

The Costs of Low Birth Weight*

Douglas Almond

Kenneth Y. Chay

David S. Lee

Columbia and NBER

UC Berkeley and NBER

UC Berkeley and NBER

May 2004

Abstract

Birth weight has emerged as the leading indicator of infant health and welfare and the central focus of infant health policy. This is because low birth weight (LBW) infants experience severe health and developmental difficulties that can impose enormous costs on society. But would the prevention of LBW generate equally sizable cost savings and health improvements? Estimates of the return to LBW-prevention from cross-sectional associations may be biased by omitted variables that cannot be influenced by policy, such as genetic factors. To address this, we compare the hospital costs, health at birth, and infant mortality rates between heavier and lighter infants from all twin pairs born in the United States. We also examine the effect of maternal smoking during pregnancy – the leading risk factor for LBW in the United States – on health among singleton births after controlling for detailed background characteristics. Both analyses imply substantially smaller effects of LBW than previously thought, suggesting two possibilities: 1) existing estimates overstate the true costs and consequences of LBW by at least a factor of four and by as much as a factor of 20; or 2) different LBW-preventing interventions have different health and cost consequences, implying that policy efforts that presume a single return to reducing LBW will necessarily be suboptimal.

* This is a revision of a paper that was circulated with the title “Does Low Birth Weight Matter? Evidence from the U.S. Population of Twin Births” (Center for Labor Economics Working Paper No. 53, September 2002). We thank David Card, Janet Currie, David Cutler, Bill Evans, Lorenz Goethe, Michael Grossman, Jon Gruber, Ted Joyce, Bob Kaestner, Larry Katz, Karen Norberg, Jack Porter, Doug Staiger, Paul Torelli, and participants of seminars at Berkeley, Illinois, Maryland, RAND, and the NBER Summer Institute in Children’s Studies for helpful comments.

I. Introduction

Infants born at low birth weight (LBW) – conventionally defined as a birth weight less than 2,500 grams – experience severe health and developmental difficulties that can impose substantial costs on society. For example, the expected costs of delivery and initial care of a baby weighing 1000 grams at birth can exceed \$100,000 (in year 2000 dollars), and the risk of death within one year of birth is over one-in-five. Even among babies weighing 2000-2100 grams, who have comparatively low mortality rates, an additional pound (454 grams) of weight is still associated with a \$10,000 difference in hospital charges for inpatient services. The open circles in Figure 1 illustrate these striking facts, plotting the associations between hospital costs at discharge and birth weight (Panels A and B) and between infant mortality and birth weight (Panel C) for singleton births in the United States.¹ Studies have also established a correlation between LBW and high blood pressure, cerebral palsy, deafness, blindness, asthma and lung disease among children, as well as with reduced IQ, test scores, and cognitive development.²

Not surprisingly, birth weight is the primary measure of a baby's health in most analyses of infant health and welfare in economic research. In some contexts, birth weight is viewed as the "output" in the study of infant health production functions and the maternal behaviors that impact infant well-being (Rosenzweig and Schultz 1983; Corman, et al. 1987; Grossman and Joyce 1990; Geronimus and Korenman 1992; Rosenzweig and Wolpin 1991; Rosenzweig and Wolpin 1995; Currie and Moretti 2003). In other contexts, birth weight is an "input" – i.e., a proxy for the initial endowment of an infant's "health human capital". Consistent with this view, research has found that LBW infants tend to have lower educational attainment, poorer self-reported health status, and reduced employment and earnings as adults, relative to their normal weight counterparts (Behrman, Rosenzweig, and Taubman 1994; Currie and Hyson 1999; Behrman and Rosenzweig 2001). Finally, birth weight has been used to evaluate the effectiveness of social policy. Research on the benefits of large-scale social programs – including welfare

¹ We discuss the samples and estimation methods underlying Figure 1 below.

² For example, see Kaelber and Pugh (1969), McCormick, et al. (1992), Paneth (1995), Nelson and Grether (1997), Lucas, Morley, and Cole (1998), Brooks, et al. (2001), Matte, et al. (2001), and Richards, et al. (2001).

and health insurance for the poor – typically use birth weight as the primary indicator of infant welfare (Currie and Cole 1993; Currie and Gruber 1996; Hanratty 1996).

Perhaps because of its central role in research on infant health and welfare, LBW is also the direct target of health policy, both in the United States and abroad. In the U.S., reducing the incidence of LBW has been a stated goal of several social programs targeting infant health, including Medicaid, and publicly-funded nutrition programs such as the Women, Infants, and Children (WIC) program.³ The National Institutes of Health has encouraged research to focus on the prevention of LBW births.⁴ The World Bank’s \$100 million “Second Tamil Nadu Integrated Nutrition Project” in India had “cut[ting] in half the incidence of low birth weight newborns” as a goal.

The logic behind focusing on LBW as a central target of policy is illustrated by an important health intervention effort in the United States – the prevention of cigarette smoking during pregnancy. Babies born to mothers who smoke during pregnancy are on average 9-10 ounces lighter than those born to non-smoking mothers. Indeed, maternal smoking has been identified as the most significant, modifiable risk factor for LBW incidence in developed countries (Kramer 1987): in the United States, smoking mothers are over two times more likely to have a LBW baby than non-smoking mothers. The strong cross-sectional association between LBW and infant mortality has led to the reasoning that the prevention of all maternal smoking is an “optimal public health outcome”, since “maternal smoking is a significant risk factor for LBW infants, which, in turn, influences infant mortality” (Vogler and Kozlowski 2002). Similarly, the correlation between LBW and the hospital costs of birth (e.g. open circles in Panels A and B of Figure 1) has been used to calculate the sizeable cost savings of interventions that encourage smoking cessation (Lightwood, et al. 1999).

³ In the United States, a motivation for the Medicaid expansion to pregnant women during the 1980s was the reduction of the incidence of low birth weight through expanded access to prenatal care (Currie and Gruber, 1996). See Institute of Medicine (1996) for the LBW motivation for WIC and Kowalseki-Jones and Duncan (2002) for an evaluation of the birth weight benefits of the Special Supplemental Nutrition Program of WIC.

⁴ For example, the chief goal of a program announcement for NIH-sponsored research (PA-99-045) is the “development of innovative strategies to prevent LBW in minority populations.” In addition, the director of the NICHD, has stated “[T]he increased black mortality from low birth weight is not due to weaker infants or poorer care. It is simply because there are so many more low birth weight and particularly very low birth weight African-American births. So the key is to try to prevent these low birth weight births.”

There is, however, an important caveat to these kinds of policy conclusions.⁵ The strong correlation between birth weight and costs (or between birth weight and mortality) may partially reflect the influence of unobserved variables. For example, some mothers could be genetically predisposed to giving birth to unhealthy babies who – *because* of this inherent disadvantage – fail to attain normal birth weights. For these babies, it may be possible to prevent LBW, but doing so will have no impact on mortality risk, if such risk were genetically determined. More generally, for cost-benefit assessments of policy, estimates of the returns to reducing LBW should control for factors that cannot be influenced by policy, such as genetics and race.

This study provides new evidence on the early health and economic costs of LBW, attempting to control for these confounding factors.⁶ Our first approach uses the population of twins born in the United States between 1983 and 2000 to examine the correlation between twin differences in birth weight and twin differences in hospital costs, other measures of health at birth, and infant mortality rates. This strategy controls for all observed and unobserved characteristics of the mother. In the case of identical twins, it guarantees that the birth weight differences are entirely driven by environmental factors (e.g., nutritional intake within the uterus) rather than by any genetic factors.

In a complementary analysis, we consider a specific external influence of LBW that is 1) arguably less directly attributable to genetics, 2) most certainly not a consequence of the inherent health of the fetus, and 3) cited as the leading modifiable cause of LBW in the United States: maternal smoking during pregnancy. We estimate the effects of maternal smoking on a variety of infant health outcomes for the population of singleton births, controlling for the detailed background variables available in birth certificate data. The estimates are predicted to be large if, indeed, smoking causes LBW, and if LBW has a causal link to adverse infant health outcomes. Here, by attributing the entire impact of maternal smoking on infant health to its impact through LBW, we generate estimates of the effects of LBW that are

⁵ Cross-sectional correlations between birth weight and various childhood outcomes are often used to calculate the costs of LBW. For example, using this approach Lewit et al. (1995) calculate that in 1988 the health care, education, and child care costs associated with the 3.5-4 million children aged 0 to 15 born at low birth weight was \$5.5-6 billion more than if the children had been born at normal weight, and that LBW accounted for 10 percent of all health care costs for children.

⁶ More specifically, as we discuss below, we are assessing the costs and consequences of intrauterine growth retardation.

arguably overstated. We recognize that, in contrast to the twin analysis, this approach cannot fully control for unobserved heterogeneity across mothers. Nevertheless, the analysis allows us to gauge the plausibility of the magnitudes derived from the analysis of twins, within an important and policy-relevant context.

Both approaches yield implied impacts of LBW that are many times smaller than those typically used in cost-benefit calculations. First, we document the significant differences in birth weights between twins – an average difference that is larger than the estimated impact of maternal smoking on birth weight. There is also substantial variation in hospital costs and health outcomes within twin pairs. We find, however, that the birth weight difference is a relatively poor predictor of differences in costs or other health outcomes. Table 1 reports the coefficients from linear regressions of several different outcomes on birth weight for different samples of twin pairs. The cross-sectional OLS estimates in the first column indicate that a one standard deviation increase in birth weight (one pound and seven ounces) is associated with a 0.51 standard deviation decrease in hospital costs of about \$19,500 (in \$2000). When mother fixed effects are included, the costs estimate falls by a factor of 6, implying a \$3,200 decrease in hospital costs.

In addition, there is evidence to suggest that even these latter estimates – which account for mother-specific heterogeneity – are overstated, due to a potential association between detrimental genes and lower birth weights.⁷ For example, the coefficients fall further when we drop the small number of twin pairs in which one of the twins is born with a congenital anomaly or dies as a result of a congenital anomaly, as shown in the third column.⁸ The estimated effects of birth weight on the incidence of assisted ventilation, an initial health measure (the five-minute APGAR score), and infant mortality, fall by additional factors of 2 to 3. Overall, the conventional OLS estimates imply that a one standard deviation increase in birth weight reduces infant mortality, increases APGAR scores, and reduces ventilator use by 0.41, 0.51, and 0.25 standard deviations, respectively, while the corresponding magnitudes from the fixed effects specification (final column) vary between 0 and 0.03 standard deviations. Due to the large sample

⁷ As we discuss below, our data cannot differentiate between identical and fraternal twins. This leads to within-twin pair variation in genes, which may be correlated with birth weight.

⁸ Among the known causes of congenital abnormalities, only about 10 percent is believed to be due to environment only. The data do not allow us to perform this exercise with the hospital discharge data.

sizes, the fixed effects estimates are precise; indeed, several of the estimates are statistically different from zero, even though their magnitudes are 13 to 20 times smaller than in the cross-section. As we discuss below, this enormous reduction in magnitudes can lead to very different cost-benefit calculations and policy conclusions.

Furthermore, we document two other facts relevant for policy purposes. First, we show that the twins population is not so different from the singleton population with respect to the association between birth weight and infant health outcomes and costs. For example, even though twins are generally born at much lower birth weights, the cross-sectional relation between birth weight and health outcomes/costs for the two populations is very similar, as illustrated in Figure 1. Second, there are important nonlinearities in the relation between birth weight and infant outcomes, implying that interventions that raise birth weights will have very different impacts, depending on whether they raise weights for moderately or extremely low weight babies. For example, the fixed effects estimates in Figure 1C show that a one pound birth weight increase from 540 grams to 1000 grams reduces the probability of infant death by 0.15 (compared to the difference of 0.60 when mother effects are excluded). But the fixed effects (as well as the conventional OLS) estimates imply very little mortality impact of a one pound weight increase at birth weights above 1,500 grams. And even though there is a negligible influence of birth weight on mortality in the 4.5 to 5.5 pound range (2,000 to 2,500 grams), the fixed effects estimates in Figure 1B imply that such a one pound increase in weight results in a cost savings of \$2,000 (compared to the \$11,000 implied by the conventional OLS estimates).

We subject our findings to a battery of robustness checks and demonstrate that they are insensitive to the years of birth examined, functional form assumptions, and the scheme used to match twins. In addition, we test the assumption of twin exchangeability that underlies the fixed effects models. Finally, we document that the five-minute APGAR score, another continuous proxy for health of the infant at birth, is a significantly better predictor of within-twin pair differences in both infant mortality and assisted ventilation than birth weight.

The findings from analyzing the impact of maternal smoking on infant health among singletons are remarkably consistent with those from the twins analysis. We estimate a substantial impact of

smoking on birth weight (and on LBW incidence) after controlling for detailed background characteristics using both regression and propensity score methods. However, the corresponding estimates of the effects of smoking on the five-minute APGAR score, assisted ventilation use after birth, and infant mortality are essentially zero. In addition, using our estimates of the association between maternal smoking and birth weight and the association between birth weight and costs, we calculate a cost savings of \$53 per infant for every mother discouraged from smoking during pregnancy. By contrast, the cross-sectional estimates typically used in the literature imply a cost savings of over \$1,000 per infant.

In two different settings – corresponding to policies that impact fetal nutrient intake and maternal smoking – we find small effects of birth weight on other outcomes of interest. There are two plausible interpretations of these results. The first is that the analysis identifies the structural impact of birth weight and that the true health benefits and cost savings of preventing LBW are greatly exaggerated by cross-sectional correlations. A more cautious interpretation is that the cross-sectional correlations, within-mother twins analysis, and the covariate-adjusted smoking analysis each isolate different determinants of birth weight, and therefore that the benefits to reducing LBW are highly dependent on the type of intervention that influences birth weight. At a minimum, this would imply that LBW status is not a consistently dependable target of policy, if the ultimate objective is to minimize infant mortality, health problems, and consequent health care costs. That is, the most cost-effective ways of preventing LBW may have little correspondence with the most cost-effective ways of minimizing infant health problems and consequent health care costs.

The next section defines the parameter of interest and describes our estimation strategies. Section III describes the data and provides descriptive statistics. Section IV presents our main findings, while Section V implements our robustness checks. We discuss the policy implications of our results in Section VI and conclude with directions for future research in Section VII.

II. Conceptual Framework and the Etiology of LBW

This section discusses our framework for quantifying the costs and consequences of LBW. In particular, we define the parameter of interest and describe our identification strategies; discuss the

etiology of LBW; justify the outcome variables used in the study; and discuss functional form issues for estimating the impact of birth weight on these outcomes.

A. Parameter of Interest

Let

$$(1) \quad h_{ij} = \alpha + bw_{ij}\beta + X_i'\gamma + a_i + \varepsilon_{ij}$$

where h_{ij} is the underlying health of newborn j for mother i , bw_{ij} is birth weight, X_i is a vector of mother-specific observable determinants of health (e.g., race, age, education), a_i reflects mother-specific unobservable determinants of health (e.g., genetic factors), and ε_{ij} represents other newborn-specific idiosyncratic factors, assumed to be independent of all observable and unobservable factors.

The central parameter of interest is β , and its magnitude is important for policy purposes. If it is large and positive, it suggests substantial benefits to interventions that raise birth weights. Estimation of β by OLS, however, is confounded by the existence of factors such as race, age, and education (elements of X_i) or genetic determinants (a_i) that simultaneously influence birth weight and infant health. The omitted variables formula implies that the OLS coefficient from a regression of the health measure on birth weight is

$$(2) \quad \beta_{OLS} = \beta + \frac{cov(bw_{ij}, X_i'\gamma)}{var(bw_{ij})} + \frac{cov(bw_{ij}, a_i)}{var(bw_{ij})}$$

Thus, even if there is a strong cross-sectional correlation between health measures and birth weight – and β_{OLS} is highly significant – the strength of the relation could be driven by the correlation between birth weight and other factors such as race, age, education, and unobservable genetic factors.

The distinction between β and the latter two terms in (2) is important. Existing LBW-reducing intervention efforts in the U.S. (nutritional programs, smoking cessation) do not seek to alter the age or education levels of mothers, and no policy can ever affect immutable factors such as race or genetics. Thus, if β_{OLS} is primarily driven by the latter two terms in (2), then it would be an exaggerated and perhaps misleading estimate of the benefits (β) of a policy that raises birth weights.

B. Identification Strategies

We use two different strategies to estimate β . First, we examine the population of twin births in the U.S., and include mother fixed effects in the regression. Twins share the same mother, so the inclusion of fixed effects effectively controls for race, age, education, family background, behaviors (e.g., smoking), as well as genetic factors and other unobservable mother-specific factors. Including mother fixed effects is equivalent to estimating the first-differenced equation:

$$(3) \quad h_{i1} - h_{i2} = \alpha_1 - \alpha_2 + (bw_{i1} - bw_{i2})\beta + \varepsilon_{i1} - \varepsilon_{i2}$$

where ‘1’ and ‘2’ denote the first- and second-born infants of a twin-pair. Under the assumption that ε_{ij} is uncorrelated with bw_{ij} , the fixed effects estimator, β_{FE} , is unbiased for β . In the case of fraternal twins, genetic differences could imply that $(\varepsilon_{i1} - \varepsilon_{i2})$ is correlated with $(bw_{i1} - bw_{i2})$ leading to a β_{FE} that is biased. Below, we find evidence that birth weight is negatively correlated with congenital defects among twins, suggesting that β_{FE} overstates the impact of birth weight *per se*.⁹ Further, below we use the birth order of twins to derive tests of the assumption of the “perfect exchangeability” of twins that underlies the twins fixed effects estimator.

Second, to complement this “within-mother” approach, we conduct a “between-mother” analysis. Specifically, we attempt to isolate variation in birth weight that is a direct result of the behavior that has been identified as the leading cause of LBW in the United States: maternal smoking (Kramer 1987). It has been argued that since smoking induces LBW, and LBW causes infant mortality, then maternal smoking must have an influence on infant mortality (Vogler and Kozlowski, 2002). To test this reasoning, we estimate the direct effect of maternal smoking on infant mortality and on other outcomes and attribute this effect entirely to the impact of smoking on birth weight.

More formally, let

$$(4) \quad bw_{ij} = \lambda + SMOKE_i \delta + X_i' \psi + u_{ij}$$

⁹ Almond, Chay and Lee (2002) show the set of conditions under which the “twins estimate” should be strong and potentially larger than the cross-sectional “between-family” relation if birth weight is a valid policy marker.

where $SMOKE_i$ is an indicator for whether the mother smoked during pregnancy, u_{ij} reflects unobservable determinants of birth weight, and is by definition orthogonal to $SMOKE_i$ and X_i . The reduced-form then becomes

$$(5) \quad h_{ij} = \alpha^* + SMOKE_i \delta\beta + X_i' \gamma^* + a_i^* + u_{ij} \beta + \varepsilon_{ij}$$

where a_i^* is the residual from regressing a_i on X_i .¹⁰ Unbiased estimation of $\delta\beta$ with OLS requires $SMOKE_i$ to be uncorrelated with a_i^* .

There is no a priori justification for a_i^* being uncorrelated with $SMOKE_i$, and the analysis cannot control for unobservable differences across families. Equations (4) and (5) implicitly assume that smoking behavior during pregnancy is “randomized” conditional on the observable characteristics X_i . While experimental studies have demonstrated an effect of smoking cessation during pregnancy on birth weight and LBW, they have found no effects on gestation length and APGAR scores.¹¹ Further, the small numbers of women in the treatment and control groups did not allow for an analysis of the impact of smoking cessation on rare outcomes such as infant mortality. Thus, all of the evidence on the infant mortality effects of maternal smoking stems from purely observational studies.

In view of this, our approach is to estimate the direct effects of maternal smoking using a large sample of singleton births while controlling for more detailed maternal and background characteristics than previous studies. We also allow these characteristics to enter flexibly in both a regression and propensity score analysis. Although this may still result in biased inference, two factors mitigate potential concerns. First, the estimated birth weight effects are both robust and similar in magnitude to those documented in the experimental literature. Second, we attribute the entire infant health effect of maternal smoking to its impact on birth weight. Thus, the OLS estimate of $\delta\beta$ will overstate the importance of the birth weight channel if maternal smoking has a negative effect on infant health that is independent of its effect on LBW. This will also be true if the omitted variables, such as genetics, are correlated both with smoking and poorer birth outcomes.

¹⁰ γ^* is $(\gamma + \psi\beta + \phi)$ where ϕ are the linear projection coefficients from regressing a_i on X_i . α^* is $(\alpha + \lambda\beta)$.

¹¹ The quasi-experimental studies involve randomly assigned interventions in which the treatment group was encouraged to cease smoking through personal visits, information dissemination, telephone calls, etc. Some examples are Sexton and Hebel (1984), Ershoff et al. (1989), and Windsor et al. (1993).

C. Etiology and Sources of Birth Weight Variation

It is generally recognized that LBW is governed by two factors: a short duration of gestation (i.e., prematurity), and a reduced fetal growth rate at a fixed gestation length, also known as intrauterine growth retardation (IUGR). Research on the etiology of birth weight suggests that both environmental and genetic factors play a role in the rate of fetal growth in utero. In a widely-cited survey of research on the determinants of low birth weight, Kramer (1987) notes that most research focuses on the factors causing IUGR, as opposed to those causing prematurity.¹² In addition, interventions targeted at preventing premature birth, including enhanced prenatal care and nutritional interventions, have been found to be ineffective (Goldenberg and Rouse 1998). By contrast, the causes of IUGR are both well-known and well-established. This study focuses on the consequences of LBW resulting from IUGR.

The twins analysis described above seems ideal for isolating birth weight variation driven by IUGR. Gestational ages for twins are essentially identical, so differences in birth weight must necessarily be due to differential intrauterine growth rates. While twins must share the same supply of nutrition from the mother, this sharing can be unequal. For example, for identical twins sharing the same placenta, the differing insertion points of the twins' umbilical cords into the placenta are believed to affect nutritional intake of each fetus, and hence contribute to intra-pair differences in fetal growth (Bryan 1992).¹³ This type of variation corresponds to an intervention that affects intrauterine nutrient consumption of the fetus. Further, the similarity of the twins and singletons cross-sectional "response functions" plotted in Figures 1 and 2 suggest that twins may not be an unusual subpopulation with respect to their responses to birth weight changes. Indeed, if anything the pooled response function of twins has a slightly sharper gradient

¹² In developing countries, it is difficult to analyze prematurity since the gestational age is often unrecorded or unreliably measured (Kramer 1987).

¹³ The most commonly cited factors for twin birth weight differences are discrepancies in the microintrauterine environment due to different placentation and blood perfusion (Zhang, et al., 2001), different nutritional sources at different parts of the uterus, and different genetic growth potentials among dizygotic twins. Although very rare, fetal transfusion syndrome among identical twins, which involves the transfer of blood from one fetus directly to the other, can cause substantial birth weight differences (also known as discordance). In this case, both the lighter and heavier twins are at greater risk of death. Some of this follows from a conversation with Dr. Louis Keith in the Obstetrics and Gynecology Department of Northwestern University. He has edited a text on multiple pregnancies.

than that of singletons for all outcomes examined in this study.¹⁴

According to the Kramer (1987) survey, nutrition plays an important role in intrauterine growth. In both developed and developing countries, low caloric intake is found to be one of the most important modifiable contributors to IUGR. The other important modifiable contributor to IUGR, particularly in developed countries, is maternal cigarette smoking. This motivates our second identification strategy to quantify the costs of LBW. Based on analyses that use the correlation between LBW and infant health and costs, interventions that target maternal smoking cessation have been identified as having large benefits-to-costs ratios. Further, the experimental intervention studies have found substantive effects of smoking cessation on birth weight but no effects on gestation length. Consequently, maternal smoking presents another compelling context for examining the effects of IUGR.¹⁵

D. Measures of Health and Costs

We use four different outcome variables (h_{ij}) to characterize the consequences and costs of low birth weight. First, we consider infant death at periods up to 1 year after birth. The advantage of this measure is that it is a somewhat objective measure of severe health problems, and is in fact an outcome of direct interest. Smoking cessation campaigns, for example, are motivated not by how smoking impacts birth weight, *per se*, but instead by how smoking-induced LBW can affect infant mortality. As another example, the U.S. National Institutes of Health has adopted the view that the difference in LBW between blacks and whites is the leading cause of the substantial black-white gap in infant mortality. It is for this reason that it has proposed that infant birth weight be targeted for intervention.

