THE COSTS OF LOW BIRTH WEIGHT*

DOUGLAS ALMOND
KENNETH Y. CHAY
DAVID S. LEE

Low birth weight (LBW) infants experience severe health and developmental difficulties that can impose large costs on society. However, estimates of the return to LBW-prevention from cross-sectional associations may be biased by omitted variables, 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 per se 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 twenty; 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 be suboptimal.

I. INTRODUCTION

Infants born at low birth weight (LBW)—conventionally defined as a birth weight less than 2500 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.[1] Studies have also established a correlation between LBW and high blood pressure, cerebral palsy, deafness, blindness, asthma,

* An earlier version of this paper 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 are grateful to three anonymous referees and Lawrence Katz for very helpful comments. We also thank David Card, Janet Currie, David Cutler, William Evans, Victor Fuchs, Lorenz Goethe, Michael Grossman, Jonathan Gruber, Theodore Joyce, Robert Kaestner, Karen Norberg, Jack Porter, Douglas Staiger, Paul Torelli, and many seminar participants for useful comments.

1. These figures are based on authors' calculations, illustrated in Figure I, which we discuss below.

The Quarterly Journal of Economics, August 2005

and lung disease among children, as well as with IQ, test scores, behavioral problems 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, Joyce, and Grossman 1987; Grossman and Joyce 1990; Geronimus and Korenman 1992; Rosenzweig and Wolpin 1991, 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 2004]. Finally, birth weight has been used to evaluate the effectiveness of social policy. Research on the benefits of large-scale social programs—including welfare 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 United States, 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

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].

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 [1999] (PA-99-045) is the "development of innovative strategies to prevent LBW in minority populations." In addition, the director of the NICHD, has

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. Maternal smoking has been identified as the most significant, modifiable risk factor for LBW incidence in developed countries [Kramer 1987].[5] The well-documented 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 has been used to calculate the sizable cost savings of interventions that encourage smoking cessation during pregnancy [Lightwood, Phibbs, and Glantz 1999].

There is, however, an important caveat to these kinds of policy conclusions.[6] 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 increase birth weights, but doing so would have no impact on mortality risk, if such risk was 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 confound-

---

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" [Alexander 2000].

5. Consistent with the literature, our calculations show that babies born to mothers who smoke during pregnancy are on average 9–10 ounces lighter, and over twice as likely to be of LBW than those born to nonsmoking mothers.

6. 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.

ing factors.[7] 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 charges, 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) 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 arguably overstated. We recognize that this approach cannot fully control for unobserved heterogeneity across mothers. Nevertheless, the analysis allows us to gauge the plausibility of the magnitudes derived from our analysis of twins, within an important and policy-relevant context.

Both approaches yield implied impacts of LBW that are many times smaller than those used in conventional cost-benefit calculations. The cross-sectional OLS estimates imply that a one standard deviation increase in birth weight is associated with a decrease in hospital costs, a reduction in infant mortality, an increase in APGAR scores (an initial assessment of the baby's health at birth), and a reduction in assisted ventilator use after birth, of 0.51, 0.41, 0.51, and 0.25 standard deviations, respectively. By contrast, controlling for mother-heterogeneity in the twins analysis, the corresponding magnitudes are 0.08, 0.03, 0.03, and 0.01.

Our analysis of the impact of maternal smoking during pregnancy on singleton health also points to a relatively weak LBW-

7. More specifically, as we discuss below, we assess the costs and consequences of intrauterine growth retardation.

health relationship. 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. But 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. Further, while the cross-sectional estimates typically used in the literature imply a cost savings of over $900 per infant for each mother who ceases smoking during pregnancy, our estimates imply a cost savings of only $53 per infant.

There are two plausible interpretations of these findings. 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 implications of our results for policy and the broader economics literature in Section VI and conclude in Section VII.

II. Conceptual Framework and Identification Strategy

This section discusses our framework for quantifying the costs and consequences of LBW. We define the parameter of interest and describe our identification strategies, discuss the potential determinants 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.

II.A. Parameter of Interest

Let

$$h_{ij} = \alpha + bw_{ij}\beta + X_i'\gamma + a_i + \varepsilon_{ij}, \tag{1}$$

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 bivariate regression of the health measure on birth weight (with no other covariates included) is

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

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 United States (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.

II.B. Identification Strategies

We use two different strategies to estimate β. First, we examine the population of twin births in the United States, 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:

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

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} might overstate the impact of birth weight per se.[8]

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

$$bw_{ij} = \lambda + \boldsymbol{SMOKE}_i\delta + X_i'\psi + u_{ij}, \tag{4}$$

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

$$h_{ij} = \alpha^* + \boldsymbol{SMOKE}_i\delta\beta + X_i'\gamma^* + a_i^* + u_{ij}\beta + \varepsilon_{ij}, \tag{5}$$

where a_i^* is the residual from regressing a_i on X_i.[9] Unbiased estimation of $\delta\beta$ with OLS requires $\boldsymbol{SMOKE}_i$ to be uncorrelated with a_i^*.

