 13.21	

Birth and Adult Health Outcomes

Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
190.76	67.82	3.36	0.66	-0.01	-0.03	-0.01	-0.02
(23.21)	(45.09)	(0.73)	(1.58)	(0.01)	(0.01)	(0.01)	(0.01)
Mean: 3411.75		Mean: 277.01		Mean: 0.03		Mean: 0.02	

Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.002	0.01	-0.02	-0.05	0.03	0.02	-0.01	0.02
(0.002)	(0.01)	(0.02)	(0.03)	(0.03)	(0.07)	(0.01)	(0.01)
Mean: 0.01		Mean: 0.12		Mean: 0.50		Mean: 0.03	

Birth Delivery				Residential Location			
C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.02	0.02	-0.04	-0.005	2127.37	1220.59	-0.01	-0.01
(0.02)	(0.04)	(0.02)	(0.030)	(811.06)	(1210.32)	(0.004)	(0.01)
Mean: 0.26		Mean: 0.30		Mean: 49794.08		Mean: 0.14	

Maternal and Paternal Characteristics

Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-16.38	74.26	0.03	0.01	-59.37	85.33	0.18	0.47
(49.99)	(96.64)	(0.01)	(0.02)	(79.84)	(161.67)	(0.10)	(0.19)
Mean: 9738.61		Mean: 0.95		Mean: 10940.85		Mean: 13.38	

Notes: Robust standard errors adjusted for within twin pair correlation are shown in parentheses. All regressions are based on the sample of twin pairs without congenital anomalies who were born between either 1960 and 1967 or 1978 and 1982, observed having a first or second birth in California between 1989 and 2002 (1,530 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions. The reported mean is the mean of the dependent variable. The pooled OLS regressions include controls for year of birth, birth order, and race.

Appendix Table 6: Fixed Effect Linear Spline Estimates of the Effect of Birthweight in Kilograms
Female Twins with First or Second Birth Observed

	Infant Mortality	**Education**		
	Death within First Year	Maximum Education	Mean Education	Education at Birth
<1500 g	-0.23	-0.02	0.09	0.49
	(0.03)	(1.47)	(1.40)	(1.54)
1500-2500 g	-0.02	-0.07	-0.03	-0.10
	(0.01)	(0.14)	(0.13)	(0.15)
2500-3000 g	-0.01	0.29	0.31	0.30
	(0.01)	(0.18)	(0.17)	(0.19)
3000g+	0.02	0.30	0.30	0.49
	(0.01)	(0.23)	(0.22)	(0.24)
F-stat of equal slopes	27.69	1.08	0.99	1.86

	Birth and Adult Health Outcomes							
	Child's Birthweigh t (in	Gestational Length (in Days)	Hypertensic	Diabetes	Anemia	# of Pregnancy Complications	# of Labor Complications	Neonatal Intensive Care Unit Transfer
<1500 g	-517.45	3.11	-0.04	0.12	0.00	0.18	-0.06	-0.05
	(596.60)	(22.09)	(0.18)	(0.14)	(0.11)	(0.44)	(0.94)	(0.19)
1500-2500 g	92.04	-0.22	-0.03	-0.02	0.00	-0.12	-0.05	0.00
	(57.09)	(2.21)	(0.02)	(0.01)	(0.01)	(0.04)	(0.09)	(0.02)
2500-3000 g	50.34	1.59	-0.01	0.01	0.01	0.01	0.13	-0.02
	(72.89)	(2.78)	(0.02)	(0.02)	(0.01)	(0.05)	(0.11)	(0.02)
3000g+	81.74	1.93	0.01	0.00	-0.03	-0.02	-0.13	0.03
	(92.85)	(3.56)	(0.03)	(0.02)	(0.02)	(0.07)	(0.15)	(0.03)
F-stat of equal slopes	0.35	0.13	0.61	0.87	1.09	1.20	0.55	0.37

Appendix Table 6 Con't: Fixed Effect Linear Spline Estimates of the Effect of Birthweight in Kilograms