Second, we examine the so-called APGAR score of initial infant health status. The National Center for Health Statistics describes this measure as a “predictor of the infant’s chances of surviving the first year of life” and a “summary measure of the infant’s condition” (NCHS *Vital Statistics* Technical

¹⁴ Conley, Strully, and Bennett (2003) argue that among identical twins sharing a placenta, the smaller twin may be even more deprived of nutrition than a comparable fraternal twin or singleton birth. They conclude that a fixed effects estimate based on identical twins will overstate the importance of size at birth.

¹⁵ Other identified contributors in Kramer (1987) were more generalized: pre-pregnancy weight, very young maternal age, maternal education, and weight gain during pregnancy. Other studies have found IUGR effects of alcohol and drug abuse, maternal nutrition, and prenatal care, but little evidence of effects on prematurity (Goldenberg and Rouse 1998).

Appendix, 1990). The APGAR score ranges from 0 to 10 and is calculated from five separate tests of newborn health made both one and five minutes after birth. At the time of birth, the doctor assesses each of the five factors, and for each factor a score from 0 to 2 is given. The five health factors are heart rate, respiratory effort, muscle tone, reflex irritability, and color. The five scores are summed to calculate the APGAR score. Since infant death can be extremely rare at birth weights above 1,500 grams, this measure is helpful when mortality indicators are too crude to measure underlying health.

Third, we consider the total charges incurred by the hospital for the delivery and initial care of the infants. This can be an important component of the societal costs of LBW. That is, even if it is possible to treat LBW infants so that they have similar health outcomes as normal weight babies, the price to pay for this parity may be the use of costly medical procedures. Hospital costs at discharge provide a comprehensive index of these procedures.

Fourth, we examine one of these costly procedures: the newborn's utilization of a ventilator for assisted-breathing shortly after delivery. In addition to being a specific component of the costs, it is also an alternative measure of any initial health problems that may be incompletely captured by the APGAR score.

In the analysis of twins, we examine the effects of birth weight on all four outcomes. Below, we find substantial intra-pair variation in these outcomes as well as in birth weight. This, along with the fact that the APGAR score is a substantive predictor of the variation in infant death and ventilator use, suggests that the relatively small birth weight effects that we document are not the result of either measurement error in the outcomes or behavioral responses of parents and hospitals – e.g., attempting to keep twins together until discharge.

Unfortunately, the hospital discharge data used in this study do not contain information on the smoking behavior of the mother during pregnancy. As a result, we cannot examine the direct association between maternal smoking and hospital costs, and, to date, we have not found any study that has been able to do this. However, the effect of maternal smoking on the use of assisted ventilation provides evidence on at least one very costly procedure.

E. Functional Form Issues

There are a number of important functional form issues involved in estimating Equations (3), (4), and (5). First, as suggested by the plots for singletons in Figure 1, the relationship between birth weight and various health and cost outcomes is nonlinear. In certain ranges – 300-1000 grams for mortality, and 800-2500 grams for costs – the relationship with birth weight is steep, but quickly flattens at higher weights. Therefore, we also estimate fixed effects models that include a set of dummy variables for discrete birth weight categories, given by

$$(6) \quad h_{ij} = \alpha + \sum_k D_{ij}^k \beta_k + X_i' \gamma + a_i + \varepsilon_{ij}$$

where D_{ij}^k is an indicator variable that is equal to one if the j^{th} infant of the i^{th} mother is in the k^{th} (out of 200) quantile of the overall birth weight distribution.¹⁶ We plot the entire set of estimated coefficients of β_k to give a complete picture of the effects at different parts of the birth weight distribution. In Figures 1A, B, and C, the solid squares plot these coefficients.

To quantify the slopes of the relationship throughout the birth weight distribution, we simply replace $\sum_k D_{ij}^k \beta_k$ in (6) with a piecewise linear spline specification with break points at 600, 800, 1000, 1500, 2000, 2500, and 3000 grams, and report the slope coefficients for each segment. For all health and cost outcomes, we use least squares to estimate these equations with and without fixed effects. For the binary outcome variables, mortality and ventilator use, we also estimate the above specifications by logit and fixed-effects conditional logit to examine sensitivity to the specification of the probability model.

Finally, for the maternal smoking analysis, we implement a propensity score procedure. The most flexible way to incorporate covariates is to simply compute smoking-nonsmoking differences for each distinct value of the observed covariates X , and average those differences. Due to the high dimension of X , we instead use a propensity score approach (Rosenbaum and Rubin, 1983). More specifically, we estimate a logit model that predicts the probability of maternal smoking during pregnancy based on all of the “pre-treatment” covariates that have been determined prior to pregnancy. These include mother’s and father’s age, education, and race, marital status, number of previous live births and

¹⁶ For the pooled singletons’ plots, we used 655 quantiles of the conditional distribution of birth weights below 3,000 grams.

terminations, prenatal care usage, months since last birth, immigrant status, county of birth, indicators for previous births over 4000 grams or LBW, indicators for alcohol use, and indicators for medical risk factors (see Torelli 2003).¹⁷ We then group the predicted probabilities of smoking for each observation – i.e., the propensity score – into 200 equal-sized bins from the overall distribution of propensity scores. To graphically illustrate the estimated “treatment” effects, we present the average outcomes for both smokers and nonsmokers in each of the 200 cells in which the propensity score is held constant.

III. Data and Summary of Differences between Twins

Here we describe the data sets used and document the degree of within-twin pair variation in twin birth weights and outcomes. We utilize two different data sources on twin births. The first is the annual, linked birth and infant death micro data produced by the National Center for Health Statistics (NCHS). These files provide detailed information on the population of U.S. twin births and death information for those who die within their first year of life. The second is the Healthcare Cost and Utilization Project (HCUP) State Inpatient Database, produced by the Department of Health and Human Services. The HCUP database contains detailed information collected from hospital discharge abstracts in participating states. We are the first to match twin pairs in either data set using a unique algorithm.

A. Linked Birth-Infant Death Data: 1983-1985 1989-1991, 1995-1997

These micro data files provide detailed information on the universe of births occurring each year in the United States, as reported on birth certificates, linked to death certificate information for the infants who die in their first year of life. The natality portion of the linked data provides socioeconomic and demographic information on each mother giving birth, including maternal age, race, educational attainment, marital status, childbearing history, prenatal care, and geographic residence. Beginning in 1989, the files also contain information on mother’s tobacco use during pregnancy. Information on the

¹⁷ Medical risk factors include anemia, cardiac disease, lung disease, diabetes, genital herpes, hydramnios/oligohydramnios, hemoglobinopathy, chronic hypertension, eclampsia, incompetent cervix, renal disease, Rh sensitization, and uterine bleeding.

father includes age, race, education, and Hispanic origin. The natality portion also includes detailed information on the newborn infant, including baby's sex, race, gestational age, birth weight, APGAR scores, use of assisted ventilation, and plurality – i.e., whether the infant was part of a multiple birth.

For the infants who die by the age of one, NCHS matches detailed information drawn from death certificates to the corresponding birth record. This information includes the infant's age at death in days, where the death occurred, and the precise cause of death. Annual linked birth/infant death files are available from 1983 through to the present, with the exception of the years 1992, 1993, and 1994, when only “unlinked” natality and mortality detail files are available. In the 1989 birth cohort, for example, approximately 97.4 percent of the 38,605 infant death records are matched to one of the 4,045,881 natality records.

Our analysis of twins only uses information for matched twin pairs. Multiple birth records are identified with the plurality variable (DPLURAL). In 1989, 2.2 percent of all births (90,022) were twin births (DPLURAL=2). While the corresponding twin pairs for each mother are not explicitly identified in the linked data, beginning with the 1989 file, records for twin pairs are generally located next to each other in the “denominator-plus” file produced by NCHS. The adjacency of twin records can be inferred from the detailed information on parental and pregnancy characteristics, which generally are repeated the same number of times as the plurality variable would indicate. The precise procedure used to confirm twin matches is described in the Data Appendix.¹⁸ Further, the birth order of the matched twins can be determined by the information provided on the number of children born to the mother at the time of birth. For the 1989 birth cohort, 58,132 matched twin pairs born to non-Hispanic, black and white native-born mothers are used in the analysis.

While births occurring between 1989 and 1991 are the primary focus of the analysis, twin pairs were also matched for the 1983 to 1985 and 1995 to 1997 birth cohorts. Before 1989, twin records are not located next to each other in NCHS's “denominator-plus” file. Therefore, the plurality indicator is

¹⁸ It should be noted that any “mismatching” of twin pairs will cause an upward bias in the twins fixed effects estimates of the effect of birth weight. Pollack, et al. (2000) use a different algorithm to match twins with their siblings in linked birth-infant death data. However, the matched twin pairs are only used to adjust their estimated standard errors for twin-pair correlation in the residuals.

used along with a string of eleven variables reflecting parental and pregnancy characteristics to match twin records with identical values. Checks of this alternative matching algorithm using variables not included in the matching string indicate that the matching is nearly as good as the record location-based approach. Moreover, the basic 1989-1991 results do not change when the covariate matching approach is used instead. For 1983 and 1984, NCHS did not include a unique identifier to enable one-to-one matching of birth and death records. Therefore, a string of 35 covariates is used to match the live birth records (where a death is known to occur) with additional information on the death provided by the “numerator file” record. Approximately 150 twin observations are lost in each year due to multiple potential matches. The elimination of these observations has only a nominal effect on the results.

For births occurring between 1995 and 1997, in which the location-based matching algorithm is used, father’s education was no longer collected by NCHS. Therefore, adjacent twin birth records for these cohorts are not checked for consistency in father’s education (see Data Appendix). However, maternal education, maternal age, race, and other factors are still checked for consistency within adjacent twin pairs.

In the fall of 2002, NCHS released the 1995-1997 Matched Multiple Birth Data Set that matches multiple births from the same delivery set. The algorithm NCHS uses to match twins is similar to the covariate approach used here for births between 1983 and 1985. NCHS does use additional information – namely mother’s date of birth and the half-week of delivery – that increases the precision of their covariate matches and is unavailable in the publicly-released linked files. However, NCHS does not make use of the information contained in the adjacency of multiple-birth records, as the algorithm applied in this study does for 1989 on.¹⁹ Despite the different matching algorithms, below we find that the empirical results based on the NCHS-matched data are nearly identical to those based on our 1995-1997 location-based matching algorithm.

¹⁹ NCHS does not maintain information on the set to which a multiple birth belongs, nor does it collect information on the name or address of the mother, which would help in identifying twins from the same birth set. Further, the NCHS-matched twins data suppresses information on the year (between 1995 and 1997) in which the multiple births occurred. This is the data used in Conley, Strully, and Bennett (2003).

B. Healthcare Cost and Utilization Project Data: 1995-2000

The HCUP state inpatient database contains detailed information from the universe of hospital discharge abstracts in participating states. In 1995, discharge data for thirteen states are available with approximately 14 million records. In 2000, 18 states provided data with approximately 16.5 million discharge records. According to HCUP, approximately 80 percent of all U.S. hospital discharges are covered in their data.

The HCUP data include detailed information on hospital stays and patient characteristics. Clinical information includes the diagnosis, course of treatment, information on the attending physician, primary surgeon, the length of hospital stay, charges incurred during hospitalization, and the disposition of the patient at discharge. Core demographic information on the patient is also provided, including the age, race, and sex, and geographic residence of each patient. Finally, information on the timing of admission, discharge, hospital at which treatment occurred, and the expected payment source is also included.

For the purposes of this study, the key data elements are the birth weight of newborn infants, the diagnosis (used to select twin births, as described in the Data Appendix), and the charges incurred in the course of treatment. The availability of individual data elements varies by State. Only five of the eighteen states provide information on birth weight: Arizona, Colorado, Maryland, New Jersey, and New York. Of these five states, our analysis uses 1995-2000 discharge data from New York and New Jersey, which compose almost one-quarter of all available discharge records in 2000. We also examined the discharge data from Maryland, which provided nearly identical results (available from the authors).²⁰

Twins can be matched using the same two approaches used for the linked birth-infant death data. As most hospital admissions are for procedures other than delivery and given the relative infrequency of twin births, the adjacency of twin records in the ASCII HCUP data is rare, and generally indicates that the twins are part of the same set. Secondly, information on the twins that should be common within twin

²⁰ As more than half of the Arizona discharge records were missing information on birth weight, Arizona is excluded from the analysis, along with the Colorado, which had the smallest sample size (15 percent of New York sample). We do not include Maryland in our analysis since the (location-based) algorithm used to match Maryland twins is different from the algorithm (location- and covariate-based) used to match New York and New Jersey twins.

sets can be used to identify unique twin pairs independent of record location. As twin discharge records cannot be linked to the maternal discharge record, information on the twin alone must be used to match multiple birth records to the appropriate set. Both matching procedures are described in greater detail in the Data Appendix. Since they led to similar findings, this study presents the results based on the covariate matching scheme, which provides larger samples than record location matching.

Between 1995 and 2000, 44,500 (non-Hispanic white and black) newborn twin records are matched in New York and New Jersey using the covariate matching approach. Average birth weight for these twins is 2440 grams, which is close to the average birth weight in the linked vital statistics data. The average length of hospital stay is 9 days; the initial hospitalization charges averaged \$15,000 per twin; and 1.4 percent of twins died prior to discharge.

C. Twins Compared to Singletons

Although twins represent a small subpopulation of all infants born in the U.S., they are of great interest in the health and biomedical literatures for several reasons. First, while constituting 3 percent of all births in the U.S. in 1997, twins accounted for 21 percent of all LBW births, 14 percent of preterm births, and 13 percent of all infant deaths (Kogan, et al., 2000).²¹ Further, Kogan, et al. (2000) finds that multiple births have accounted for an increasing share of all LBW infants over the past 20 years. Since LBW is strongly associated with increased risk of infant mortality and subsequent developmental difficulties, several researchers have suggested that the rising incidence of twin births is an important public health problem.²² Finally, several studies suggest that interventions that decrease infant mortality among twins should be applicable to other high-risk groups (e.g., Fowler, et al., 1991).

Table 2 provides summary statistics for both singleton and twin births for one of the years of data used in our analysis, 1989. It highlights some of the main similarities and dissimilarities between the two

²¹ In the 1989-1991 sample, twins account for over 2 percent of all births, 15 percent of LBW births, and 10 percent of all infant deaths in the U.S.

²² For example, Behrman and Rosenzweig (2001) calculate that twinning results in a 12 percent reduction in lifetime earnings when compared to singleton births due to their (28-ounce) lower birth weights. They conclude that the growing incidence of twin births resulting from the increased use of fertility procedures among older women imposes significant costs on children's future development.

populations. Mothers of singleton and twin infants have comparable demographic characteristics, along race, age and education. Owing to the large sample sizes, the differences are statistically significant, but the means are similar – mothers of twins are slightly more likely to be black, older, more educated, and married.

There are two important differences between the two populations. First, twins tend to be lighter, with mean birth weights about 950 grams lower than singleton newborns. As Table 2 shows, the entire distribution of twin birth weights is shifted leftward in comparison to the singleton distribution. Indeed, the median twin would be considered a LBW baby. The prevalence of twins with low weights is helpful for the identification of the birth weight effects. This is especially relevant since the cross-sectional relation between birth weight and adverse outcomes among singletons is steeper at very low weights (Figure 1). Second, along many measures, twins tend to be less healthy at birth than singletons. They are much more likely to die within 1-day, 1-week, 1-month, and 1-year of the birth. Also, their APGAR scores are lower, and their gestation lengths are shorter, with the average twin being born prematurely.

In view of these differences, we do two things when reporting our results. First, we compare the fixed effects estimates to the pooled cross-section estimates for the same twins population. This ensures that the differences between the two sets of estimates can be attributed to the presence of omitted variables, rather than to a changing population. Second, we compare the pooled cross-section estimates for the twins to that for the singleton population. If the birth weight-health cross-sectional relations in these two populations are substantially different, it can suggest that the twins' results may not be readily generalizable to the larger population of births. If they are similar, then one cannot reject the hypothesis that twins' and singletons' outcomes are characterized by the same equation (1), and subject to similar omitted variables biases as in (2).

D. Within-Twin Pair Variation

Table 3 presents a variance decomposition of birth weight and the health and cost outcomes for U.S. twins. Column 1 reports the total variance; column 2 the residual variance from a regression controlling for a linear term in gestation length (in weeks); column 3 the residual variance from a

regression controlling for gestation length fixed effects; and column 4 the within-twin pair variance – i.e., the residual variance after controlling for mother fixed effects.

The first row shows the results for birth weight for the 1989-1991 birth cohorts of twins.²³ The third column entry implies that variability in intrauterine growth rates account for $(19.08/44.43=)$ 43 percent of the overall cross-sectional variation in birth weight, with gestation length accounting for the remaining 57 percent. The final column reports that 40 percent of the variation in IUGR is due to within-twin pair differences in birth weight.²⁴ This suggests that – conditional on gestation length – the factors that cause differences in growth rates between twins of the same mother may also be important in explaining across-family differences in twin growth rates.

The remaining rows show substantial within-twin pair variability in the outcome variables, also. The second through fourth rows show that intra-pair differences in mortality account for 45 to 68 percent of the variation in infant death adjusted for weeks of gestation. The within twin-pair variation in APGAR scores and ventilator use is also significant. Since the discharge data do not contain gestation length, we cannot compute a similar ratio for hospital costs. Nevertheless, the ratio of within-pair variation to the overall variation (column 1) is 21 percent, which is on the same order of magnitude as the corresponding ratios for the other variables (17, 42, 25, 32, 34, and 28 percent, respectively).

The first three columns in Table 4 show the differences in outcomes between the heavier twin and his/her lighter co-twin for the 1989-1991 NCHS birth cohorts.²⁵ The first row shows that the difference in birth weights is 300 grams, on average. To put this magnitude in perspective, consider that maternal smoking during pregnancy – believed to be the most important modifiable cause of LBW – has been shown to have about a 200 gram impact on birth weight. The next three rows show that twin differences in birth weight occur throughout the birth weight distribution, leading to large differences in the likelihoods of the lighter twin weighing less than both 1500 grams and 1000 grams.

²³ The scaling of the variables is described in the table notes.

²⁴ It should be noted that this is probably a lower bound figure. Since gestation length in the NCHS data is measured in weeks, the residual variation in birth weight after adjusting for the number of days of gestation would likely be less than 19.08. Also, the within-twin pair variation is slightly greater for the 1995-1997 birth cohorts.

²⁵ For about 5 percent of the pairs, the twins have identical birth weights.

The remaining rows show that the lighter twin also tends to experience worse health outcomes. But the implied impact of birth weight is substantially smaller than that implied by the cross-sectional correlations. For example, this 300 gram difference is associated with 6.8 more deaths (within the first year of life) per 1,000 live births. The coefficient from the cross-sectional regression in Table 1 implies (0.1168 times 300 grams =) 35 deaths for the same decrease in weight. Similarly, the heavy-light comparison yields a 0.03 difference in the 5-minute APGAR score, whereas the corresponding cross-sectional estimate from Table 1 would have predicted a 0.32 difference (300 times 0.105 divided by 100).

Further, the sixth row shows that the lighter twin is significantly more likely to have a congenital anomaly, suggesting that in fraternal twin pairs the lighter twin may be genetically disadvantaged.²⁶ The bottom rows show that the infant mortality differences fall sharply when the three percent of twin pairs in which one twin suffers from a congenital anomaly are dropped from the sample. The one-year mortality difference drops to about 2.6 deaths per 1000 live births. While the difference is statistically significant, it is very small in magnitude. Given the large sample sizes, statistical significance is not a reliable indicator of the importance of an effect.

Before proceeding, we note that we can identify the birth order of the twins in a pair.²⁷ This information allows us to test twin exchangeability – an implicit assumption of the fixed effects specification. For example, we can test whether the effect of birth weight when the firstborn is heavier is similar to the effect when the second-born is heavier. The last three columns of Table 4 show that the firstborn twin tends to be healthier at birth than the second-born twin. Given these systematic differences, it is important to test the implications of the exchangeability assumption.

²⁶ A congenital anomaly is a defect that is present at birth. These include genetic defects and disorders, but also include conditions that could plausibly be the result of damage incurred during fetal development. As a result, we always present the estimates from both the full sample of twins and from the sample that excludes twin pairs with congenital anomalies, and we use the former sample for all cost-benefit calculations.

²⁷ Birth order could not be determined for about 10 percent of the twin pairs.

IV. Empirical Results

A. Main Twins Results

Table 5 is an expanded version of Table 1, providing more details on the results from the twins analysis. The health outcomes data are for all twins born in the United States between 1989 and 1991, and the hospital costs data are for twins born in New York and New Jersey between 1995 and 2000. In the second set of rows, the cross-sectional estimate (or “Pooled OLS”) of -0.0739 shows that the majority of the association between infant mortality and birth weight is driven by mortality risk within the first day of life. When mother fixed effects are included, the coefficient falls by more than a factor of 10. Thus, while the cross-section implies that a 300 gram birth weight difference is associated with 22 more deaths within 24 hours of birth per 1,000 live births, the within-pair estimator suggests an impact of only 2 deaths per 1,000 live births.

The fourth column provides evidence that the fixed effects estimates themselves are largely driven by the small number of twin pairs in which at least one infant is born with a congenital anomaly. For infant mortality within a year of birth, the impact of birth weight falls further, by more than a factor of 2.7. For one-day mortality, the fixed effects estimate falls by an additional factor of 24, and the point estimate is no longer statistically significant. The fixed effects estimates for the other outcomes also fall to varying degrees when these twin pairs are excluded, and the fixed effects estimates for ventilator use are statistically insignificant in both samples. By contrast, the third column shows that that the exclusion of these anomalous pairs (less than 3 percent of the sample) has a relatively minor impact on the cross-sectional estimates.²⁸

The fixed effects estimates for the full sample are five to 20 times smaller in magnitude than the cross-sectional estimates, and they are 14 to 40 times smaller than in the cross-section for the sample excluding “anomalous” pairs.²⁹ The one exception is for infant death occurring 28 days or more after birth (postneonatal); in which the fixed effects estimates are two to three times smaller than the cross-

²⁸ Indicators for congenital anomalies are not available on the HCUP discharge data.

²⁹ Adding controls for infant gender and birth order to the regressions results in nearly identical findings.

sectional estimates. Below, we find that the majority of the postneonatal mortality effect is driven by death due to sudden infant death syndrome and death due to accidents and homicide.

Figures 1 and 2 present our graphical analysis of the birth weight effects that allow for very flexible relationships between birth weight and the outcomes of interest. In each figure, the solid circles are the means of the outcome, by quantiles (200 of them) of the birth weight distribution. The solid squares are the corresponding fixed effects estimates – that is, the estimated coefficients β_k from the specification in Equation (6).³⁰ When data on congenital anomalies are available, the solid triangles are the fixed effects estimates when infants with those anomalies (and their co-twins) are excluded from the sample.