8. 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.

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

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.[10] 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 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.

II.C. 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 given gestation length, also known as intrauterine growth retardation (IUGR). Empirically, gestation length is an important determinant of birth weight. The first row of Table I, which reports mean squared errors from regressions

10. The quasi-experimental studies involve randomly assigned interventions in which the treatment group is encouraged to cease smoking through personal visits, information dissemination, telephone calls, etc. Some examples are Sexton and Hebel [1984], Ershoff, Dolan-Mullen, and Quinn [1989], and Windsor et al. [1993].

TABLE I
COMPONENTS OF VARIANCE FOR BIRTH WEIGHT AND OUTCOMES AMONG TWINS

	Mean squared error in OLS regressions				Ratio
Dependent variable	(1)	(2)	(3)	(4)	(4)/(3)
1989–1991 U. S. twins					
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 ≥30 min.	0.0370	0.0348	0.0338	0.0102	0.30
1995–2000 NY-NJ twins					
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	

The hospital cost data are from 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 Infant Birth-Death Detail Files. 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 100s of grams, and hospital costs are in 10,000s 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.

with birth weight as the dependent variable, shows that the inclusion of gestation variables (columns (2) and (3)) explains over half of the overall variance (44.43) in birth weight in a pooled sample of twins.[11] Despite the significant contribution of gestation length to variation in birth weight, the emphasis of the literature has instead been on IUGR. 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. This is perhaps because the causes of prematurity are less well-understood. For example, interventions targeted at preventing premature birth, including enhanced prenatal care and nutritional interventions, have been found to be ineffective [Goldenberg and Rouse 1998].

11. Since gestation length is measured at the level of weeks, 19.08 in the third column is arguably an upper bound on the residual variation in birth weight remaining after controlling for gestation length fixed effects.

By contrast, studies focusing on IUGR tend to point to two important factors. According to the Kramer [1987] survey, nutrition plays an important role in intrauterine growth. In both developed and developing countries, low caloric intake is one of the most important modifiable contributors to IUGR. In addition, Kramer finds that the other important modifiable contributor to IUGR, particularly in developed countries, is cigarette smoking during pregnancy.[12]

Our two empirical approaches attempt to use sources of variation in birth weight attributable to IUGR. The two sources of variation are quite distinct from one another, each having advantages and limitations. With smoking-induced variation in birth weight, we can directly assess whether the health impacts of a specific, modifiable, environmental factor match what would be predicted by a strong, structural link between LBW and health. The main limitation is that the covariates we use in the regression and propensity score analysis may not account for all of the between-mother differences in genetics and other environmental factors that influence IUGR.

By contrast, the twins analysis controls for all mother-specific factors that simultaneously impact IUGR and other health outcomes. In addition, intrapair variation in birth weight must necessarily be due to differences in IUGR, since twins have identical gestational ages. An important caveat, however, is that the exact identities and relative contributions of the environmental factors causing intrapair variation are unknown, as is their modifiability.

For example, if twin differences arise solely because one twin's nourishment is of higher caloric value than the other's, then the twins analysis could plausibly be used to estimate the relation between health outcomes and LBW induced by low caloric intake of the mother—a factor that could be directly influenced by a health intervention. But if, for example, the twin differences arise instead because one twin receives a greater supply of blood from the mother than the other, then the usefulness of the twins analysis for policy predictions would be considerably diminished.

While the medical literature has identified differences in nutritional intake as the predominant cause of birth weight dif-

12. Other identified contributors in Kramer [1987] were more generalized: prepregnancy weight, very young maternal age, maternal education, and weight gain during pregnancy.

ferences between twins, there exist other causes, some of which are not amenable to policy intervention.[13] Further, even if caloric intake was the sole cause of birth weight differences between twins, the results from the twins analysis may not be valid for predicting the effects of increased caloric intake on the health outcomes of singleton births. We return to this issue below.

Whatever the precise mechanism, within-twin-pair differences in birth weight are large and account for a substantial fraction of the overall birth weight variation due to IUGR in a pooled sample of twins. Column (4) of the first row in Table I shows that the mean squared error from a regression of birth weight on mother fixed effects is 7.54. This implies that 40 percent of the birth weight variation due to differential fetal growth rates is due to within-mother differences.[14] Thus, the twins analysis can be viewed as measuring the relation between health outcomes and 40 percent of the total variation in IUGR. The average difference in birth weight between heavier and lighter twins is 290 grams, with the 10th, 25th, 50th, 75th, and 90th percentiles equal to 30, 85, 220, 400, and 620 grams, respectively. To put these differences in perspective, note that maternal smoking during pregnancy—considered the most important modifiable cause of LBW in the United States—is believed to reduce birth weight by about 200 grams, on average.