	Birth Delivery		Residential Location		Maternal and Paternal Characteristics			
	C-Section Delivery	Public Payment for Delivery	Median Household Income (1999) of	Poverty Rate (1999) of Zipcode	Maternal Age (in Days)	Father Present	Paternal Age (in Days)	Paternal Education
<1500 g	0.384	-0.98	-316.13	0.033	361.06	0.80	1598.14	-3.77
	(0.469)	(0.44)	(14690.10)	(0.073)	(1289.23)	(0.24)	(3479.60)	(4.31)
1500-2500 g	-0.047	-0.03	2609.66	-0.010	-21.88	0.02	-290.91	0.22
	(0.045)	(0.04)	(1431.88)	(0.007)	(123.37)	(0.02)	(207.16)	(0.25)
2500-3000 g	0.002	-0.04	-957.36	-0.001	91.04	-0.04	5.10	0.40
	(0.057)	(0.05)	(1825.39)	(0.009)	(157.50)	(0.03)	(261.17)	(0.32)
3000g+	-0.081	-0.02	-3159.56	0.009	265.60	-0.01	1005.01	-0.19
	(0.073)	(0.07)	(2349.80)	(0.012)	(200.63)	(0.04)	(332.24)	(0.41)
F-stat of equal slopes	0.45	1.54	1.86	0.74	0.54	4.79	3.83	0.64

Notes: The point estimates represent the estimated slope within the relevant birthweight interval (e.g., 0-2500g). Robust standard errors adjusted for within twin pair correlation are reported in parentheses. The reported F-stat tests whether the four segments of the linear spline have equal slopes. All regressions are based on the sample of twins having first or second births in California between 1989 and 2002 (3,396 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions.

Appendix Table 7: Pooled OLS and Fixed Effect Estimates: Effect of Low Birthweight
Female Twins with First or Second Birth Observed

Education

Maximum Education		Mean Education		Education at Birth	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.17	-0.14	-0.18	-0.13	-0.17	-0.12
(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)
Mean: 13.09		Mean: 12.87		Mean: 12.87	

Birth and Adult Health Outcomes

Child's Birthweight (in Grams)		Gestational Length (in Days)		Hypertension		Diabetes	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-132.78	-42.74	-2.18	-0.53	0.004	0.01	0.005	-1.360E-19
(14.14)	(23.87)	(0.48)	(0.92)	(0.004)	(0.01)	(0.003)	(0.005)
Mean: 3399.89		Mean: 277.19		Mean: 0.02		Mean: 0.01	

Anemia		# of Pregnancy Complications		# of Labor Complications		Neonatal Intensive Care Unit Transfer	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
-0.002	-0.004	0.01	0.02	-0.03	-0.05	0.002	-0.001
(0.002)	(0.004)	(0.01)	(0.02)	(0.02)	(0.04)	(0.004)	(0.008)
Mean: 0.01		Mean: 0.11		Mean: 0.49		Mean: 0.02	

Birth Delivery / **Residential Location**

C-Section Delivery		Public Payment for Delivery		Median Household Income (1999) of Zipcode		Poverty Rate (1999) of Zipcode	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
0.003	-0.02	0.02	0.03	-663.43	-416.93	0.001	0.003
(0.010)	(0.02)	(0.01)	(0.02)	(463.03)	(596.18)	(0.002)	(0.003)
Mean: 0.24		Mean: 0.37		Mean: 47418.04		Mean: 0.15	

Maternal and Paternal Characteristics

Maternal Age (in Days)		Father Present		Paternal Age (in Days)		Paternal Education	
Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
16.47	8.52	-0.01	0.003	3.49	37.76	-0.21	-0.07
(32.69)	(51.58)	(0.01)	(0.010)	(50.99)	(85.65)	(0.07)	(0.11)
Mean: 9125.28		Mean: 0.96		Mean: 10283.04		Mean: 12.89	

Notes: The number reported in the first row of each set of estimates is the low birthweight (birthweight < 2500 grams) coefficient. Robust standard errors adjusted for within twin pair corelation are shown in parentheses. The reported mean is the mean of the dependent variable. All regressions are based on the sample of twins and observed having a first or second birth in California between 1989 and 2002 (3,396 twin pairs). For those twins whose first and second births are both observed, only the second birth is included in the regressions. The pooled OLS regressions include controls for year of birth, birth order, and race.