In Figure 1A, the cross-section reveals a peak in costs for hospital services of \$160,000 for a twin infant weighing nearly 800 grams at birth, and a steady and significant decline in costs with increases in birth weight.³¹ The fixed effects estimates reveal a substantially smaller effect throughout the 800 to 2,000 gram range of birth weights. The contrast between the cross-sectional and fixed effects associations is even more pronounced in the 2,000 to 2,500 gram range. As Figure 1B illustrates, for the cross-section the decrease in costs continues to be substantial until the 2,500 gram mark. By comparison, the fixed effects estimates are relatively flat throughout that range of birth weights.

Figure 1C shows qualitatively similar results for mortality rates within one year of birth. In the cross-section, the relationship between birth weight and mortality is steepest in the 300 to 1,000 gram range. This is also the range of birth weights in which the differences between the cross-sectional and fixed effects gradients are the largest – that is, the fixed effects slopes are substantially smaller in magnitude than the cross-sectional slopes. On the other hand, the fixed effects and cross-sectional gradients are similar in the 1,000 to 1,500 gram range,

The results for the 5-minute APGAR score in Figure 2A show significant differences between the cross-sectional and fixed effects “response functions” throughout the birth weight distribution. In the

³⁰ The average of the mother-specific intercepts is used in reporting the results.

³¹ The rise in costs in the 300 to 800 gram range is likely due to a simple selection effect, as babies born in that weight range are highly likely to die soon after birth, and therefore accumulate fewer charges for hospital services. In the HCUP data, we find infant death rates before discharge that are very consistent with this selection effect.

cross-section, the APGAR-birth weight relation is strongly positive even at birth weights well above 1,000 grams. The fixed effects estimates reveal a somewhat positive relation in the 300-800 gram range, but very little relation at birth weights above 800 grams.³²

There are even more striking differences between the cross-sectional and fixed effects profiles for the use of a ventilator for assisted breathing in Figure 2B. In the cross-section, ventilator use is strongly negatively related to birth weight in the 800 to 2500 gram range.³³ By comparison, the fixed effects estimates reveal a much more modest decline in ventilator use. As shown in Figure 2C, this is particularly true for the use of a ventilator for more than 30 minutes, which is considered a stronger indicator of initial health problems than overall ventilator use. There, the fixed effects estimates reveal essentially no relation between birth weight and the use of this expensive procedure.

Figures 1C, and 2A-2C, also show that excluding the small number of infants with congenital anomalies tends to further flatten the birth weight-outcome profiles. Based on this pattern – and the qualitative similarity between the ventilator and hospital costs profiles – it seems plausible that the fixed effects profile for costs at discharge is also partially driven by these congenital anomaly cases.

Finally, Figures 1A-1C and 2A-2C also present the cross-sectional relations for the population of singleton births, represented by the open circles. For every outcome that we examine, the birth weight profiles for the singletons' population are virtually identical to the cross-sectional profiles for the twins' population. This is strongly consistent with the notion that the difference between the singleton cross-sectional profile and the twins fixed effects profile is primarily due to omitted variables, rather than due to inherent differences between singleton and twin births. Thus, extrapolating the twin fixed effects findings to the singleton population may not be a completely unjustified exercise.

³² We find very similar results for the one-minute APGAR score (see Almond, Chay, and Lee 2002).

³³ Again, the hump-shaped pattern at extremely low birth weights can be explained by a selection effect in which these babies die before having the opportunity to use a ventilator for assisted breathing.

B. Magnitudes of the Effects

Table 6 presents estimates of the implied impact of birth weight in terms of its effect on the outcomes per gram of birth weight. Since the effects are highly nonlinear, we estimate piecewise linear splines with knot points at 600, 800, 1000, 1500, 2000, 2500, and 3000 grams. The table reports the slope coefficients for each segment of the spline, for both the cross-sectional OLS and fixed effects specifications applied to the full samples of twins.

First, we consider the results for costs incurred by the hospital before discharge. In Table 5 the cross-sectional estimate implies that each gram increase in birth weight is associated with a \$30 reduction in hospital costs, whereas the fixed effects estimate implies only a \$5 dollar cost reduction. Table 6 shows that even as the marginal effect of birth weight on costs varies significantly across the birth weight distribution, the fixed effects estimates are considerably smaller in magnitude than the corresponding “pooled” estimates. For the birth weight range of 800 to 2,500 grams, the fixed effects coefficients are 3.1 to 5.7 times smaller than the pooled coefficients. In the 2,500-3,000 gram range, the pooled coefficient implies a highly significant five dollar reduction in costs for every added gram in weight, but the corresponding fixed effects estimate implies absolutely no cost savings.

The next sets of columns in Table 6 show very similar results for any ventilator use and ventilator use for more than 30 minutes. In fact, the fixed effects slope coefficients are statistically insignificant throughout the birth weight distribution. For the APGAR score, the fixed effects coefficients are 3 to 20 times smaller than the cross-section coefficients.

For infant mortality in the final two columns, the differences between the cross-section and fixed effects estimates are particularly large at birth weights below 1,000 grams. For example, in the 600 to 800 gram range, the cross-section estimate is -2.11 per 1,000 live births, while the fixed effects estimate is -0.42 . On the other hand, in the 1000-1500 gram range the cross-section and fixed effects estimates are -0.23 and -0.17 , respectively.

When interpreting their magnitudes, these two sets of estimates must be placed within the context of the highly nonlinear relationship illustrated in Figure 1C. For example, consider a 200 gram increase in birth weight. In the 600-800 gram range, the cross-section estimate implies a decrease in the

probability of death of 42 percent; the corresponding fixed effects estimate implies a decrease of 8.4 percent. On the other hand, that same 200 gram increase in the 1000-1500 gram range implies decreases in the probability of death of 4.5 and 3.4 percent for the cross-section and fixed effects estimates, respectively. Also, it is perhaps misleading to point to the relatively small differences between the two estimates in the 2,000-2,500 gram range as evidence of minimal omitted variables bias. After all, the birth weight impact implied by the cross-sectional estimate is already small: a 200 gram increase in that range leads to a decrease in the probability of death of 0.0028.

It is also misleading to rely solely on statistical significance in assessing the magnitudes of the effects. The fixed effects estimate in the 1500-2000 gram range, for example, has a t-ratio of 3.17, but the implied impact of a 200 gram decrease in birth weight – the effect size associated with maternal smoking – is an increase in the probability of death of only 0.0078.

Table 7 summarizes the excess costs associated with low birth weight, as implied by both the cross-sectional and fixed effects estimates. For every infant born under 2,500 grams in the NCHS birth certificate files (1989-1991), we calculated how much money would be “saved” by raising their weight to 2,500 grams. That is, we imputed the hospital costs in excess of the costs associated with 2,500 grams for every infant, using the slope coefficients in the first two columns of Table 6. Table 7 reports the average imputed value within each birth category. In each category above 600 grams, the fixed effects estimates imply significantly smaller excess costs associated with low birth weight.³⁴ For example, the cross-sectional estimates imply that increasing the weight of a typical baby born in the 800-1000 gram range to the “normal” weight range would save over \$127,000 in hospital charges. However, the fixed effects estimates imply a cost savings of \$37,000.

Also, the degree of omitted variables bias appears to grow larger in higher birth weight categories. For example, the implied excess costs among infants born in the 1500-2000 gram range are 3.7 times smaller for the fixed effects estimates relative to the cross-section. For the 2000-2500 gram range, the fixed effects estimates are 5.7 times smaller than the cross-sectional estimates.

³⁴ The smaller excess costs among infants born below 600 grams are due to the “perverse” birth weight effects for babies born under 800 grams that are documented in Figure 1A and Table 6.

These differences have striking implications for the calculation of the aggregate costs of LBW in the United States population. The distribution of birth weights can be applied to the numbers in Table 7 to impute the aggregate excess costs for all infants born below 2,500 grams. For twins born in the U.S. between 1989 and 1991, the cross-section estimates imply aggregate costs of LBW of \$2.22 billion, while the fixed-effects estimates imply aggregate costs of \$691 million. Table 7 also shows the birth weight distribution for the 2.7 million singletons born in 1989 to non-Hispanic, black and white, native-born mothers. For singleton births in 1989, the (twins) cross-section estimates imply aggregate costs of LBW of \$3.40 billion, while the fixed effects estimates imply aggregate costs of \$1.07 billion.

C. Maternal Smoking Results

Figure 3 graphically depicts a propensity score analysis of the empirical association between maternal smoking during pregnancy and birth weight for the 497,139 singleton births in Pennsylvania between 1989 and 1991.³⁵ First, we estimated a logit model of the propensity to smoke conditional on a multitude of “pre-treatment” maternal and family background characteristics.³⁶ Then we grouped the estimated propensity scores into 200 equal-sized cells and separately calculated the average birth weight of the babies of nonsmoking and smoking mothers in each cell. Thus, the smoking-nonsmoking birth weight difference in each cell provides the estimated birth weight effect of maternal smoking at a constant propensity score. According to the propensity score theorem (Rosenbaum and Rubin 1983), this effectively adjusts for all bias due to observable confounders.

Panel A presents the average birth weights of babies of non-smoking mothers (the crosses) and of smoking mothers (the open circles), by values of the estimated propensity score. The graph reveals a visible birth weight gap that is relatively uniform across the entire range of estimated propensity scores, even as baseline birth weights decline at high propensity scores. Panel B plots the 200 smoking-nonsmoking differences in birth weight, which appear to be tightly scattered at around –200 grams. This

³⁵ We focus on Pennsylvania since there is complete smoking information for over 95 percent of the mothers in 1989-1991. In the singleton sample, about 21 percent of Pennsylvanian women report smoking during pregnancy. It should be noted that we found very similar results for the other states in which smoking reporting was comprehensive – i.e., Florida, Georgia, Illinois, Michigan, North Carolina, and Ohio.

³⁶ For the details of the specification, see Torelli (2000).

difference is consistent with the regression-adjusted estimate of the birth weight effect of maternal smoking reported in the first row of Table 8.³⁷ In addition, it is similar in magnitude to the birth weight effects documented in the experimental intervention literature (e.g., Sexton and Hebel 1984)

Panel C is analogous to Panel B, but the outcome is instead whether or not the birth weight is less than 2,500 grams. The graph shows a higher incidence of LBW among smoking mothers, by about 3 to 4 percent, which is again consistent with the OLS estimates reported in the second row of Table 8. The LBW differential appears to be slightly larger at propensity scores above 0.7. On the other hand, Panel D shows no evidence of a systematic relation between maternal smoking and the incidence of birth weights below 1,000 grams. Table 8 shows that while there are significant raw differences in the incidences of birth weights below 1,500 and 1,000 grams (column 1), these differences disappear after regression adjustment for maternal and background characteristics. In fact, after controlling for observable confounders (column 2), maternal smoking only appears to affect birth weights for infants with weights above 1,500 grams. In addition, the next rows of Table 8 reveal very little effect of maternal smoking on gestation lengths and the likelihood of a premature birth.³⁸ Thus, maternal smoking appears to influence intrauterine growth rates, but not gestation durations.

Figures 4A and 4B are the analogue to Figures 3A and 3B, except that the dependent variable is now infant mortality within one year of birth. The graphs reveal no systematic impact of maternal smoking on infant mortality throughout the propensity score distribution. The bottom rows of Table 8 show that while there is a significant raw difference in the infant and neonatal mortality rates of smokers and nonsmokers, this difference disappears after regression adjustment.

Panels C and D of Figure 4 show zero-impact of maternal smoking on the 5-minute APGAR score. Surprisingly, Panels E and F also reveal no impact of maternal smoking on the probability that the newborn requires the use of a ventilator for more than 30 minutes. Again, the regression-adjusted estimates in Table 8 are quite consistent with the visual impressions left by the figures.

³⁷ The regression analysis adjusts for the exact same variables that are used to estimate the logit propensity score model. The same transformations and interactions of the variables are also included.

³⁸ Live births of less than 32 weeks of gestation account for the vast majority of neonatal deaths and disorders.

It appears that – once mother’s characteristics are included in the analysis – birth weight stands out as the *only* variable that has a strong empirical association with maternal smoking. As noted earlier, if smoking has other deleterious effects on infant health independent from its effect through lowering birth weight, then the implied impact of birth weight on all of these outcomes are probably upwardly biased. That is, however small the implied birth weight effect is, it would have to be even smaller if maternal smoking is expected to independently raise infant mortality, lower the APGAR score, or make it more likely that the infant will need a ventilator. Also, we found no effect of maternal smoking on fetal death rates (results available from authors). This implies that, for the case of maternal smoking, conditioning the analysis on the population of live births is not prone to nonrandom sample selection bias.

Finally, we use the estimates in Table 6 and Table 8 to calculate the savings in initial hospital costs achieved by encouraging women to cease smoking during pregnancy. First, we derive estimates of the cost savings based on the conventional approach used in the literature – that is, we multiply the cross-sectional estimates of the association between costs and birth weight by the raw effects of maternal smoking on LBW incidence. Suppose that maternal smoking reduces birth weight by 200 grams, on average, and that its effect on the incidences of low, very low, and extremely low birth weights are the same as in column 1 of Table 8. Then the costs-birth weight gradients in Table 6 imply an estimated cost savings of \$907 for every woman who ceases smoking before pregnancy.

Next, we derive estimates of the cost savings based on the fixed effects estimates of the costs-birth weight gradients in Table 6 and the regression-adjusted effects of smoking on the incidences of low, very low, and extremely low birth weights shown in column 2 of Table 8. Again, presume that maternal smoking reduces birth weight by 200 grams. The estimated cost savings falls by a factor of 17 to only \$53 for each woman who is prevented from smoking during pregnancy.³⁹ Below, we discuss the implications of this finding for cost-benefit analyses of smoking intervention programs targeted at pregnant women.

³⁹ Ideally, we could estimate the direct effect of maternal smoking on hospital costs. Unfortunately, the hospital discharge data do not contain information on the smoking behavior of mothers. However, the adjusted estimate of the effect of smoking on ventilator use provides evidence that at least one costly procedure is not affected by maternal smoking during pregnancy.

D. An Alternative Measure of Initial Health: the APGAR score

The results above suggest a much weaker association between birth weight and early health and cost outcomes than previously recognized. Here, we examine whether an alternative measure of the initial health of the infant performs better than birth weight in predicting other health outcomes. In particular, we consider the 5-minute APGAR score, which has the same practical advantages as birth weight: 1) it is relatively easy to collect; 2) it is already available in United States birth record data; and 3) it is a continuous measure that does not depend on a rare event (such as mortality).

In Table 9 we report the coefficients from cross-sectional and fixed effects regressions of the other health outcome measures on both birth weight and the APGAR score for the 1989-1991 birth cohorts of twins. We use the t-ratios of the estimated coefficients (shown in square brackets) to compare the predictive power of the two indices. The APGAR score performs notably better than birth weight in predicting within-twin pair differences in both one-year and one-day mortality. This is particularly true when twin pairs in which one sibling exhibits a congenital anomaly are dropped from the analysis.

The APGAR score clearly outperforms birth weight as a predictor for our proxy of costs, the use of a ventilator for assisted breathing. While the cross-sectional coefficients for birth weight are highly significant, the fixed effects coefficients are essentially zero and statistically insignificant for both any ventilator use and ventilator use for more than 30 minutes. By contrast, the fixed effects and cross-sectional coefficients on the APGAR score are quite similar and highly significant. Perhaps this should not be surprising since respiratory effort of the newborn is one of factors used to calculate the APGAR score.

It is also interesting to note that the birth weight effects in Table 8 suggest large returns to maternal smoking cessation. On the other hand, the effect of maternal smoking on the APGAR score implies much smaller benefits, which is consistent with the findings for the effects of smoking on the other outcome variables (ventilator use and infant mortality).

V. Robustness of the Findings

This section explores the sensitivity of the results to: 1) zygosity and gender of the twins; 2) years of birth; 3) different race and education groups; 4) functional form issues; and 5) different matching algorithms.

A. Zygosity and Gender

The data that we examine do not allow us to distinguish between monozygotic (identical) and dizygotic (fraternal) twins. Thus it should be noted that for some fraction of our sample, genetic differences between dizygotic twins of the same mother may be contributing to within-pair variation in birth weight. We provide a partial assessment of the differences in effects by zygosity, by examining the necessarily dizygotic infants of boy-girl twin pairs. For the 1989-1991 birth cohorts, 59,308 twins are in boy-girl pairs, and 129,728 twins are in same sex pairs. This suggests that approximately 54 percent of the same sex pairs are identical and 46 percent are fraternal.⁴⁰

Examining the results by twin gender-types is also helpful in assessing a potential measurement issue in the infant mortality analysis. That is, if some twins are accidentally “switched” either at the hospital or by the parents once at home, then the death record could, in principle, be matched to the incorrect birth record. As an extreme example, if all twins were randomly switched with their sibling, then one would expect to observe a zero fixed effects estimate of the effect of birth weight on infant mortality.⁴¹ Naturally, this is not an issue for boy-girl twins. Thus, if random switching is a substantive problem for same-gender twins, then the fixed effects estimates for such twin pairs would be predicted to be significantly smaller than the fixed effects estimates for boy-girl twin pairs.

⁴⁰ If the effects of birth weight on infant mortality are constant across the birth weight distribution and do not interact with twin’s gender, then the so-called “Weinberg rule” may be used to derive the effects of birth weight among identical twins (e.g., Conley et al. 2003). However, we find that neither necessary condition holds in our data, and therefore do not use the weighted average of the mixed-gender and same-gender fixed effects estimates to derive fixed effects estimates for “identical” twins.

⁴¹ That said misclassification error due to the assignment of twins to an incorrect birth or death certificate is an unlikely source of variation in twin pair birth weight differences. Specific safeguards are in place to prevent this. For example, often two separate teams are present at each twin delivery to examine each newborn. Babies are labeled with tags almost immediately after birth and remain tagged until they leave the hospital. Nevertheless, we thank Dr. Christopher Almond for providing us with this information.

Table 10 reports the coefficients on a linear term in birth weight from linear probability models of three different mortality measures. It shows that there are small differences in the fixed effects estimates of boy-boy and girl-girl twin pairs for the full sample of twins. However, these differences effectively disappear when the sample excludes the small number of twin pairs in which one sibling has a congenital anomaly. Moreover, the fixed effects estimates for the mixed-gender twin pairs are nearly identical to those for the same-gender pairs. Thus, these findings are consistent with effects that do not differ substantially by zygosity, and they also suggest that the results are not an artifact of the switching of siblings within a twin pair.

B. Results across Years of Birth

Next we examine the stability of the infant mortality-birth weight relationship across different cohorts of twins born between 1983 and 1997. Figure 5 plots the cross-sectional (Panel A) and fixed effects (Panel B) estimates of the birth weight profiles of infant mortality for the 1983-1985, 1989-1991, and 1995-1997 birth cohorts. Panel B shows fixed effects profiles that are remarkably similar over time. By contrast, the cross-sectional profiles in Panel A change considerably over time and show a sharp improvement in the survival rates of more recent birth cohorts at each birth weight below 1,500 grams. For example, in the early 1980s, nearly 3-in-4 twins born at 700 grams died within a year of birth; by the late 1990s, the mortality risk had fallen to 1-in-3. In addition, while an increase in birth weight from 800 to 1,000 grams is associated with a 0.25 decrease in the probability of death in 1983-85, it is associated with a 0.10 probability decrease in 1995-97.

Table 11 presents the linear probability estimates of the birth weight effects on infant and neonatal mortality for the same sets of birth cohorts. The results confirm the visual impressions left by the figures. While the fixed effects estimates are quite similar across birth cohorts, the magnitude of the cross-sectional estimates is over 34 percent smaller for the 1995-97 birth cohort than for the 1983-85 cohort.

The instability of the cross-sectional estimates is consistent with the notion that the mortality-birth weight, cross-sectional relation is largely driven by omitted variables. In particular, unobserved

secular factors over time – including advances in medical technologies and changes in the quality of health care – appear to have a large effect on the cross-sectional associations. By contrast, the stability of the fixed effects estimates is consistent with the view that the twins’ analysis eliminates the bias caused by these omitted variables. It also implies that our findings above are not unique to the 1989-1991 period.

C. Functional Form Issues

We examine the sensitivity of the results to three functional form issues, in particular for the binary dependent variables in the study (mortality, ventilator use). First, we test the appropriateness of the piecewise linear spline specification used in Table 6 by simply comparing the predicted values from the spline specification to those from the more flexible dummy variable specification used to create Figures 1 and 2. Figure 6 shows the predictions from the spline models based on the cross-section estimates (indicated by the solid line) and based on the fixed effects estimates (indicated by the open circles) for infant mortality (Panel A) and ventilator use (Panels B and C). In each case, these predicted values are coincident with the profile given by the dummy variable specification (connected solid circles and connected solid squares).⁴²

Second, we examine the appropriateness of the linear probability model (LPM) that we have used for infant mortality and ventilator use, by comparing their estimates to those from a conditional (“fixed-effects”) logit specification. For the cross-section, a logit with dummy variables yields exactly the same estimates as does the LPM dummy variable specification; thus, we focus on the fixed effects specifications.

Figures 6B and 6C show that for any ventilator use, and ventilator use for more than 30 minutes, the conditional logit with the linear spline yields a virtually identical profile as both the LPM linear spline and the LPM with birth weight dummy variables. This implies that our results are not sensitive to the choice of a linear probability or logit model.

⁴² For a valid comparison, all predicted values are given in terms of probabilities. For the conditional logit result, since no “fixed effect” is identified, the intercept can be arbitrarily chosen. We chose to set the intercept such that the predicted probability of death at 2500 grams is the same as that given by the cross-sectional estimates.

For infant mortality (Figure 6A), while the conditional logit and LPM profiles are very similar throughout most of the birth weight distribution, there is a noticeable difference between the two specifications in the 300 to 1000 gram range. The logit specification is steeper than the LPM in the 300-600 gram range, but flatter in the 600-1000 gram range. This difference is not an artifact of the spline specification, as a conditional logit with birth weight dummies (open triangles) is coincident with the conditional logit spline specification (small solid circles). Overall, it appears that our LPM results actually *understate* our primary mortality finding – that the fixed effects make the most difference where the cross-sectional mortality-birth weight relationship is the steepest (600-1000 grams).

Finally, utilizing the information we have on the birth order of twins, we provide a simple test of the fixed effects specification itself. There is an implicit restriction in (3) that can be relaxed to yield:

$$(7) \quad h_{i1} - h_{i2} = \alpha_1 - \alpha_2 + bw_{i1}\beta_1 - bw_{i2}\beta_2 + \varepsilon_{i1} - \varepsilon_{i2}$$

where the ‘1’ and ‘2’ denote the first- and second-born twin. β_1 and β_2 can be separately estimated by regressing the twin difference in the mortality indicator on each sibling’s birth weight. If the two estimates are different, it suggests that the “twin exchangeability” assumption underlying the fixed effects specification does not hold and that the fixed effects specification is inadequate.

In practice, we utilize a more flexible parameterization, using the full set of birth weight dummy variables – as in equation (6) – for both twins’ birth weights. Figure 7 plots the estimates of the dummy-variable analogues of β_1 and β_2 (the open circles and squares) along with the cross-sectional mortality-birth weight profiles of the first- and second-born twins. In both cases, the estimated profiles are nearly identical. This is consistent with the hypotheses that the exchangeability assumption holds and that the fixed effects specification is valid.⁴³

⁴³ In a correlated random effects framework, this specification test can be reformulated as a test of similar mortality effects in twin pairs in which the first-born is heavier and in twin pairs in which the second-born twin is heavier. In Almond, Chay, and Lee (2002), we implement these tests for both the linear regression specification and the piecewise spline specification of infant mortality and APGAR scores.

D. Alternative Matching Algorithm

We also examine the sensitivity of our results to the method of matching twins, for the birth cohorts (1995 to 1997) in which two different methods are possible. In particular, we compare our matching procedure to the procedure used by the NCHS in the 1995-1997 Matched Multiple Births Data set, which was released in Fall 2002. Figure 8 plots both the cross-sectional and fixed effects estimates of the (one-year) infant mortality-birth weight profiles, derived from the two matching algorithms. The figure shows that the two sets of estimates are nearly identical, with minute differences in the 300 to 500 gram range for the fixed effects profiles. Thus, the two matching procedures lead to identical findings.

E. Results by Race and Education

Finally, we examine whether our findings vary by mother's race and education. Table 12 presents the cross-sectional and fixed effect estimates of the effect of birth weight on infant and neonatal mortality for babies of black and white mothers, and by mother's educational category (for white mothers only). In the first two sets of rows, the cross-sectional, birth weight-mortality correlation is about 50 percent larger in magnitude for infants of black women than for those of white women. However, the fixed effects estimates for black and white infants are similar

Figure 9 plots the cross-sectional and fixed effects birth weight profiles, separately for the infants of black and white mothers. Both sets of profiles are very similar by race. These results imply that most of the racial difference in the cross-sectional OLS estimates of the birth weight effects is due to the racial difference in birth weight distributions. The average birth weight of a black twin (2,191 grams in 1989-1991) is 280 grams less than the average weight of a white twin (2,471 grams). This and the strong nonlinearity of the mortality-birth weight relation in the cross-section results in OLS estimates that are markedly different by race.

The remaining rows of Table 12 show that the cross-section OLS estimates also differ by the educational attainment of the mother; but these differences fall substantially after mother fixed effects are included, especially for the sample that excludes twin pairs with congenital anomalies. There is still a difference in the birth weight coefficients between high school dropouts and high school graduates.

However, when we construct a figure analogous to Figure 9, we find cross-sectional and fixed effects birth weight profiles that are nearly identical by educational category. Thus, it again appears that differences in birth weight distributions across mother's education account for the differences in estimates that assume linear effects. We conclude that our findings are largely the same across different subpopulations of twins.

VI. Interpretation and Implications

Our primary finding is that in two very different settings – corresponding to policies that impact fetal nutrient intake and maternal smoking – controlling for mother-specific heterogeneity leads to dramatically smaller effects of birth weight on several infant outcomes of interest. In Almond, Chay, and Lee (2002), we find that the cross-sectional estimate of the effect of birth weight on infant mortality is largely insensitive to detailed controls for mother's observable characteristics – including age, education, race, marital status, medical and behavioral risk factors, and prenatal care histories. Together, these findings imply that unobserved factors, including genetics, lead to a severe overstatement of the importance of birth weight in cross-sectional analysis.

There are two plausible interpretations of our results. The first is that, due to mother-specific omitted variables, the payoff to LBW-prevention – in terms of cost savings or improvements in health – has been greatly exaggerated by the well-documented cross-sectional relationships. For example, consider the study of Lightwood et al. (1999), which provides an estimate of the direct medical costs associated with maternal smoking. For their costs estimate, they multiply two numbers: the excess costs at discharge associated with LBW babies, and the excess risk of LBW due to smoking during pregnancy.⁴⁴ The analysis yields an estimated \$263 million (in \$1995) of total hospital costs, per year, for the United States. Our analysis suggests that the cross-sectional estimates of excess costs due to LBW is

⁴⁴ This is the approach most commonly used in the literature to calculate costs. For example, see Marks et al (1990) and Windsor et al. (1993).

upwardly biased by a factor of 3.2, which implies that the \$263 million estimate overstates the “true” costs of maternal smoking by \$180 million.⁴⁵

The case of maternal smoking also highlights the policy relevance of the highly non-linear relations between birth weight and infant outcomes that we document in our analysis. For example, the second column of Table 7 suggests that maternal smoking does not affect the incidence of births below 1,500 grams. In this case, the \$263 million annual cost estimate is exaggerated by a factor of over four, and its overstatement of the true costs would total \$200 million per year. This sizeable reduction in the estimated economic benefits of smoking cessation among pregnant women has clear implications for the cost-effectiveness of different prenatal care interventions that target maternal smoking. It also has relevance for federal and state decisions on the optimal allocation of health care resources that improves infant health at minimal cost – for example, how to spend the U.S. states’ \$206 billion settlement with the tobacco industry, and the cost-effectiveness of Proposition 99 in California.

Since our analysis is non-experimental, we hesitate to conclude that it identifies the true “structural” relation between birth weight and early health outcomes and costs. However, we are confident in a second, more cautious interpretation of the findings – that the “impact” of LBW is highly dependent on which environmental factor is influencing birth weight. Cross-sectional birth weight variation is likely to be directly or indirectly influenced by immutable factors (genetics and race), socioeconomic factors (education, income levels), maternal behaviors (caloric intake, smoking behavior), and other environmental factors (intrauterine environment of the fetus). Within-twin pair variation – which represents 40 percent of the variability in birth weight due to variable intrauterine growth rates – holds constant all mother-specific and pregnancy-specific factors, and is partly generated by differences in fetal nutritional intake. Our maternal smoking analysis attempts to isolate birth weight variation due to smoking behavior during pregnancy – the leading modifiable cause of LBW in the United States.

⁴⁵ Our estimate of the excess costs due to LBW is the average of the rows in the second column of Table 7 (for the cross-section) (\$21,428) and in the third column (fixed effects) (\$6722), using the first column as weights. We do not use the (less than 600 gram) segment, since Lightwood et al. (1999) exclude these births from their analysis.

At a minimum, our findings indicate that there does not exist a uniform causal link between birth weight and hospital costs (ventilator use, APGAR score, and infant mortality). Some interventions may indeed succeed in both raising birth weights and improving health outcomes, while others may only be effective in raising birth weights, with little or no other effects on health. In other words, the most effective ways of preventing LBW may have little correspondence with the most effective ways of minimizing infant health complications and consequent health care costs.

If birth weight is a questionable target of policy, is there any other measure that could take its place? One possibility is to focus directly on medical costs, both during delivery and in the first several years of the child's life. Another possible target of policy – in the case that detailed cost data are unavailable – is the APGAR score. Another possibility is gestation length, or the incidence of premature birth. These latter two indicators are readily available on large-scale databases, such as those drawn from U.S. birth certificates.

More generally, other methods of infant health assessment may need to be developed. For example, Wilcox (2001) emphasizes that small preterm births with extremely low weights – less than 1000 grams – are much more informative for infant mortality than LBW births. Even though the fixed effects estimates in Figure 1C are much smaller than the cross-sectional estimates, there is nevertheless a substantive correlation between mortality and birth weight in this extremely low weight range. This suggests that interventions that impact the number of infants weighing less than 1000 grams could be effective in reducing infant mortality; however, their effects at the population level may still be small, since only one-half of one percent of all babies in the U.S. are born with a weight below 1000 grams.

VII. Directions for Future Research

We conclude by pointing out some important limitations in the scope of our analysis, and suggested areas of future research. First, while we have focused on two important influences of birth weight variation (nutritional factors and maternal smoking behavior), there are undoubtedly other factors that influence birth weight. Thus, we cannot rule out the possibility that the two sources of variation we

have examined are the exception and not the rule.⁴⁶ We suspect that this can only be resolved through an accumulation of evidence from studies that can control for confounding immutable factors, such as genetics. Mother-level unobserved heterogeneity appears to be an important confounder.

Second, our outcomes focus on the status of the infant at birth, or shortly after birth. Therefore, we cannot make any conclusions about the relation between LBW (or birth weight more generally) and longer run outcomes such as cognitive development, educational attainment, labor market outcomes, and adult health. In particular, our analysis focuses on prenatal investments in infant health, and not the interactions between birth weight, overall health, and postnatal investments in the child. The relatively small effects that we find for early infant health status suggest that previous studies' findings of significant impacts of birth weight on later adult outcomes is more likely to be driven by an interaction between postnatal investments and initial birth weight, rather than due to the effects of birth weight (or prenatal investments) *per se*.⁴⁷ A more direct investigation of this hypothesis seems to be a useful direction for future research.

⁴⁶ We note that if these are anomalies, they are not the only ones. For example, Chay and Greenstone (2003a, 2003b) find a strong association between sharp reductions in particulates air pollution across counties and declines in infant mortality, with only small effects on birth weight, or the incidence of LBW. Almond, Chay, and Greenstone (2003) document that the dramatic reduction in black infant mortality rates (IMR) in the United States from 1965 to 1971 accounts for the greatest convergence in black-white IMRs in the entire post-World War II era. However, they find a comparatively small change in the birth weight distribution of black infants relative to whites during this period.

⁴⁷ For example, women who quit smoking during pregnancy have high relapse rates following delivery (Fingerhut et al. 1990). The handful of studies that have used twin differences to estimate the effect of birth weight on adult health have found mixed evidence on its significance (e.g., Poulter et al. 1999, Ijzerman et al. 2000, Zhang et al. 2001, Hübinette et al. 2001). This is not surprising given the small numbers of twin pairs in these studies (132 to 492 twin pairs). Based on 404 monozygotic, female twin pairs, Behrman and Rosenzweig (2001) find an association between twin differences in birth weight and differences in educational attainment and wages. Interestingly, they find that the birth weight effects are larger at normal weights than at weights below 2,500 grams. Finally, Stein et al. (1975) find that the birth cohort born during the 1944-1945 Winter famine in Holland had largely similar adult outcomes to cohorts born before and after the famine.

References

- Almond, Douglas, Kenneth Y. Chay, and Michael Greenstone, "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in Mississippi," mimeograph, University of California, Berkeley, 2003.
- Almond, Douglas, Kenneth Y. Chay, and David S. Lee, "Does Birth Weight Matter? Evidence from the U.S. Population of Twin Births," Center for Labor Economics Working Paper No. 53, September 2002.
- Alexander, Duane "Child Health: Birth Outcomes and Infant Mortality," Speeches and Reports, National Institute for Child and Human Development.
- Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman, "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment," Journal of Political Economy, CII (1994), 1131-1174.
- Behrman, Jere R., and Mark R. Rosenzweig, "The Returns to Increasing Body Weight," PIER Working Paper No. 01-052, 2001.
- Brooks, Anne-Marie, et al., "Impact of Low Birth Weight on Early Childhood Asthma in the United States," Archives of Pediatric and Adolescent Medicine, 155 (2001), 401-406.
- Bryan, Elizabeth, Twins and Higher Order Births: A Guide to their Nature and Nurture, (London, UK: Edward Arnold, 1992)
- Chay, Kenneth Y., and Michael Greenstone, "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," Quarterly Journal of Economics, CXVIII (2003a), 1121-1167.
- _____, "Air Quality, Infant Mortality, and the Clean Air Act of 1970," National Bureau of Economics Research Working Paper, No. 10053 (2003b).
- Conley, Dalton, Kate W. Strully, Neil G. Bennett, The Starting Gate: Birth Weight and Life Chances (Berkeley, University of California Press, 2003).
- Currie, Janet, and Nancy Cole, "Welfare and Child Health: The Link Between AFDC Participation and Birth Weight," American Economic Review, LXXXIII (1993), 971-985.
- Currie, Janet, and Jonathan Gruber, "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," Journal of Political Economy, CIV (1996), 1263-1296.
- Currie, Janet and Rosemary Hyson, "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birth weight," AEA Papers and Proceedings, LXXXIX (1999), 245-250.
- Currie, Janet and Enrico Moretti, "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," CXVIII (2003), 1495-1532.
- Fingerhut, LA, JC Kleinman, and JS Kendrick, "Smoking Before, During, and After Pregnancy," American Journal of Public Health, 80 (1990), 541-544.

- Fowler, Mary Glenn, et al., "Double Jeopardy: Twin Infant Mortality in the United States, 1983 and 1984," American Journal of Obstetrics and Gynecology, CLXV (1991), 15-22.
- Geronimus, Arline T., and Sanders Korenman, "The Socioeconomic Consequences of Teen Childbearing Reconsidered," Quarterly Journal of Economics, CVII (1992), 1187-1214
- Goldenberg, Robert L. and Dwight J. Rouse, "Prevention of Premature Birth," New England Journal of Medicine, 339 (1998), 313-320.
- Grossman, Michael, and Theodore J. Joyce, "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City," Journal of Political Economy, XCVIII (1990), 983-1007.
- Hanratty, Maria J., "Canadian National Health Insurance and Infant Health," American Economics Review, LXXXVI (1996), 276-284.
- Hübinnette, A., "Birth Weight, Early Environment, and Genetics: A Study of Twins Discordant for Acute Myocardial Infarction," The Lancet, 357 (2001), 1997-2001.
- Ijzerman, Richard G., et al., "Evidence for Genetic Factors Explaining the Birth Weight-Blood Pressure Relation: Analysis in Twins," Hypertension, XXXVI (2000), 1008-1012.
- Institute of Medicine, WIC Nutrition Risk Criteria: A Scientific Assessment (Washington, D.C.: National Academy Press, 1996).
- Kaelber, Charles T., and Thomas F. Pugh, "Influence of Intrauterine Relations on the Intelligence of Twins," New England Journal of Medicine, CCLXXX (1969), 1030-1034.
- Kogan, M.D., G.R. Alexander, M. Kotelchuck, et al., "Trends in Twin Birth Outcomes and Prenatal Care Utilization in the United States, 1981-1997," Journal of the American Medical Association, 284 (2000), 335-341.
- Kowaleski-Jones, Lori and Greg J. Duncan, "Effects of Participation in the WIC Program on Birth weight: Evidence from the National Longitudinal Survey of Youth," American Journal of Public Health, 92 (2002), 799-804.
- Kramer, Michael S., "Intrauterine Growth and Gestational Duration Determinants," Pediatrics, 80 (1987), 502-511.
- Lewit, Eugene M., "The Direct Cost of Low Birth Weight," The Future of Children, V (1995), 35-56.
- Lightwood, James M., Ciaran S. Phibbs, and Stanton A. Glantz, "Short-term Health and Economic Benefits of Smoking Cessation: Low Birth Weight," Pediatrics, 104 (1999), 1312-1320.
- Lucas, A., R. Morley, and T.J. Cole, "Randomised Trial of Early Diet in Preterm Babies and Later Intelligence Quotient," British Medical Journal, 317 (1998), 1481-1487.
- Marks, JS, et al., "A Cost-Benefit/Cost Effectiveness Analysis of Smoking Cessation for Pregnant Women," American Journal of Preventative Medicine, 6 (1990), 282-289.

- Matte, T.D., M. Bresnahan, M.D. Begg, and E. Susser, "Influence of Variation in Birth Weight within Normal Range and within Sibships on IQ at Age 7 Years: Cohort Study," British Medical Journal, 323 (2001), 310-314
- McCormick, MC, J Brooks-Gunn, et al., "The Health and Developmental Status of Very Low-Birth weight Children at School Age," Journal of the American Medical Association, 267 (1992), 2204-2208.
- National Institutes of Health, "NIH Guide: Low Birth Weight in Minority Populations," PA-99-045, 1999.
- Nelson, K.B. and J.K. Grether, "Cerebral Palsy in Low-Birth weight Infants: Etiology and Strategies for Prevention," Mental Retardation and Developmental Disabilities Research Reviews, 3 (1997), 112-117.
- Paneth, Nigel S., "The Problem of Low Birth Weight," The Future of Children, V (1995), 19-34.
- Pollack, Harold, Paula M. Lantz, and John G. Frohna, "Maternal Smoking and Adverse Birth Outcomes Among Singletons and Twins," American Journal of Public Health, 90 (2000), 395-400.
- Poulter, N.R., et al., "Association between Birth Weight and Adult Blood Pressure in Twins: Historical Cohort Study," British Medical Journal, 319 (1999), 1330-1333.
- Richards, Marcus, et al., "Birth Weight and Cognitive Function in the British 1946 Birth Cohort: Longitudinal Population Based Study," British Medical Journal, CCCXXII (2001), 199-203.
- Rosenbaum, Paul and Donald Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," Biometrika, 70(1983), 41-55.
- Rosenzweig, Mark R., and T. Paul Schultz, "Estimating a Household Production Function: Heterogeneity and the Demand for Health Inputs, and Their Effects on Birth weight," Journal of Political Economy, XCI (1983), 723-746.
- Rosenzweig, Mark. R., and Kenneth I. Wolpin, "Inequality at Birth: The Scope for Policy Intervention," Journal of Econometrics, L (1991), 205-228.
- _____, "Sisters, Siblings, and Mothers: The Effects of Teen-age Childbearing on Birth Outcomes," Econometrica, LXIII (1995), 303-326.
- Sexton, Mary, and J. Richard Hebel, "A Clinical Trial of Change in Maternal Smoking and Its Effect on Birth Weight," Journal of the American Medical Association, CCLI (1984), 911-915.
- Stein, Z., M. Susser, G. Saenger, and F. Marolla, Famine and Human Development: The Dutch Hunger Winter of 1944-45 (New York, NY: Oxford University Press, 1975).
- Torelli, Paul A., "Maternal Smoking and Birth Outcomes: A Nonparametric Approach," mimeo, University of California, Berkeley, 2000.
- Vogler, George P., and Lynn T. Kozlowski, "Differential Influence of Maternal Smoking on Infant Birth Weight," Journal of the American Medical Association, 287 (2002), 241-242.

Wilcox, Allen J., “On the Importance – and the Unimportance – of Birth Weight,” International Journal of Epidemiology, 30 (2001), 1233-1241.

Windsor, RA, et al., “Health Education for Pregnant Smokers: Its Behavioral Impact and Cost Benefit,” American Journal of Public Health, 83 (1993), 201-206.

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

Data Appendix

Matching twins in linked birth/infant death data

Two approaches are used to match twins in the annual linked birth/infant death data. The primary approach uses the fact that beginning in 1989, multiple-birth records are generally located next to each other in the “denominator-plus” microdata file provided by NCHS. Additionally, with the rich set of covariates included on the microdata file, twin pairs can be matched when two unique strings of parental and pregnancy covariates are identified. This second approach is necessary for births occurring before 1989, when multiple births were not located next to each other in the “denominator-plus” file.

A. Record Location Matching Approach

The record location-based matching approach begins by selecting all “higher-order” pregnancies identified with the DPLURAL variable. For 1989, the DPLURAL variable indicates that 90,222 of the 4,045,881 births (2.2 percent) were twin births. The repetition of paternal and pregnancy characteristics indicate that adjacent twin records are part of the same twin set. Adjacent twin records are considered matched twins if the following covariates are each identical: mother’s age, state of birth, state of residence, race, and education, marital status of the parents, father’s age, father’s education⁴⁸, and the month in which prenatal care began. Of the 81,757 adjacent twin records that have both information on maternal education and birth weight, 74,472 (91%) have identical information for these nine covariates. Additional information on these files not used in matching suggests that incorrect twin matches are uncommon: of the 74,742 matched twin pairs for 1989, all but 37 of the pairs were born in the same month. While a longer string of covariates could be used to increase the likelihood that adjacent twin records are indeed part of the same twin set, incomplete or miscoding of the second twin record can cause the incorrect rejection of true twin matches.

This risk of Type 1 error needs to be balanced against the risk generated by relying more on the adjacency of twin records and less on the correspondence of parental and pregnancy covariates. Relying more on adjacency increases the risk of Type 2 error for several reasons. First, there appear to be singleton birth records that were coded as twins. (This could possibly result from the fact that the plurality field, DPLURAL, should be coded as “2” for twins, whereas “2” is often used as a code for “no”). It is also possible that one twin in the pair was not born alive, and therefore will appear in the fetal death data rather than in the “denominator-plus” natality data. Finally, it could be the case that while both twin birth records are present in the natality data, they are not adjacent in the “denominator-plus” file. All of these possibilities increase the risk of incorrectly matching adjacent twin records when a more sparse set of covariate checks is used.

B. Covariate Matching Approach

Before 1989, multiple birth records are not located together in the “denominator-plus” files. Therefore, an alternative matching algorithm is needed. A long string of parental and pregnancy covariates can instead be used to search across the entire set of twin records (i.e., where DPLURAL = 2) for a unique match. A string of eleven covariates was created that attempts to balance the risk of incorrectly matching twin pairs from different mothers (by including fewer covariates) against the risk of not matching correct twin pairs where there is incomplete or incorrect information for one of the twin pairs (by including more covariates in the matching). The eleven covariates used to match twins are: maternal education, maternal age, maternal race, marital status, mother’s state of birth, state of birth occurrence, county of birth

⁴⁸ Note for births occurring in 1995, father’s education is no longer collected and therefore this covariate is dropped from the matching algorithm.

occurrence, father's age, father's education, month of pregnancy that prenatal care began, and the number of prenatal care visits. Twins are considered matched when there are exactly two records with `DPLURAL = 2` and with equal values for this set of eleven covariates. For 1989, 74,886 twin pairs are matched, which is nearly identical to the number of pairs matched with the location-based method (74,472). Moreover, the econometric results using these two different matching techniques are nearly identical.

Finally, the birth order of the matched twins is not explicitly noted in the linked birth/infant death data. As pronounced health differences exist between first- and second-born twins, birth order should be allowed for in the econometric estimation. On the "denominator-plus" file, the `DTOTORD` variable gives the number of children born to the mother at the time of birth as well as the number of terminations of pregnancy. In most cases, `DTOTORD` differs by 1 for matched twin pairs. In these cases, the matched twin with the lower value for `DTOTORD` is considered the first-born twin.

Matching twins in HCUP state inpatient discharge data

The first step to matching twin records in the discharge data is to identify birth records where the newborn is part of a twin set. In contrast to the linked birth/infant death data, there is no dedicated field that identifies multiple births. In the hospital discharge data, the fields containing the diagnosis code `DX1` (principal diagnosis) and `DX2` (first secondary diagnosis) are used to select liveborn infants that are twins (ICD9-9-CM code equals `V31`, `V32`, or `V33`).

Like the NCHS data, the HCUP data do not provide information on the twin set to which individual twin records belong. Nor are twin discharge records linked to the maternal discharge record, which would permit use of information on the mother in identifying which twin records are part of a twin set. Nevertheless, the two basic approaches used to match twins in the linked birth/infant death data can be applied to the discharge records – i.e., using location of the discharge record in the `ascii` data and using a string of covariates that sibling twins should have in common. Note, the absence of information on the mother also means that the birth order of the twins cannot be imputed, as it was in the linked birth/infant death data (using the information on the number of children born to the mother).

A. Record Location Matching Approach

The `ascii` HCUP discharge data is sorted by a unique record identifier for each discharge record (not unique to each patient)⁴⁹. The basic record location approach checks whether twin records, approximately .3 percent of all discharge records, have a unique ID number that are in sequence. These adjacent twin records are then considered to be part of the same set.

Multiple matches of an individual twin discharge record can occur when it is both preceded and followed by other twin records. To determine which of the two potential matches for of this individual twin record to its neighbors is invalid (as is necessarily the case), information that should be common within valid twin sets is checked. This information includes the race of the twin, the zip code of residence of the mother, and the identity of the attending physician. If any of this information differs for an individual record with two potential neighboring matches, the match is rejected. Another reason more than two twin records might be contiguous is that triplets or quadruplets were miscoded as twin births. When more than two twin records are contiguous in the `ascii` data and also have identical information for race, zip code, and identity of the attending physician, these records are considered to be higher-order multiple births miscoded as twins and deleted.

⁴⁹ For 1998-2000, this variable is called `KEY`; for 1995-1997 it is called `SEQ_SID`

The record location matching approach matches approximately 72 percent of the twin records identified as twins using the diagnosis codes.

B. Covariate Matching Approach

The covariate approach to matching twins does not use any information on the position of records within the file. Instead, it uses a string of five variables that will be common within twin sets in order to isolate unique pairs. The variables used to match twins are: the zip code of the patients' residence, the HCUP hospital ID number from which the twin was discharged, the month the hospital admission occurred, whether the admission was on a weekend, and the race of the infant twin. Approximately 82 percent of discharge records with twin diagnosis codes can be uniquely matched with this approach.

If one is willing to assume that twins would also have the same attending physician in discharge records, then the ID number of the attending physician can be included in the covariate string. Approximately 76 percent of the records with twin diagnosis codes can be matched when attending physician is included in the variable string used for the covariate match.

Table 1: Summary of Basic Results for United States Twins

Outcome Variables	Coefficient on Birth Weight		
	Pooled	Fixed Effects	
		w/ congenital	w/o congenital
Hospital costs at discharge (in 2000 dollars)	-29.95 (0.84) [-0.506]	-4.93 (0.44) [-0.083]	--
Mortality, 1-year (per 1,000 births)	-0.1168 (0.0016) [-0.412]	-0.0222 (0.0016) [-0.078]	-0.0082 (0.0012) [-0.031]
Mortality, 28 days (per 1,000 births)	-0.1050 (0.0016) [-0.415]	-0.0154 (0.0013) [-0.061]	-0.0041 (0.0008) [-0.018]
5-min. APGAR score (0-10 scale, divided by 100)	0.1053 (0.0011) [0.506]	0.0117 (0.0012) [0.056]	0.0069 (0.0011) [0.034]
Ventilator incidence (per 1,000 births)	-0.0837 (0.0015) [-0.228]	-0.0039 (0.0017) [-0.011]	-0.0020 (0.0016) [-0.006]
Ventilator \geq 30 min. (per 1,000 births)	-0.0724 (0.0013) [-0.252]	0.0006 (0.0013) [0.002]	0.0016 (0.0012) [0.006]

Notes: The hospital cost data are from the 1995-2000 annual Healthcare Cost and Utilization Project (HCUP) State Inpatient Database for New York and New Jersey. All other data come from the National Center of Health Statistics 1989-1991 Linked Birth-Infant Death Detail Files. The Pooled column contains the OLS estimates based on the pooled sample of twins. The Fixed Effects columns include mother fixed effects in the regressions. The “w/ congenital” column includes twin pairs in which one or both twins either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The “w/o congenital” column excludes these twin pairs from the analysis. The standard errors are in parentheses and are corrected for heteroskedasticity. The Pooled results also correct the standard errors for within-twin-pair correlation in the residuals. The brackets contain the effects as measured in standard deviation units – i.e., the effect of a one standard deviation increase in birth weight (667 grams).

Table 2: Sample Means for Singletons and Twins in the 1989 Linked Natality-Mortality Detail File

	Sample Means		T-ratio for Diff In Sample Means
	Singletons	Twins	
<u>Mother's Demographic Characteristics</u>			
Fraction Black	0.195	0.219	14.1
Education	12.8	13.0	22.2
Fraction High School Dropout	0.184	0.157	17.3
Fraction College Graduate	0.187	0.212	15.2
Age	26.3	27.4	51.3
Fraction Teenager	0.129	0.078	45.8
Fraction 30 or Older	0.289	0.360	35.5
Fraction Married	0.736	0.746	5.1
<u>Mother's Risk Factors</u>			
Fraction Prenatal Care	0.982	0.982	0.4
Number of Prenatal Visits	11.2	12.4	49.7
Pregnancy Associated Hypertension	0.030	0.067	35.1
Anemia of Mother	0.018	0.035	21.0
Fraction Smoke during Pregnancy	0.212	0.201	5.4
<u>Characteristics of Birth</u>			
Fraction Male	0.512	0.504	4.1
Breech Birth	0.035	0.256	119.5
Abnormal Conditions of Newborn	0.055	0.148	62.9
Assisted Ventilation (<30 minutes)	0.011	0.024	20.4
Assisted Ventilation (>=30 minutes)	0.006	0.038	40.1
Congenital Anomaly	0.019	0.027	12.4
Sample Size	2,655,977	58,132	

Table 2 (cont'd)

	Sample Means		T-ratio for Diff In Sample Means
	Singletons	Twins	
<u>Infant Birth Weight (grams)</u>			
Mean	3,369 (591)	2,417 (670)	339.6
Median	3,402	2,495	
25 th percentile	3,060	2,070	
10 th percentile	2,693	1,503	
5 th percentile	2,410	1,080	
1 st percentile	1,430	503	
Fraction Low birth weight (<2500g)	0.061	0.504	
Gestation in weeks	39.3 (2.65)	36.0 (3.85)	201.2
1-minute APGAR score (0-10)	8.02 (1.35)	7.30 (1.97)	79.9
5-minute APGAR score (0-10)	8.97 (0.81)	8.52 (1.40)	70.2
<u>Infant Deaths (per 1,000 live births)</u>			
Within 1-year of birth (infant mortality)	8.46	38.71	37.7
Within 24 hours of birth	2.73	19.27	29.0
Within 7 days	3.99	26.92	34.1
Within 28 days (neonatal)	4.99	30.62	35.8
28 days to 1 year (postneonatal)	3.49	8.19	12.5
<u>Fraction of dead with birth weight<2500 g</u>			
Infant mortality	0.57	0.93	
Within 24-hour mortality	0.89	0.98	
Neonatal mortality	0.76	0.97	
Postneonatal mortality	0.30	0.79	
Sample Size	2,655,977	58,132	

Notes: Data come from the National Center of Health Statistics 1989 Linked Birth-Infant Death Detail File. The sample is restricted to non-Hispanic, black and white mothers born in the United States. The standard deviations of the means are in parentheses.

Table 3: Components of Variance for Birth Weight and Outcomes among Twins

Dependent Variable	Mean Squared Error in OLS Regressions				Ratio
	(1)	(2)	(3)	(4)	(4)/(3)
<u>1989-1991 U.S. Twins</u>					
Birth weight	44.434	21.307	19.080	7.535	0.40
Mortality (1-year)	0.0356	0.0287	0.0219	0.0149	0.68
Mortality (1-day)	0.0183	0.0152	0.0102	0.0046	0.45
Mortality (28-day)	0.0283	0.0224	0.0158	0.0090	0.57
5-min. APGAR	1.9254	1.4078	1.1744	0.6510	0.55
Ventilator \geq 30 min.	0.0370	0.0348	0.0338	0.0102	0.30
<u>1995-2000 NY-NJ Twins</u>					
Hospital costs	14.410	--	--	2.958	--
Controls for					
Gestation length (linear)	No	Yes	--	--	
Gestation length dummies	No	No	Yes	--	
Mother fixed effects	No	No	No	Yes	

Notes: See notes to Table 1. Columns (1) – (4) provide the means squared error from OLS regressions that include no controls, a linear control for gestation length (in weeks), gestation length fixed effects, and mother fixed effects, respectively. The final column provides the ratio of Column (4) to Column (3) – that is, the fraction of overall variation in outcomes, for fixed gestation lengths, that is due to within-twin-pair differences instead of between-twin-pair differences. The hospital costs data do not contain gestation lengths. Birth weight is measured in 100's of grams, and hospital costs are in 10,000's of dollars. The sample size for birth weight, mortality, and assisted ventilation is 187,948. The sample size for 5-minute APGAR score is 158,700, and the sample size for hospital costs is 44,410.

Table 4: Sample Means for Twin Pairs in 1989-1991

	<u>Heavier versus Lighter Twin</u>			<u>First versus Second Born Twin</u>		
	<u>Sample Means</u>		T-ratio for Difference	<u>Sample Means</u>		T-ratio for Difference
	Heavy	Light		First	Second	
Birth weight	2,563.3	2,263.6	98.7	2,422.5	2,388.0	10.7
Birth weight < 2500 grams	0.399	0.620	-96.2	0.499	0.526	-10.9
Birth weight < 1500 grams	0.075	0.114	-28.5	0.095	0.101	-4.14
Birth weight < 1000 grams	0.035	0.050	-15.2	0.043	0.045	-2.32
Male	0.548	0.460	37.8	0.507	0.503	1.68
Congenital anomaly	0.0237	0.0275	-5.18	0.0256	0.0256	0.02
Breech birth	0.249	0.258	-4.27	0.214	0.294	-37.4
Abnormal conditions	0.154	0.157	-2.03	0.151	0.165	-8.27
Ventilation < 30 min.	0.0249	0.0266	-2.35	0.0233	0.0290	-7.29
Ventilation >= 30 min.	0.0384	0.0377	0.76	0.0364	0.0418	-5.79
1-minute APGAR	7.39	7.29	10.14	7.54	7.09	43.5
5-minute APGAR	8.56	8.53	4.44	8.59	8.47	16.4
5-min. APGAR <= 8	0.264	0.270	-3.00	0.242	0.300	-24.4
<u>Infant Deaths (per 1,000 births)</u>						
<u>Including Congenital Anomalies</u>						
Within 1-year of birth	32.4	38.4	-6.84	35.7	37.9	-2.40
Within 24 hours	16.3	18.5	-3.48	18.0	18.2	-0.25
Within 28 days	25.6	29.6	-5.15	27.8	29.7	-2.41
28 days to 1 year	6.84	8.83	-4.78	7.90	8.14	-0.54
<u>Excluding Congenital Anomalies</u>						
Within 1-year of birth	28.1	30.7	-3.19	29.8	31.4	-1.83
Within 24 hours	14.3	14.8	-0.97	15.1	15.1	0.10
Within 28 days	22.0	23.2	-1.69	22.9	24.4	-1.96
28 days to 1 year	6.13	7.52	-3.51	6.94	7.02	-0.21
Sample Size	90,179	90,179		85,142	85,142	

Notes: See notes to Table 2. Data are from the 1989-1991 Linked Birth-Infant Death Detail Files. The first set of columns present sample means separately for the heavier and lighter twin across twin pairs. The second set of columns present sample means separately for the first- and second-born twin across twin pairs. Observations missing data on birth order are dropped from this sample.

Table 5: Pooled OLS and Twins Fixed Effects Estimates of the Effect of Birth Weight

Birth weight coefficient	<u>Including Congenital Anomalies</u>		<u>Excluding Congenital Anomalies</u>	
	Pooled OLS	Fixed Effects	Pooled OLS	Fixed Effects
<u>Mortality, 1-year</u>	-0.1168 (0.0016)	-0.0222 (0.0016)	-0.1069 (0.0017)	-0.0082 (0.0012)
Adj. R-squared	0.169	0.585	0.164	0.629
Sample Size	189,036	189,036	183,727	183,727
<u>Mortality, 1-day</u>	-0.0739 (0.0015)	-0.0071 (0.0010)	-0.0675 (0.0015)	-0.0003 (0.0006)
Adj. R-squared	0.132	0.752	0.127	0.809
Sample Size	189,036	189,036	183,727	183,727
<u>Mortality, neonatal</u>	-0.1050 (0.0016)	-0.0154 (0.0013)	-0.0962 (0.0016)	-0.0041 (0.0008)
Adj. R-squared	0.173	0.683	0.169	0.745
Sample Size	189,036	189,036	183,727	183,727
<u>Mortality, postneonatal</u>	-0.0117 (0.0005)	-0.0069 (0.0010)	-0.0107 (0.0005)	-0.0042 (0.0009)
Adj. R-squared	0.008	0.059	0.007	0.067
Sample Size	189,036	189,036	183,727	183,727
<u>5-min. APGAR score</u>	0.1053 (0.0011)	0.0117 (0.0012)	0.1009 (0.0011)	0.0069 (0.0011)
Adj. R-squared	0.255	0.663	0.248	0.673
Sample Size	159,070	159,070	154,449	154,449
<u>Ventilator incidence</u>	-0.0837 (0.0015)	-0.0039 (0.0017)	-0.0810 (0.0015)	-0.0020 (0.0016)
Adj. R-squared	0.052	0.706	0.050	0.716
Sample Size	189,036	189,036	183,727	183,727
<u>Ventilator >= 30 min.</u>	-0.0724 (0.0013)	0.0006 (0.0013)	-0.0701 (0.0014)	0.0016 (0.0012)
Adj. R-squared	0.063	0.724	0.062	0.739
Sample Size	189,036	189,036	183,727	183,727
<u>Hospital Costs</u>	-29.95 (0.84)	-4.93 (0.44)	--	--
Adj. R-squared	0.256	0.796		
Sample Size	44,410	44,410		

Notes: See notes to Tables 1 and 3. The data come from the 1989-1991 Linked Birth-Infant Death Detail Files and the 1995-2000 HCUP Inpatient Database for New York and New Jersey. The first two columns use samples that include twin pairs in which one or both twins either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The second two columns exclude these twin pairs from the analysis. The HCUP data does not contain information on congenital anomalies. The standard errors are in parentheses and are corrected for heteroskedasticity and within-twin-pair correlation in the residuals. For mortality and ventilator use the birth weight coefficients have been scaled up by 1,000. For APGAR score, the coefficients are scaled up by 100.

Table 6: Piece-wise Linear Spline Estimates of the Effect of Birth Weight on Hospital Costs and Infant Health

Dependent Variable		Hospital Costs		Ventilator (any)		Ventilator (> 30 min.)		APGAR (5-minute)		Mortality (1-year)	
Units		Dollars		Usage per 1000 births		Usage per 1000 births		Score times 1000		Deaths per 1000 births	
		per gram		per gram		per gram		per gram		per gram	
Birth Weight segment	Percent	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.
< 600g	0.18	428.09 (39.56)	247.79 (99.35)	0.7995 (0.0640)	0.2718 (0.1012)	0.7434 (0.0578)	0.2361 (0.0969)	12.26 (0.39)	4.41 (1.02)	-0.5754 (0.0640)	-0.4400 (0.1560)
600-800 g	0.16	186.59 (62.03)	-270.30 (133.79)	0.3027 (0.1022)	-0.0052 (0.0914)	0.2369 (0.0970)	0.0261 (0.0891)	9.38 (0.58)	2.19 (1.06)	-2.1122 (0.0913)	-0.4197 (0.2139)
800-1000 g	0.19	-212.77 (47.89)	-49.75 (67.31)	-0.2164 (0.0863)	-0.0772 (0.0797)	-0.2136 (0.0815)	-0.0549 (0.0792)	3.04 (0.40)	0.16 (0.81)	-1.1302 (0.0752)	-0.4332 (0.1669)
1000-1500 g	0.55	-121.20 (7.57)	-36.88 (9.89)	-0.2216 (0.0218)	-0.0005 (0.0243)	-0.2124 (0.0201)	0.0168 (0.0232)	2.12 (0.09)	0.60 (0.22)	-0.2258 (0.0144)	-0.1693 (0.0336)
1500-2000 g	1.11	-74.27 (2.50)	-24.07 (2.95)	-0.2121 (0.0110)	-0.0063 (0.0137)	-0.1891 (0.0096)	-0.0018 (0.0118)	0.98 (0.04)	0.18 (0.09)	-0.0279 (0.0051)	-0.0391 (0.0123)
2000-2500 g	3.95	-18.19 (0.84)	-3.21 (0.89)	-0.0671 (0.0052)	-0.0091 (0.0067)	-0.0469 (0.0038)	0.0008 (0.0050)	0.51 (0.02)	0.14 (0.04)	-0.0146 (0.0022)	-0.0113 (0.0051)
2500-3000 g	15.54	-5.19 (1.45)	0.58 (0.56)	-0.0206 (0.0034)	-0.0009 (0.0050)	-0.0128 (0.0020)	0.0014 (0.0032)	0.27 (0.02)	0.05 (0.03)	-0.0076 (0.0020)	-0.0043 (0.0034)
> 3000 g	78.31	6.66 (3.31)	1.37 (0.83)	-0.0035 (0.0033)	-0.0012 (0.0053)	-0.0027 (0.0016)	-0.0033 (0.0031)	-0.02 (0.02)	-0.02 (0.03)	0.0120 (0.0038)	-0.0019 (0.0036)
Observations		44410	44410	189036	189036	189036	189036	159070	159070	189036	189036

Notes: Costs data come from 1995-2000 NY/NJ HCUP data. The remaining columns use the 1989-1991 linked Birth-Infant Death files. Entries are the slope coefficients within each birth-weight segment from piecewise linear spline specifications. The Percent column contains the percentages of all singleton births in 1989 with birth weights in the corresponding birth weight segment. Pooled is ordinary least squares applied to the pooled data. The F.E. columns include mother fixed effects. Standard errors are in parentheses and are corrected for both within-twin-pair correlation and heteroskedasticity.

Table 7: Implied Excess Hospital Costs Associated with Low Birth Weight,
Pooled versus Fixed Effects Estimates

Birth Weight Segment	Percent of Singleton Births	Excess Hospital Costs	
		Pooled	Fixed Effects
< 600g	0.18	\$51,803	\$61,213
600-800 g	0.16	\$131,589	\$67,816
800-1000 g	0.19	\$127,190	\$36,846
1000-1500 g	0.55	\$74,706	\$22,309
1500-2000 g	1.11	\$25,137	\$6,806
2000-2500 g	3.95	\$3,417	\$604

Notes: The entries represent the average reduced hospital costs associated with increasing an infant's birth weight from the given birth weight category to above 2,500 grams. These are calculated using the estimated spline coefficients in Table 6. The Percent of Singleton Births column provides the percentages of all singleton births in 1989 with birth weights in the corresponding birth weight segment. The costs figures are in year 2000 dollars.

Table 8: OLS Estimates of Effect of Maternal Smoking during Pregnancy,
Pennsylvania Singletons 1989-1991

	Coefficient on Smoking Indicator	
	Unadjusted	Adjusted
<u>Impact on Birth Weight</u>		
Birth weight (in grams)	-284.8 (2.07)	-203.2 (2.11)
Birth weight < 2500 grams (per 1,000 births)	64.0 (1.04)	34.7 (1.03)
Birth weight < 2000 grams (per 1,000 births)	19.93 (0.63)	6.30 (0.62)
Birth weight < 1500 grams (per 1,000 births)	8.58 (0.44)	0.59 (0.44)
Birth weight < 1000 grams (per 1,000 births)	3.99 (0.31)	-0.44 (0.32)
<u>Impact on Other Outcomes</u>		
Gestation length (in weeks)	-0.328 (0.010)	-0.079 (0.010)
Premature Birth (< 32 weeks) (per 1,000 births)	13.05 (0.55)	-0.27 (0.54)
5-min. APGAR score (0-10 scale)	-0.067 (0.003)	-0.010 (0.003)
Ventilator incidence (per 1,000 births)	6.28 (0.45)	1.20 (0.48)
Ventilator >= 30 min. (per 1,000 births)	2.61 (0.29)	-0.11 (0.30)
Mortality, 1-year (per 1,000 births)	4.57 (0.36)	0.10 (0.38)
Mortality, 28 days (per 1,000 births)	1.76 (0.28)	-1.16 (0.30)

Notes: The data come from the 1989-1991 Linked National Natality-Mortality Detail Files. The sample consists of the 497,139 singletons born in Pennsylvania between 1989 and 1991. The table entries are the coefficient estimates on an indicator equal to one if the mother smoked during pregnancy. The Unadjusted column presents the raw difference in sample means between smoking and nonsmoking mothers. The Adjusted column presents the coefficient estimates from a regression that includes all of the “pre-treatment” variables included in the propensity score model in Figures 3 and 4 (see Appendix). The standard errors are in parentheses and are corrected for heteroskedasticity.

Table 9: Estimated Effects of Birth Weight and 5-Minute APGAR Score on Mortality and Assisted Ventilation

	<u>Mortality within 1-Year</u>			<u>Mortality within 1-Day</u>			<u>Ventilator Incidence</u>			<u>Ventilator >= 30 Minutes</u>		
	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca
Birth weight	-0.0511 (0.0013) [-40.52]	-0.0188 (0.0017) [-11.31]	-0.0072 (0.0012) [-5.85]	-0.0161 (0.0008) [-19.54]	-0.0048 (0.0010) [-4.89]	-0.0004 (0.0007) [0.68]	-0.0626 (0.0017) [-36.28]	-0.0010 (0.0019) [-0.55]	-0.0006 (0.0018) [-0.36]	-0.0587 (0.0015) [-39.33]	0.0026 (0.0014) [1.81]	0.0026 (0.0013) [2.02]
5-Min. APGAR	-0.0623 (0.0009) [-72.02]	-0.0227 (0.0012) [-19.38]	-0.0108 (0.0010) [-10.43]	-0.0547 (0.0009) [-58.22]	-0.0186 (0.0010) [-18.51]	-0.0094 (0.0008) [-11.56]	-0.0244 (0.0010) [-25.32]	-0.0261 (0.0011) [-23.13]	-0.0271 (0.0012) [-23.28]	-0.0168 (0.0008) [-20.72]	-0.0145 (0.0010) [-14.95]	-0.0148 (0.0010) [-15.19]
Sample Size	159,070	159,070	154,449	159,070	159,070	154,449	159,070	159,070	154,449	159,070	159,070	154,449

Notes: The data are for twins from the 1989-1991 Linked Natality-Mortality Detail Files. Observations with missing data on either birth weight or APGAR scores are excluded from the sample. The “F.E., nca” column further excludes twins that either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The birth weight coefficients are scaled up by 1,000 (effect of 1,000 gram increase). The standard errors are in parentheses and have been corrected for heteroskedasticity and within-twin-pair correlation. The t-ratios of the coefficients are in brackets.

Table 10: Fixed Effects Estimates of Birth Weight Effect for Same-Sex and Boy-Girl Twin Pairs

Birth weight coefficient	Both Girl Twins			Both Boy Twins			Boy-Girl Twins		
	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca
Mortality, 1-year	-0.1066 (0.0029)	-0.0226 (0.0027)	-0.0085 (0.0019)	-0.1352 (0.0028)	-0.0272 (0.0031)	-0.0107 (0.0023)	-0.1079 (0.0029)	-0.0222 (0.0028)	-0.0096 (0.0018)
Mortality, 1-day	-0.0687 (0.0026)	-0.0060 (0.0015)	-0.0001 (0.0007)	-0.0842 (0.0025)	-0.0110 (0.0020)	-0.0019 (0.0013)	-0.0684 (0.0027)	-0.0058 (0.0021)	-0.0004 (0.0012)
Mortality, neonatal	-0.0962 (0.0028)	-0.0166 (0.0022)	-0.0060 (0.0013)	-0.1213 (0.0027)	-0.0198 (0.0026)	-0.0058 (0.0016)	-0.0969 (0.0029)	-0.0135 (0.0022)	-0.0032 (0.0012)
Control for Gender	--	--	--	--	--	--	Yes	Yes	Yes
Sample Size	64,056	64,056	62,335	65,672	65,672	63,481	59,308	59,308	57,911

Notes: The data are from the 1989-1991 Linked Natality-Mortality Detail Files. The “F.E., nca” column excludes twins that either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The birth weight coefficients are scaled up by 1,000. The standard errors are in parentheses and have been corrected for heteroskedasticity and within-twin-pair correlation.

Table 11: Twins Pooled and Fixed Effects Estimates for 1983-1985, 1989-1991, and 1995-1997

	<u>Mortality within One Year</u>			<u>Mortality within 28 Days</u>		
	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca
<u>1983-1985</u>						
Birth Weight	-0.1375 (0.0019)	-0.0177 (0.0019)	-0.0089 (0.0016)	-0.1234 (0.0019)	-0.0118 (0.0016)	-0.0040 (0.0013)
Dep. var. mean	47.0	47.0	43.1	37.6	37.6	34.5
Adj. R-squared	0.199	0.620	0.656	0.198	0.702	0.744
Sample Size	140,388	140,388	139,811	140,388	140,388	139,811
<u>1989-1991</u>						
Birth Weight	-0.1168 (0.0016)	-0.0222 (0.0016)	-0.0082 (0.0012)	-0.1050 (0.0016)	-0.0154 (0.0013)	-0.0041 (0.0008)
Dep. var. mean	37.3	37.3	31.3	29.3	29.3	24.4
Adj. R-squared	0.169	0.585	0.629	0.173	0.683	0.745
Sample Size	189,036	189,036	183,727	189,036	189,036	183,727
<u>1995-1997</u>						
Birth Weight	-0.0907 (0.0015)	-0.0177 (0.0014)	-0.0057 (0.0010)	-0.0815 (0.0014)	-0.0138 (0.0011)	-0.0034 (0.0007)
Dep. var. mean	27.2	27.2	22.6	21.8	21.8	18.0
Adj. R-squared	0.133	0.554	0.615	0.133	0.633	0.713
Sample Size	214,168	214,168	209,155	214,168	214,168	209,155

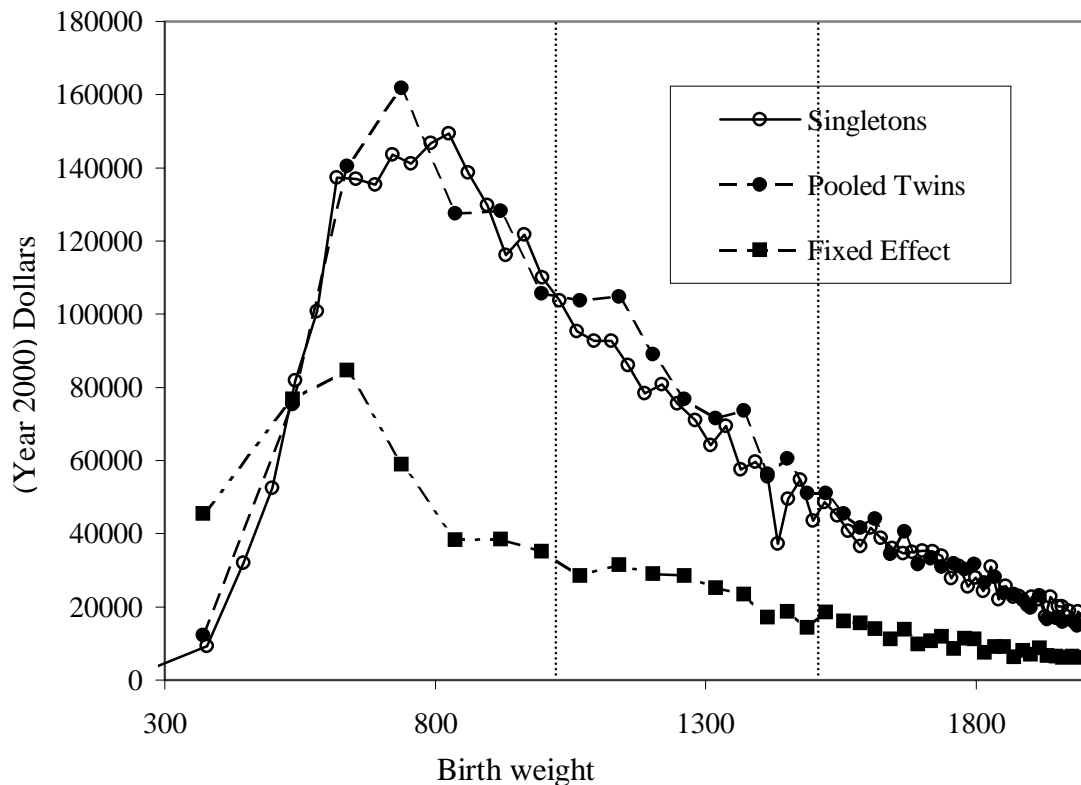
Notes: The data are from the 1983-1985, 1989-1991, and 1995-1997 Linked Natality-Mortality Detail Files. The "F.E., nca" column excludes twins that either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The birth weight coefficients are scaled up by 1,000. The standard errors are in parentheses and have been corrected for heteroskedasticity and within-twin-pair correlation.

Table 12: Twins Pooled and Fixed Effects Estimates by Mother's Race and Education

Birth weight coefficient	Mortality within One Year			Mortality within 28 Days		
	Pooled	F.E.	F.E., nca	Pooled	F.E.	F.E., nca
White	-0.1035 (0.0019)	-0.0212 (0.0018)	-0.0066 (0.0012)	-0.0946 (0.0019)	-0.0155 (0.0014)	-0.0037 (0.0008)
Black	-0.1558 (0.0034)	-0.0263 (0.0041)	-0.0146 (0.0033)	-0.1383 (0.0034)	-0.0147 (0.0032)	-0.0056 (0.0020)
<u>White Education</u>						
HS Dropout	-0.1318 (0.0055)	-0.0381 (0.0066)	-0.0162 (0.0052)	-0.1173 (0.0054)	-0.0278 (0.0050)	-0.0072 (0.0032)
HS Graduate	-0.1080 (0.0031)	-0.0281 (0.0032)	-0.0083 (0.0021)	-0.0987 (0.0031)	-0.0197 (0.0026)	-0.0045 (0.0014)
Some College	-0.1002 (0.0039)	-0.0099 (0.0033)	-0.0008 (0.0022)	-0.0940 (0.0039)	-0.0078 (0.0026)	-0.0004 (0.0016)
College Graduate	-0.0853 (0.0037)	-0.0148 (0.0026)	-0.0056 (0.0016)	-0.0790 (0.0036)	-0.0118 (0.0022)	-0.0041 (0.0012)

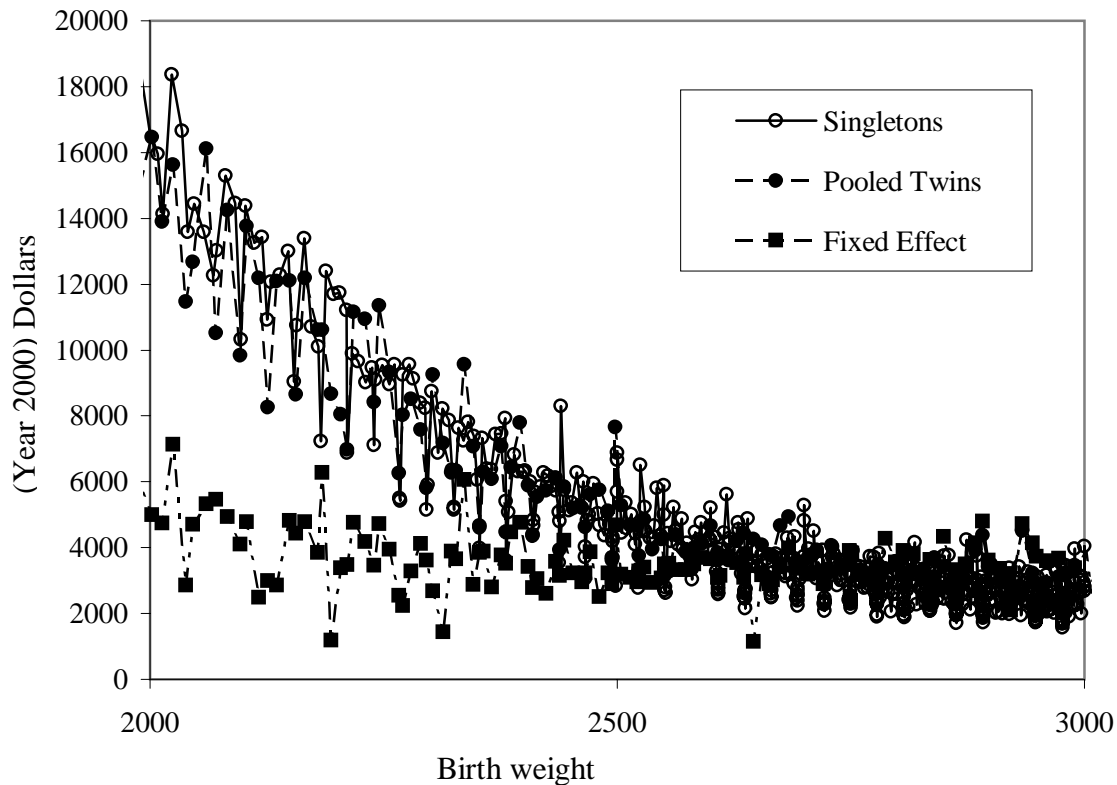
Notes: The data are from the 1989-1991 Linked Natality-Mortality Detail Files. The "F.E., nca" column excludes twins that either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The birth weight coefficients are scaled up by 1,000. The standard errors are in parentheses and have been corrected for heteroskedasticity and within-twin-pair correlation.

Figure 1A: Hospital Costs and Birth Weight, under 2000 grams



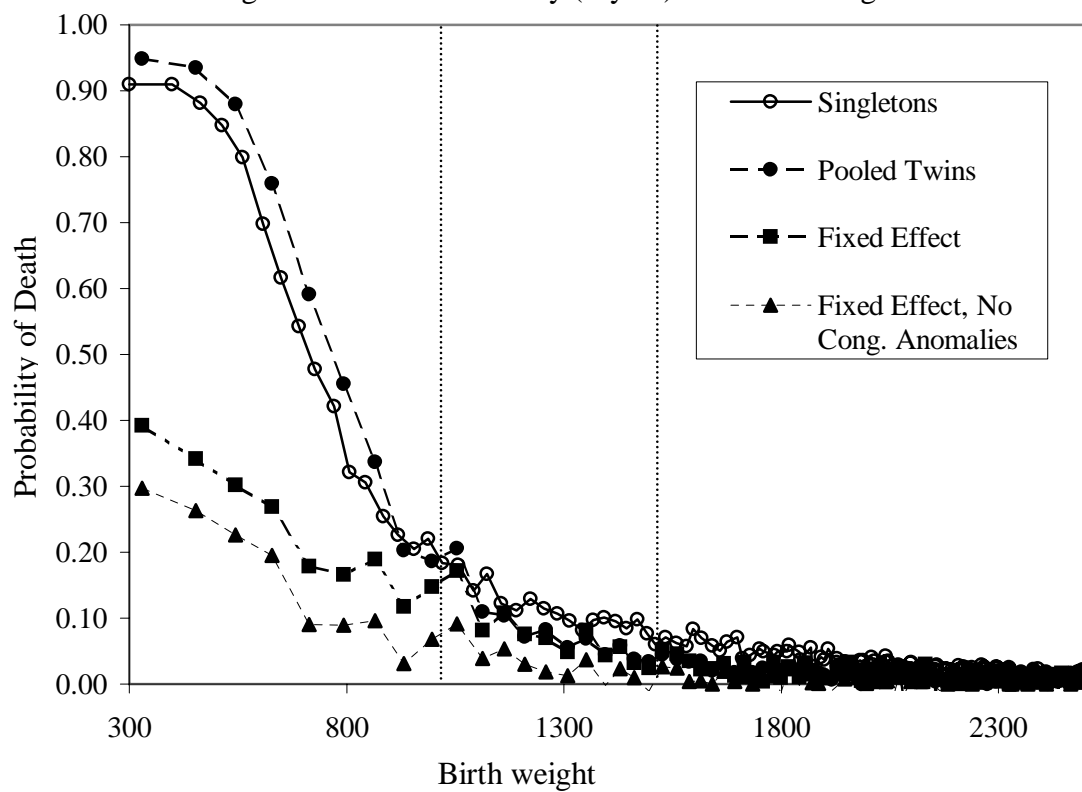
Note: 1995-2000 NY/NJ Hospital Discharge Microdata

Figure 1B: Hospital Costs and Birth Weight, 2000-3000 grams



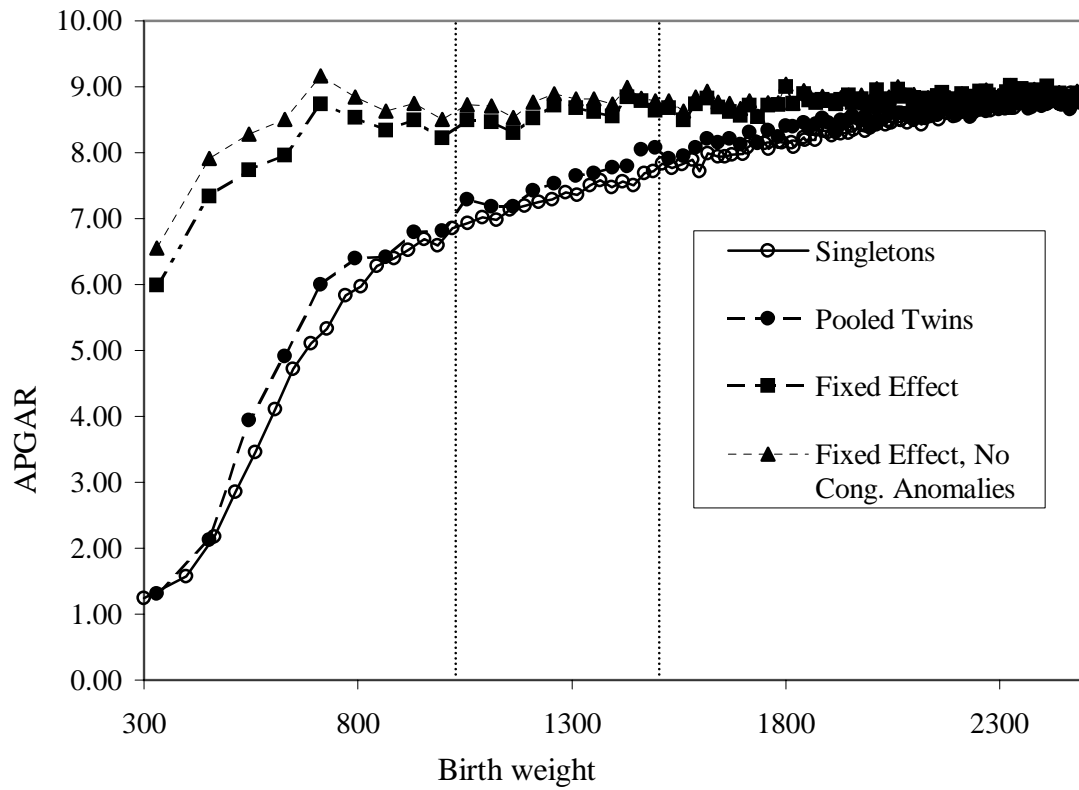
Note: 1995-2000 NY/NJ Hospital Discharge Microdata

Figure 1C: Infant Mortality (1-year) and Birth Weight



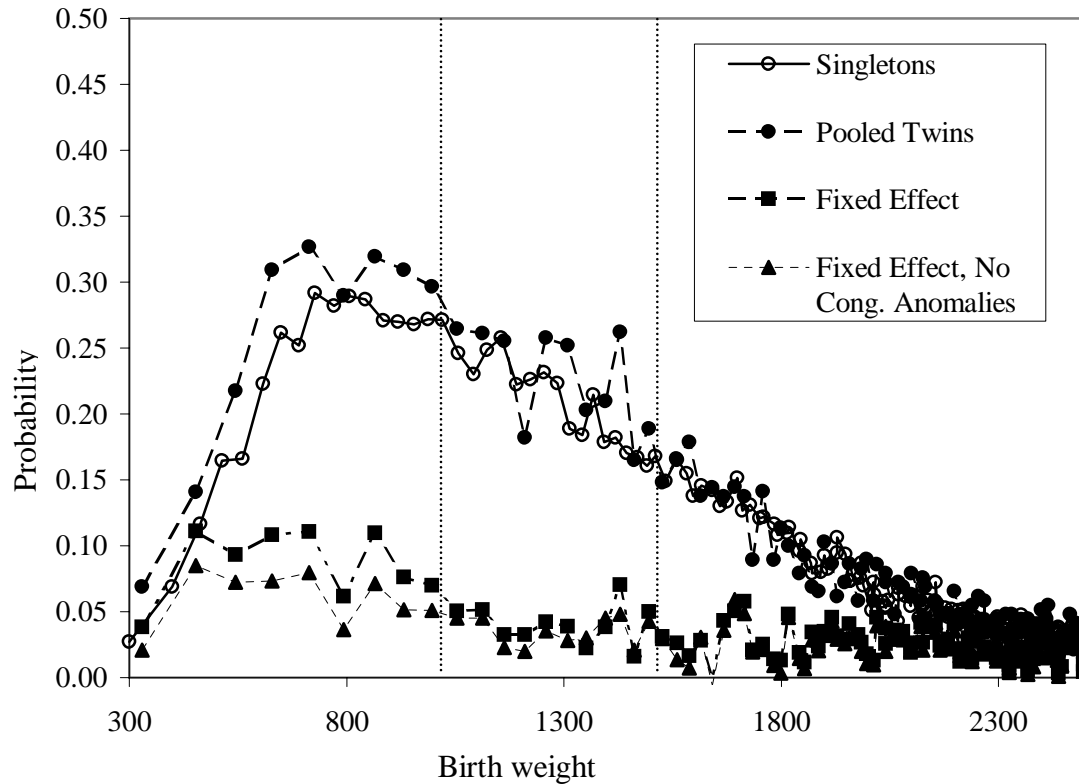
Note: Linked Birth-Death certificate data, 1989

Figure 2A: 5-minute APGAR score and Birth Weight



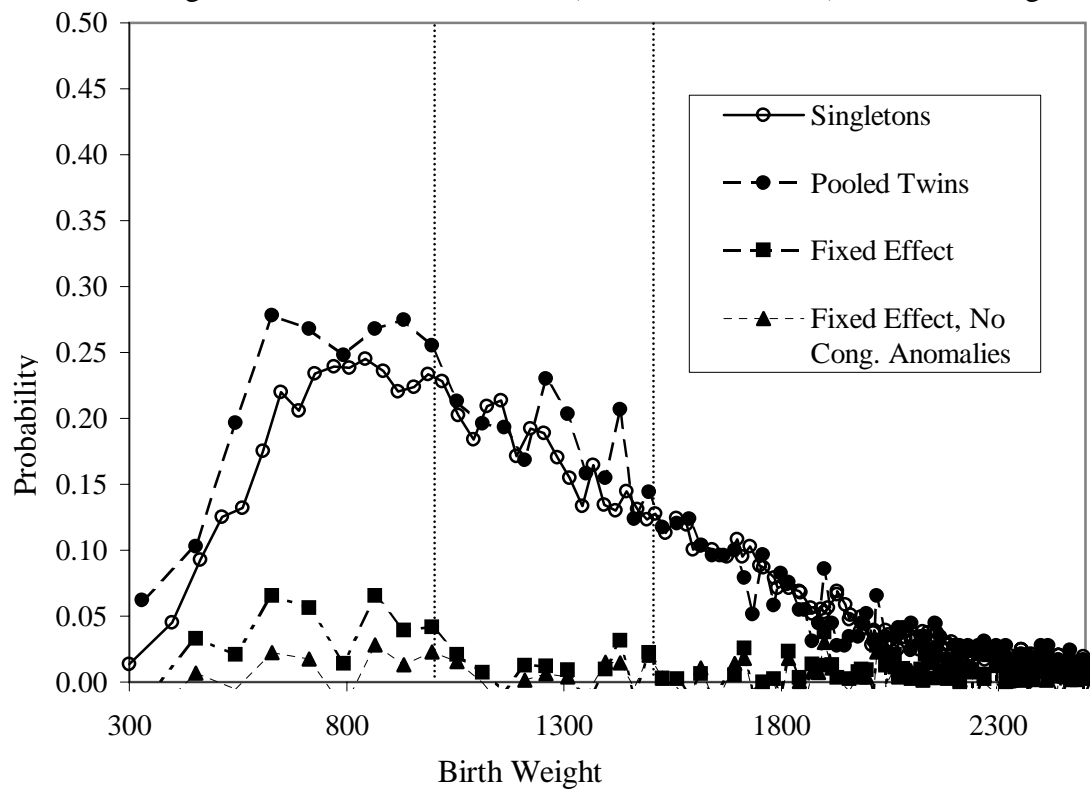
Note: Birth certificate data, 1989

Figure 2B: Assisted Ventilation and Birth Weight



Note: Birth certificate data, 1989

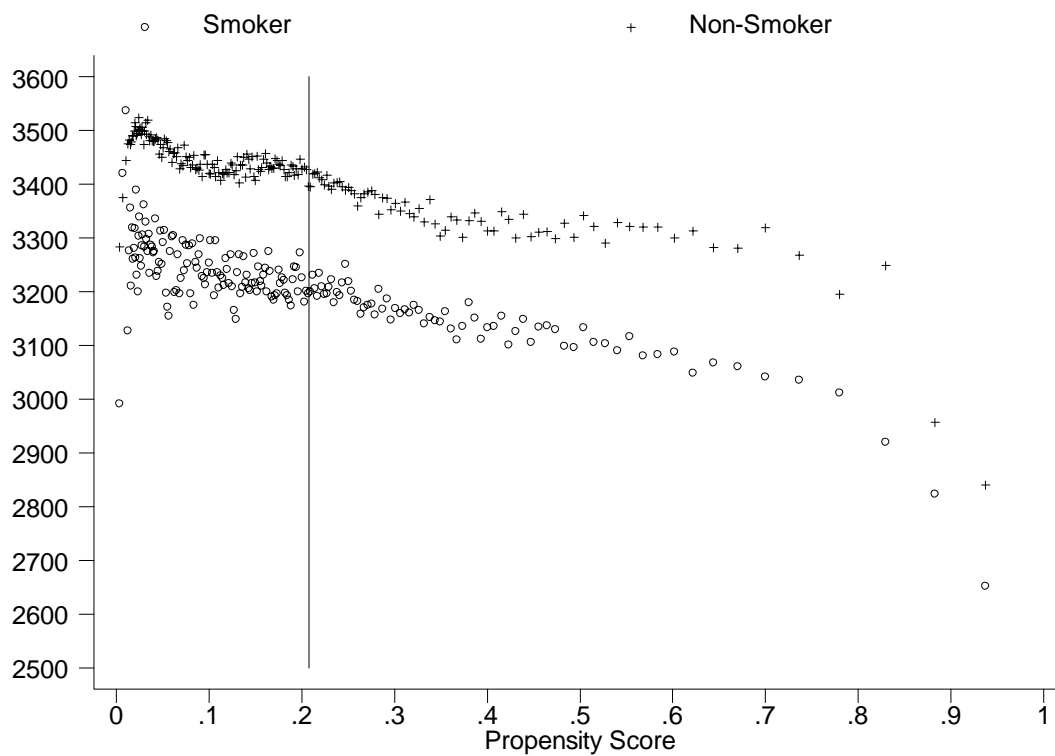
Figure 2C: Assisted Ventilation (30 minutes or more) and Birth Weight



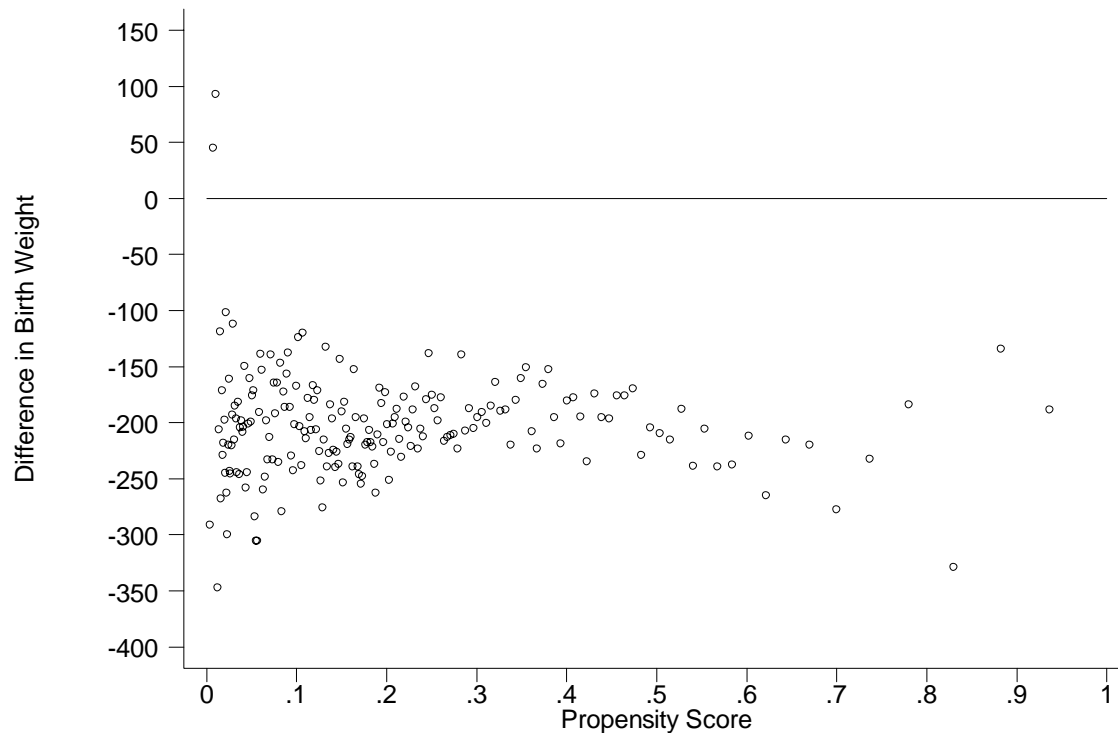
Note: Birth certificate data, 1989

Figure 3: Estimated Birth Weight Effects of Maternal Smoking by the Propensity Score, Pennsylvania Singletons 1989-1991

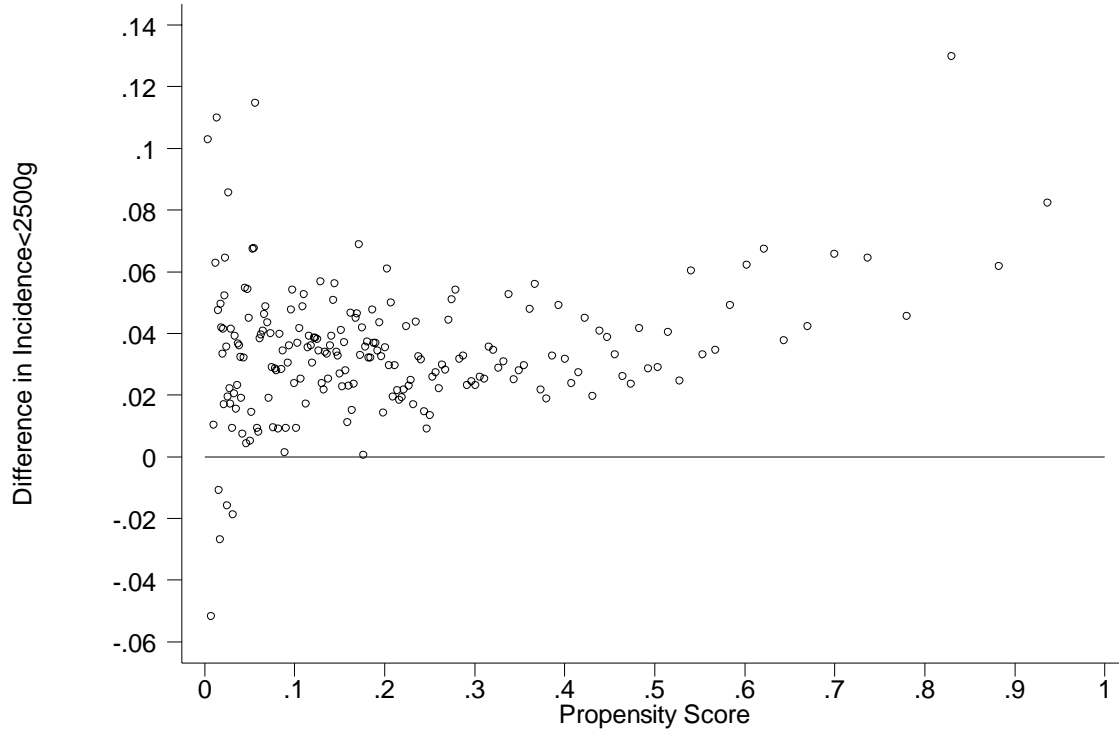
A. Average Birth Weight of Infants of Smoking and Non-Smoking Mothers by the Propensity Score



B. Difference between Smokers and Non-Smokers in Average Birth Weight by the Propensity Score



C. Difference between Smokers and Non-Smokers in Low Birth Weight (<2500 grams) Incidence



D. Difference between Smokers and Non-Smokers in Incidence less than 1000 grams

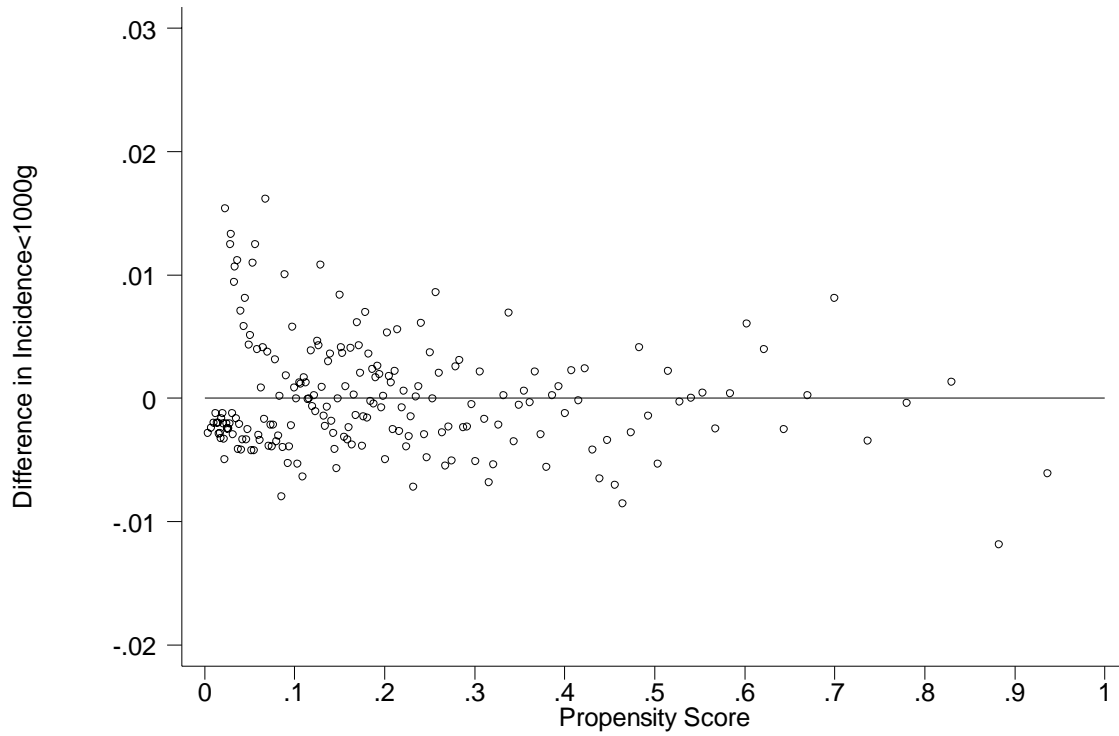
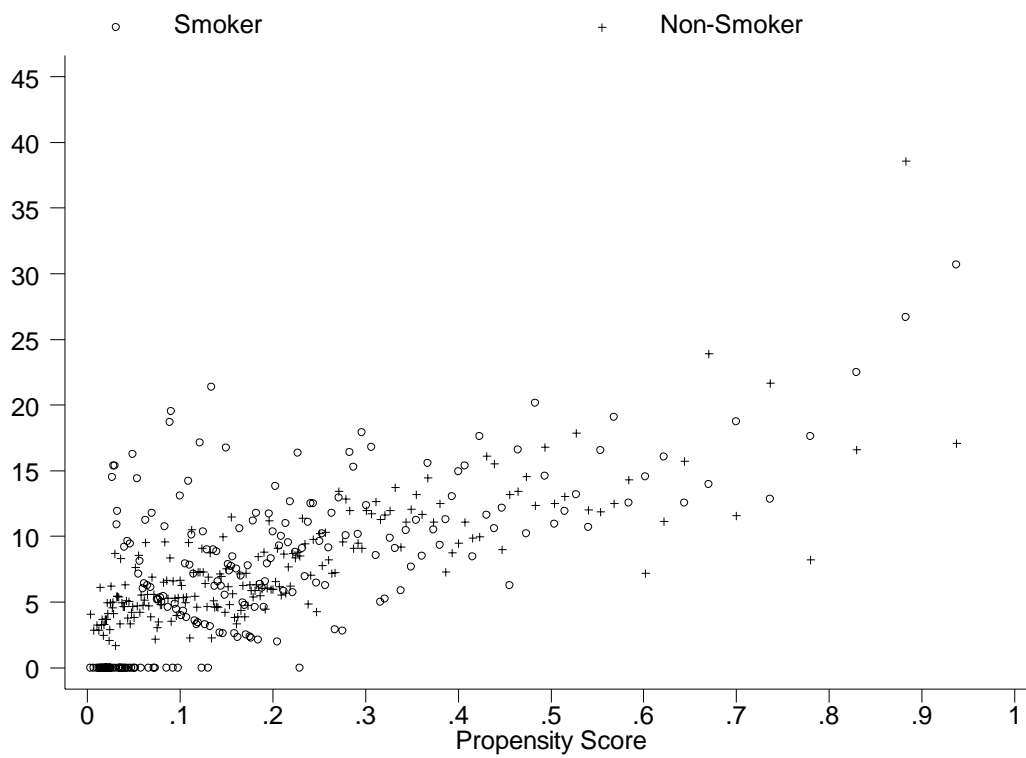
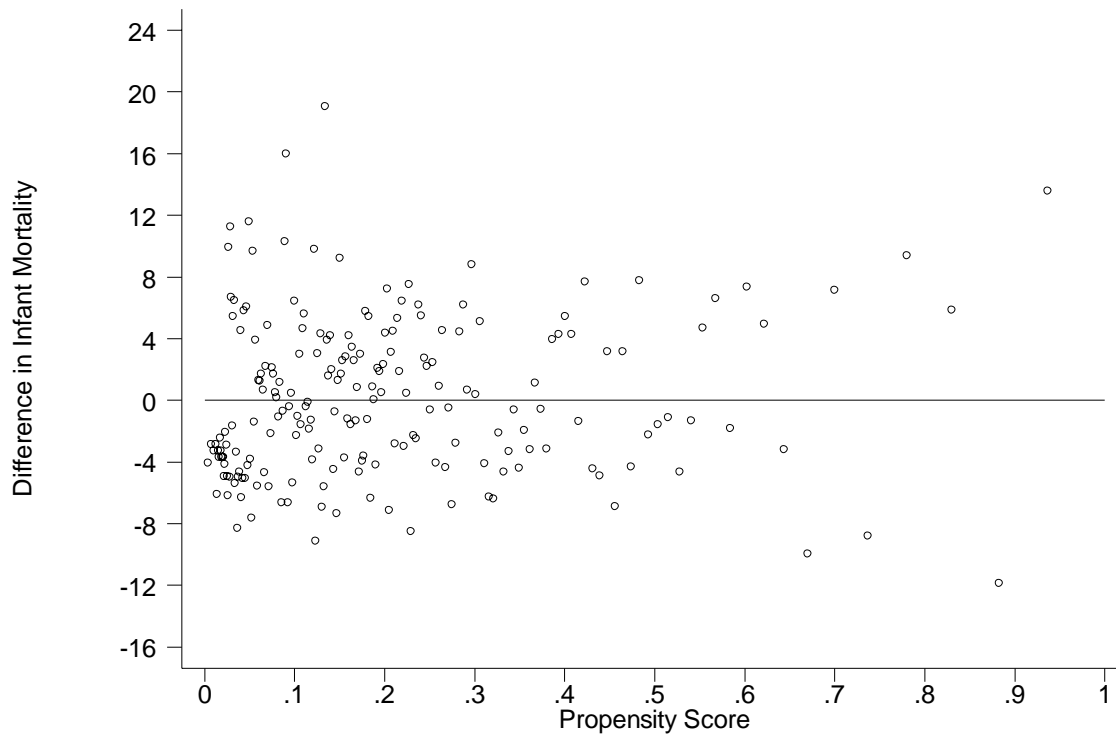


Figure 4: Estimated Effects of Maternal Smoking on Health Outcomes by the Propensity Score, Pennsylvania Singletons 1989-1991

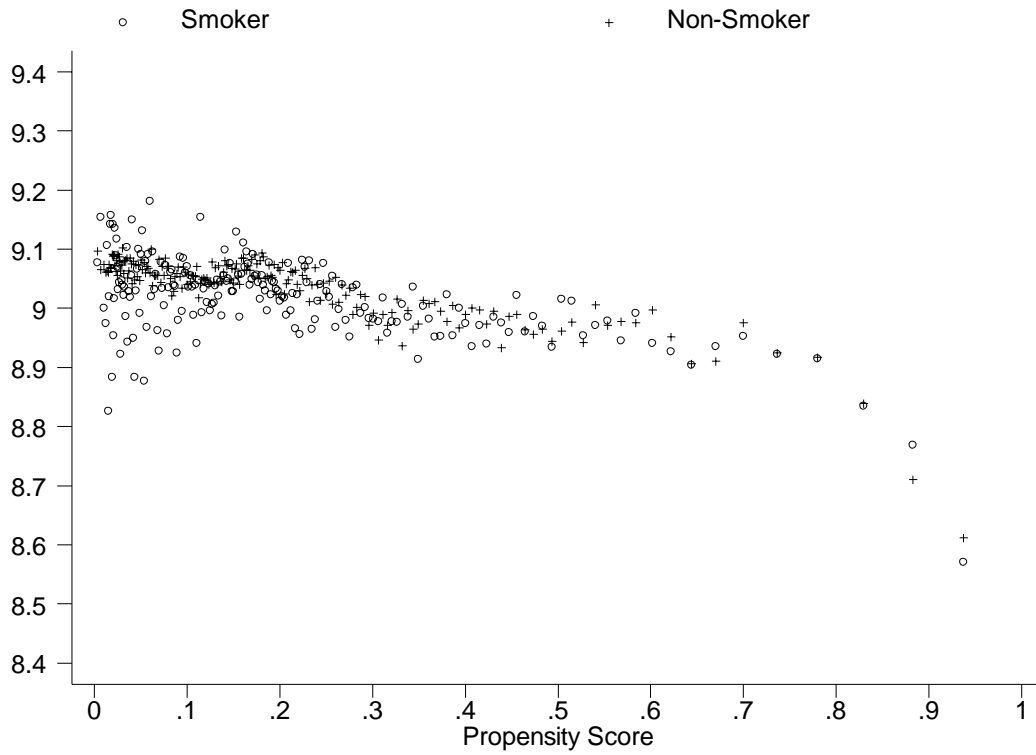
A. Infant Deaths within 1-Year of Birth (per 1,000 live births) for Smokers and Non-Smokers



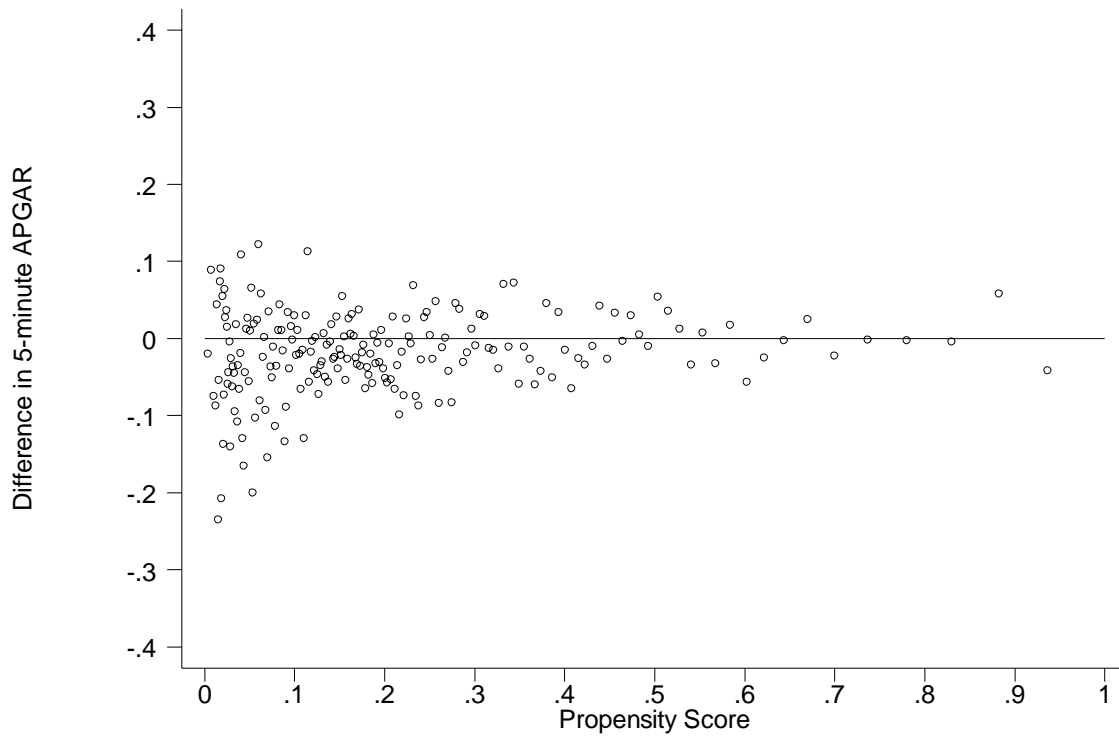
B. Difference between Smokers and Non-Smokers in Infant Mortality Rates



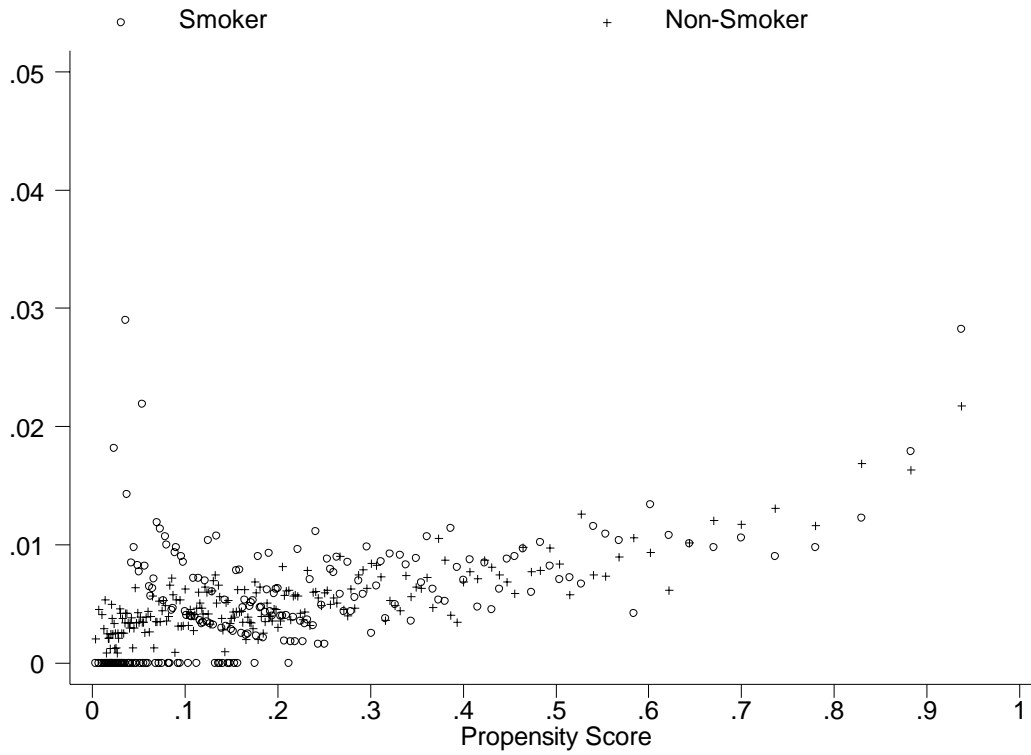
C. 5-Minute APGAR Score for Smokers and Non-Smokers by the Propensity Score



D. Difference between Smokers and Non-Smokers in the 5-Minute APGAR Score



E. Incidence of Assisted Ventilation ≥ 30 Minutes for Smokers and Non-Smokers



F. Difference between Smokers and Non-Smokers in Assisted Ventilation ≥ 30 Minutes

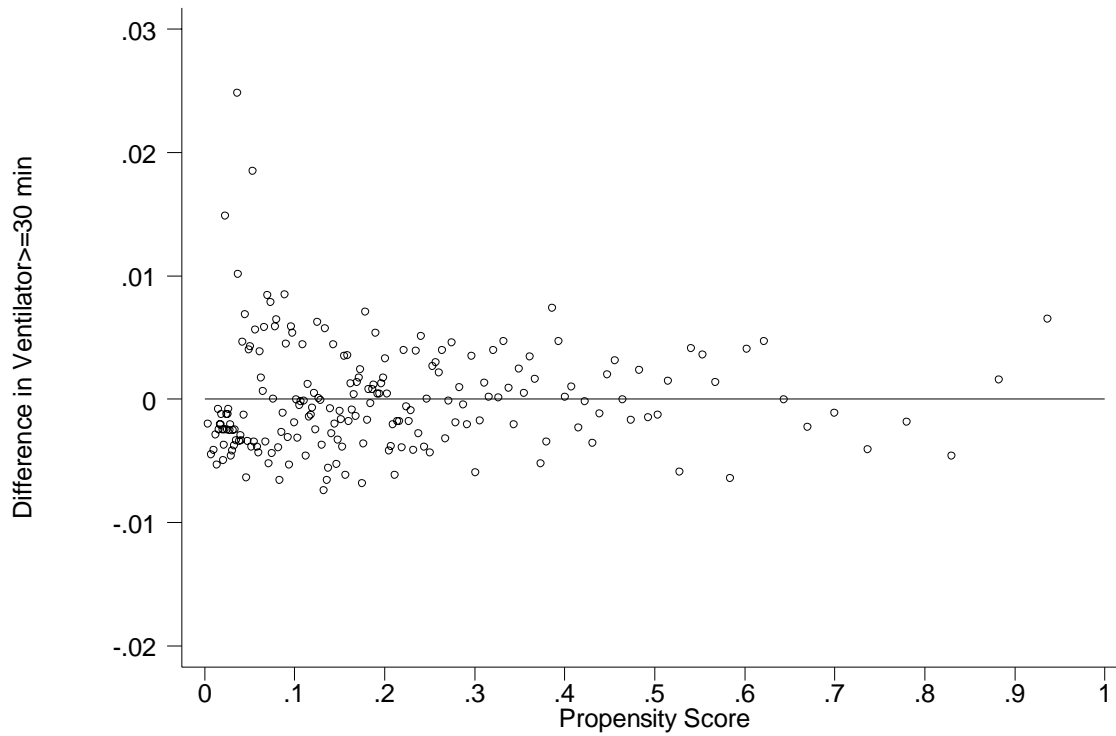
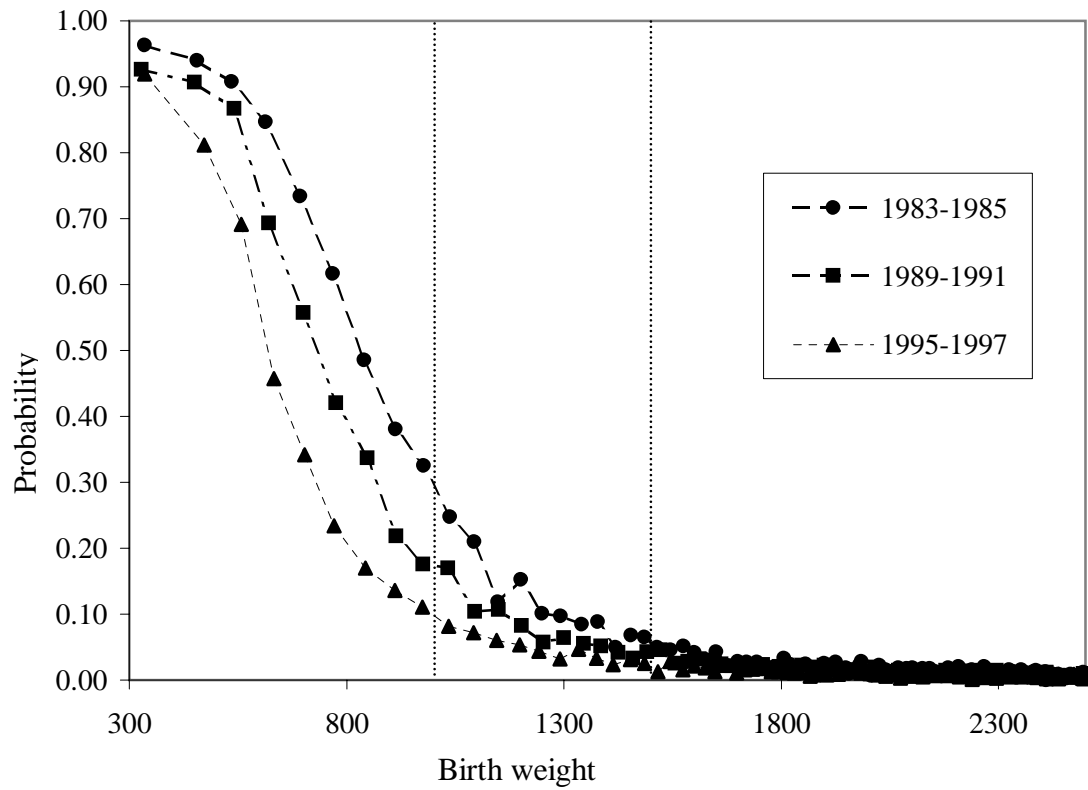
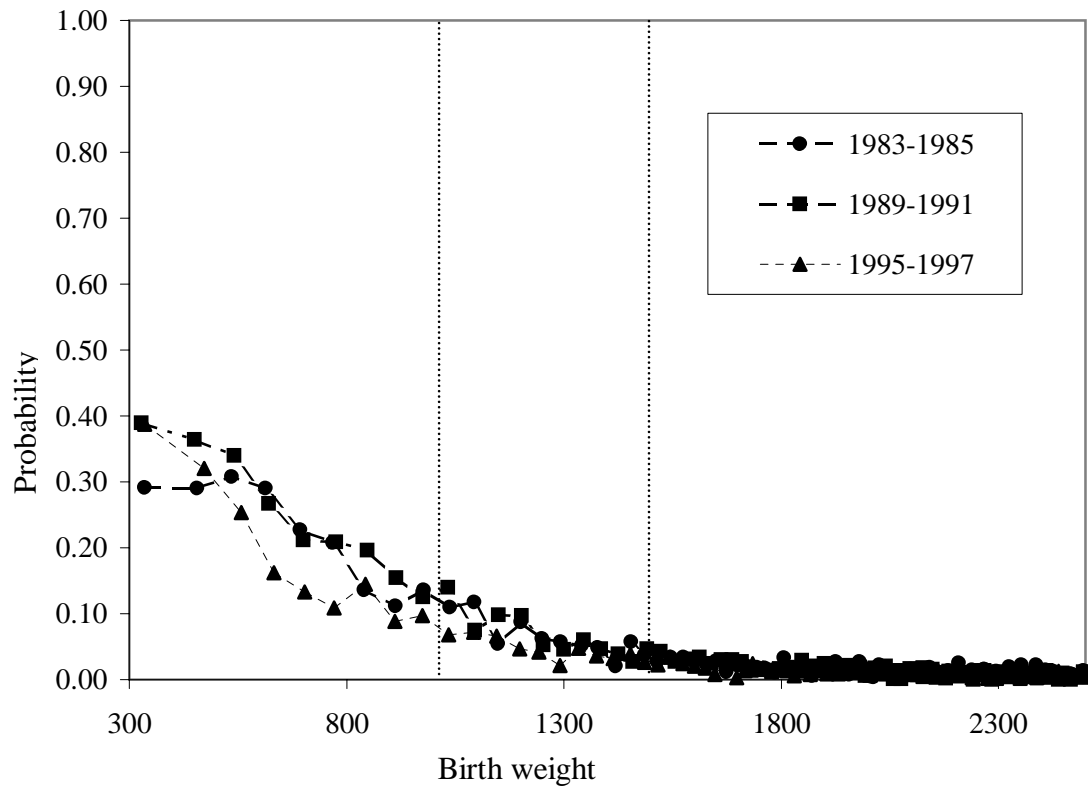


Figure 5A: Infant Mortality and Birth Weight, Pooled Twins, 1983-1997



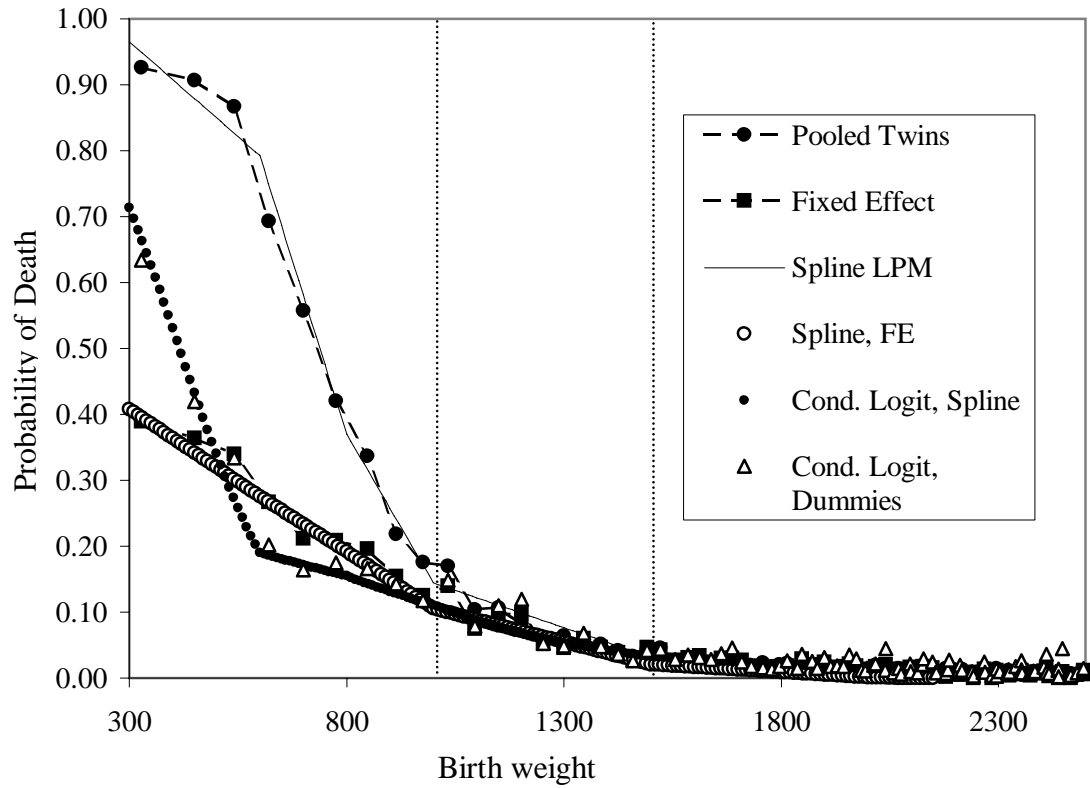
Note: Linked Birth-Death certificate data, 1983-1997

Figure 5B: Infant Mortality and Birth Weight, Fixed Effects, 1983-1997



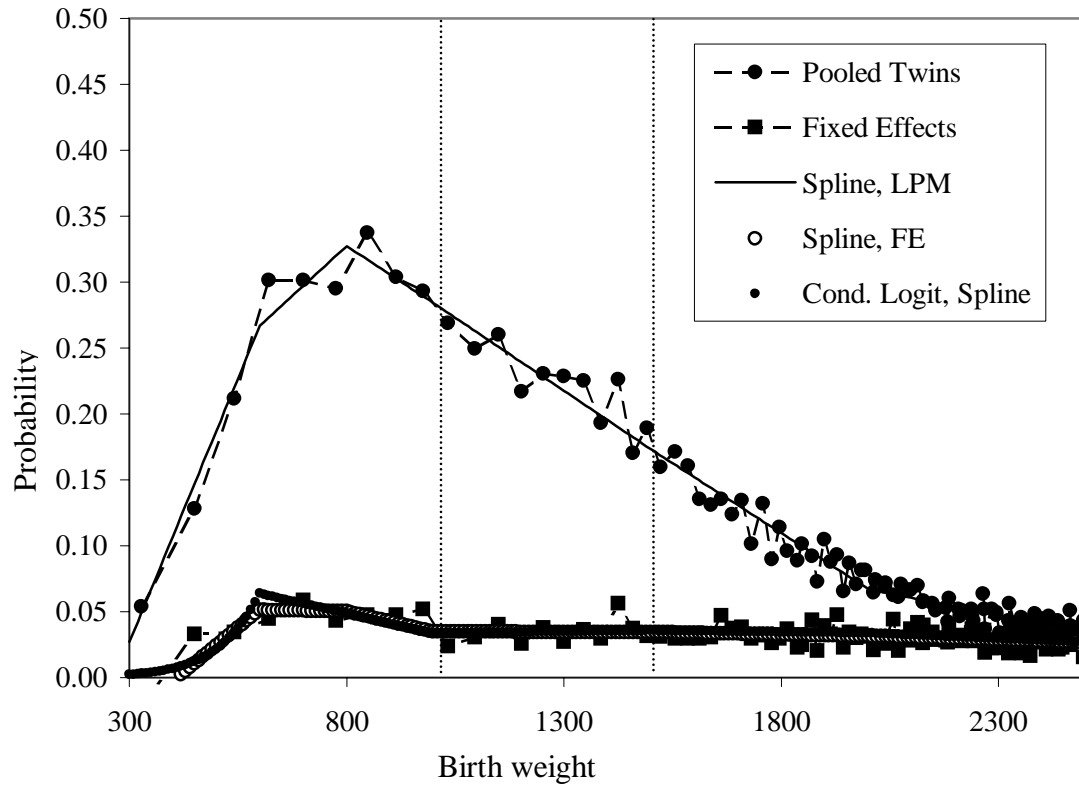
Note: Linked Birth-Death certificate data, 1983-1997

Figure 6A: 1-year Infant Mortality and Birth Weight



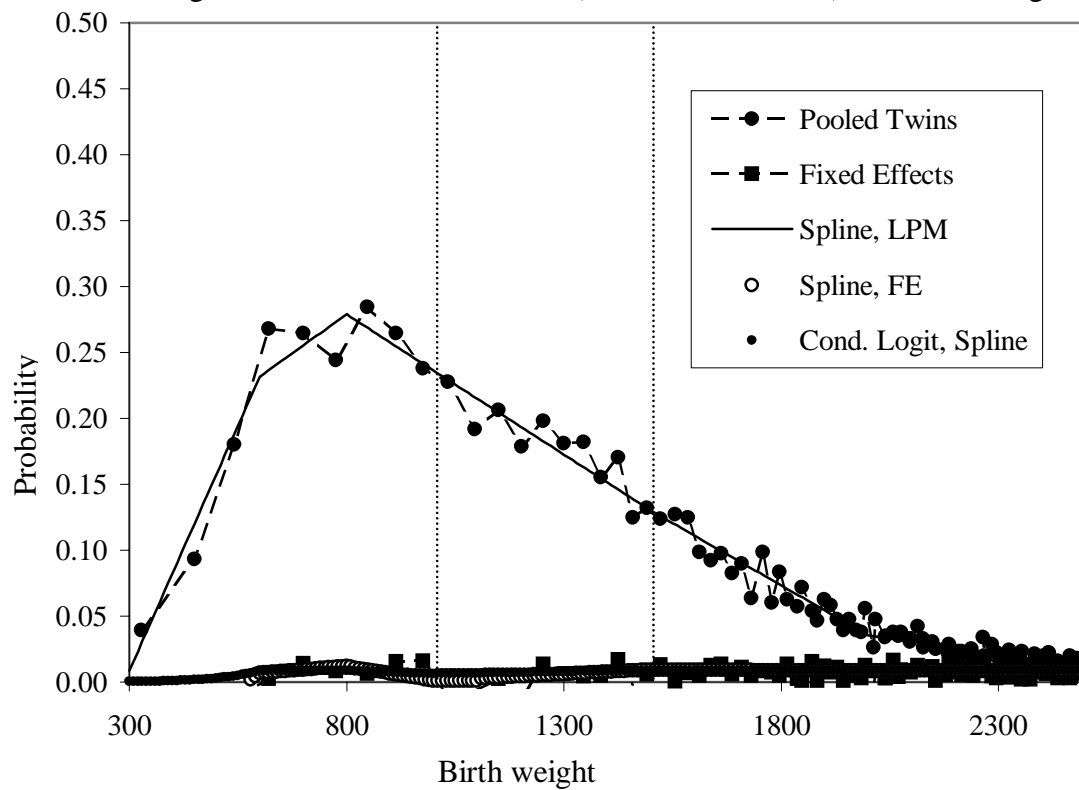
Note: Linked Birth-Death certificate data, 1989-1991

Figure 6B: Assisted Ventilation and Birth Weight



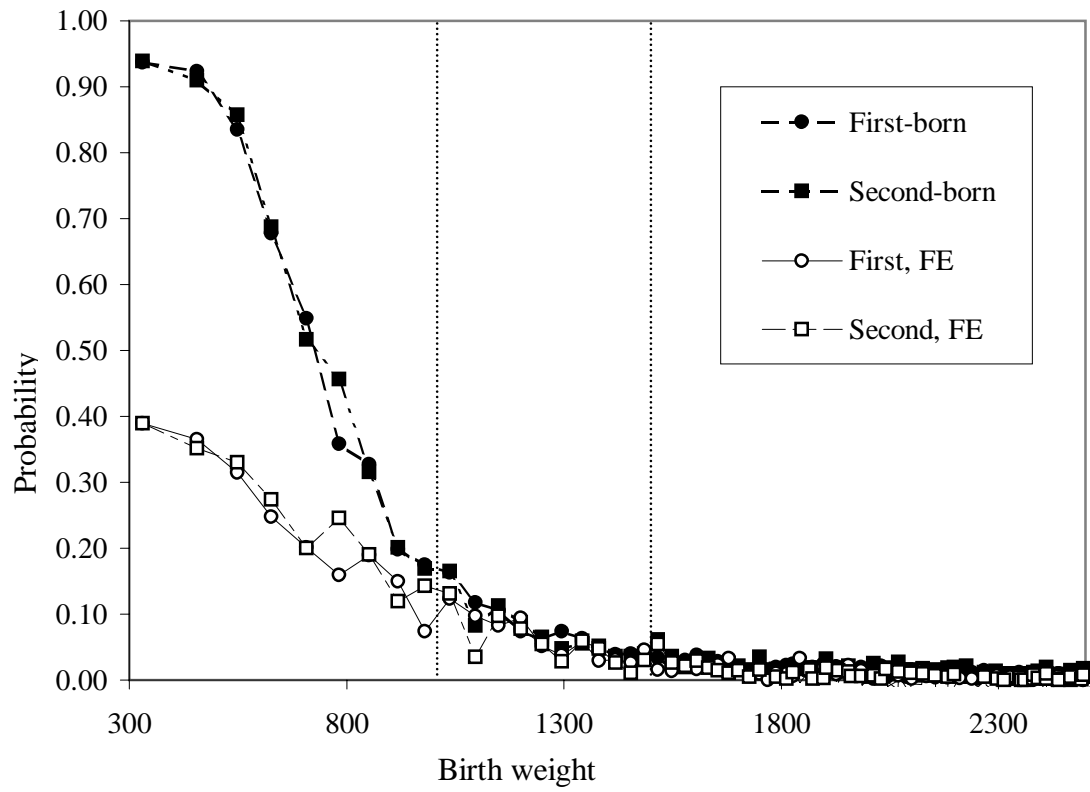
Note: Birth certificate data, 1989-1991

Figure 6C: Assisted Ventilation (30 minutes or more) and Birth Weight



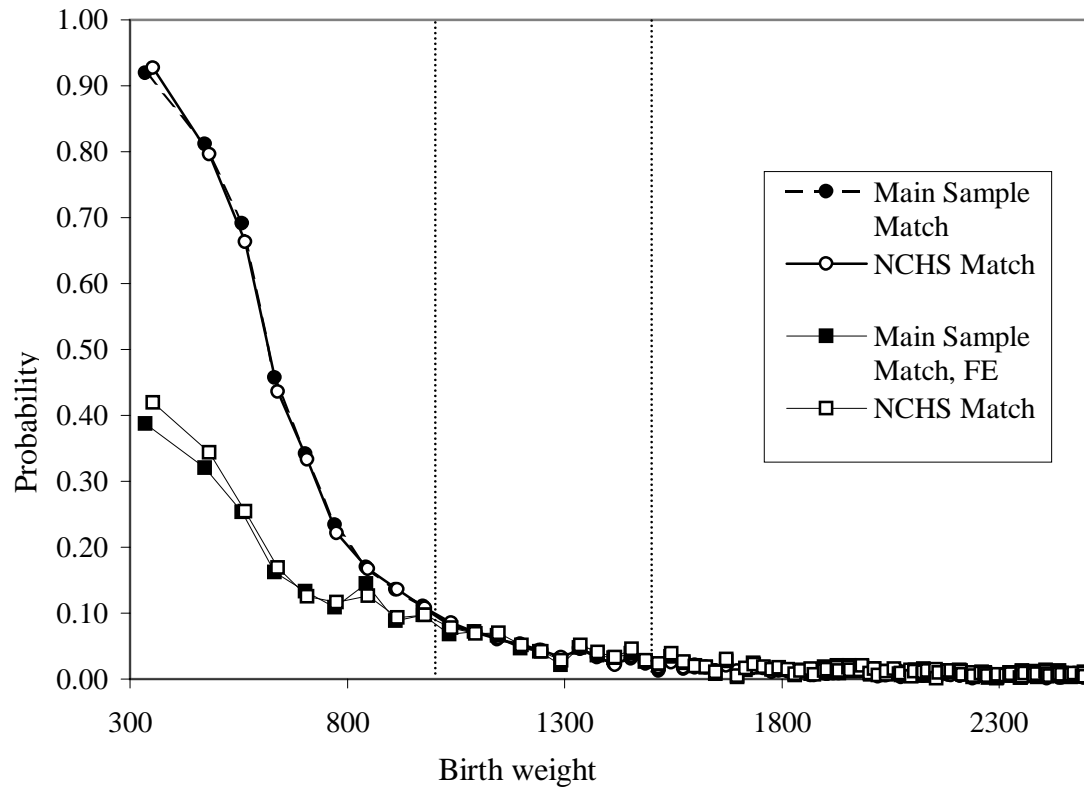
Note: Birth certificate data, 1989-1991

Figure 7: Infant Mortality and Birth Weight, First- and Second-born, 1989-1991



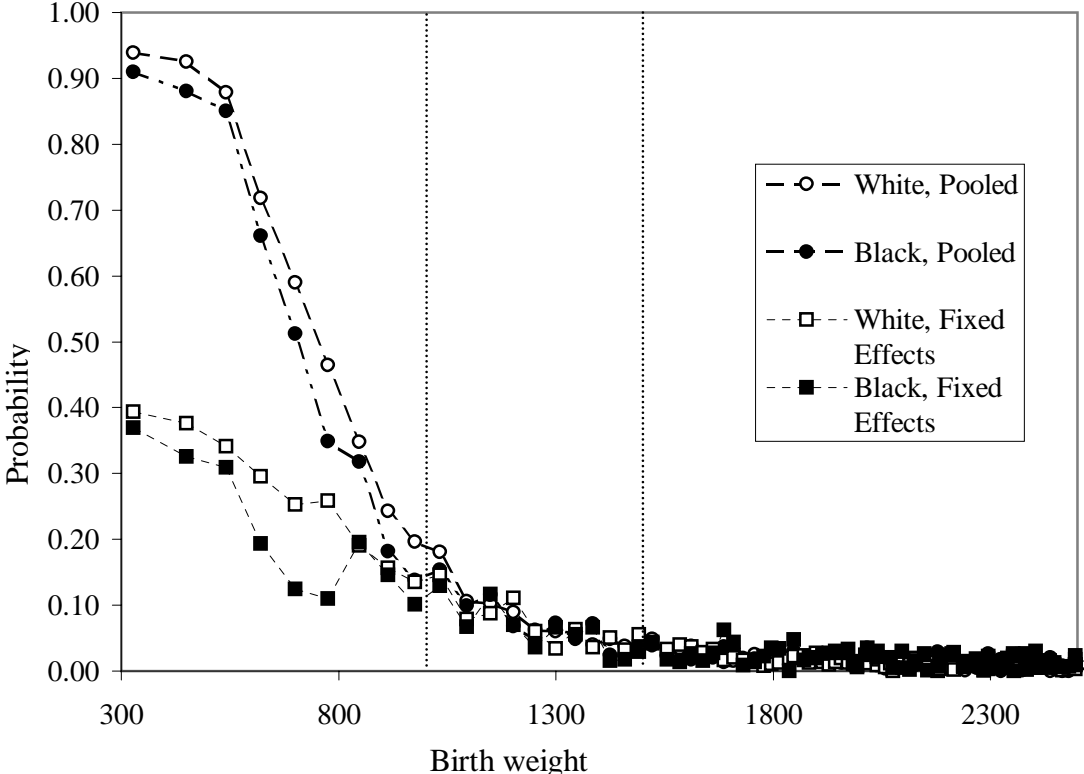
Note: Linked Birth-Death certificate data, 1989-1991

Figure 8: Infant Mortality and Birth Weight, Comparison of Matching Algorithms, 1995-1997



Note: Linked Birth-Death certificate and NCHS Match data, 1995-1997

Figure 9: Infant Mortality and Birth Weight, Pooled Twins, 1989-1991, By Race



Note: Linked Birth-Death certificate data, 1989-1991