13. While twins share the same supply of nourishment from the mother, this sharing is often unequal. For example, for identical twins sharing the same placenta, differing insertion points of the twins' umbilical cords into the placenta are believed to affect nutritional intake of each fetus, and hence contribute to intrapair differences in fetal growth [Bryan 1992]. More generally, the most commonly cited factors for twin birth weight differences are different implantations of placentae in the uterus leading to different nutritional intakes and blood perfusion [Zhang, Brenner, and Klebanoff 2001], different nutritional sources at different umbilical cord insertion points within the placenta, 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. Partially due to such haemodynamic effects, monozygotic (MZ) twins that share the same placenta exhibit larger birth weight differences, on average, than MZ twins with dichorionic placentae [Derom, Derom, and Vlietinck 1995]. Some of this comes from a conversation with Dr. Louis Keith in the Obstetrics and Gynecology Department of Northwestern University, who has edited a text on multiple pregnancies [Keith, Papiernik-Berkhauser, and Keith 1995].

14. Again, due to imperfect measurement of gestational age, the column (3) entries may be overstated, which implies that the ratios in the final column may be understated.

II.D. Measures of Health and Costs and Within-Twin Pair Variation

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 one year after birth. The advantage of this measure is that it is a somewhat objective indicator 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.

Our second measure is 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 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 gives a score from 0 to 2 for each factor. The five health factors are heart rate, respiratory effort, muscle tone, reflex irritability, and color, and the five scores are summed to calculate the APGAR score.

Since infant death can be extremely rare at birth weights above 1500 grams, this measure may provide additional information on underlying infant health at birth. Although it is highly correlated with infant mortality, there is substantial variability in APGAR scores among infants who survive their first year of life, and this variability is highly correlated with different measures of health at birth.[15] Also, using data from the National Maternal and Infant Health Survey (NMIHS), we find that the

15. In our sample, controlling for infant death in the first year removes 1) under 30 percent of the overall variation in the five-minute APGAR score, 2) 7 percent of the variation in APGAR scores at fixed gestation lengths, 2 percent of the variation in scores within families. Further, among infants who survive their first year of life, APGAR scores are highly correlated with 1) gestation lengths across families in which both twins survive, and 2) assisted ventilator use after birth both across families and within twin pairs in which both survive.

five-minute APGAR score is a significant predictor for measures of health, cognitive ability, and behavioral problems of the child at age three, after controlling for family background variables and infant birth weight.[16] We conclude that APGAR scores contain information on infant health that is independent of infant mortality.

Our third outcome measure is the newborn's utilization of a ventilator for assisted-breathing shortly after delivery. This indicator incorporates health providers' subjective assessment of the infant's immediate medical needs and hence—indirectly—their underlying health. When it is utilized, it can also comprise a large share of the overall costs of delivery. It is important to note that assisted ventilation is administered not simply on the basis of whether a newborn is premature. If this were the case, then there would exist little correlation between its use and IUGR in the first place. On the contrary, even after controlling for gestation length, birth weight is a strong predictor of ventilator use.[17] This is consistent with what we learned from consultation with Dr. David Rubenstein, director of the Neonatal Intensive Care Unit at Columbia-Presbyterian hospital—that ventilators were originally developed for *term* babies, and then subsequently expanded to treat preterm babies as well.[18]

The second through sixth rows of Table I show that there is substantial *within-twin-pair* variation in these three measures of health. Intrapair differences in mortality account for 45 to 68 percent of the variation in infant death conditional on gestation

16. The NMIHS is one of the only data sets that contain information on APGAR score and health outcomes subsequent to the first year of life. In the analysis we adjust for mother's age (cubic), race, schooling (cubic), household income, child's sex, state of birth, and birth weight. We find that the APGAR score has a negative, statistically significant association with the likelihood of having various chronic childhood diseases (e.g., asthma and hearing problems) at age three. Further, the APGAR score is positively and significantly associated with achieving various developmental milestones (results available from the authors).

17. With our twins sample, a regression of ventilator use on birth weight among pairs where both twins survived yields a coefficient of -0.084 (per 1000 births, per gram) with a standard error of 0.0018. When gestation length dummies are included the coefficient is -0.426 with a standard error of 0.00314. Furthermore, among the more than three million singleton births in 1989 with information on ventilator use, IUGR infants (those with birth weights less than the tenth percentile for a given gestational age) are twice as likely to be given assisted ventilation and three times as likely to require more than 30 minutes of assisted ventilation than non-IUGR infants.

18. The primary cause of respiratory distress among premature infants is surfactant deficiency [Behrman, Kliegman, and Jenson 2000]. There was a sizable increase in the use of ventilators among preterm infants after the introduction of surfactant therapy in late 1989, which led to greatly improved survival rates among infants born severely premature.

length (last column). The within twin-pair variations in APGAR scores and ventilator use are similar in magnitude. These findings suggest that the factors that cause differences between twins of the same mother may also be important in explaining across-family differences for infants of the same gestational age. We also find that the APGAR score is a substantive predictor of the within-twin-pair variation in infant death and ventilator use. Thus, the relatively small birth weight effects that we find do not appear to be the result of either measurement error in the outcomes or behavioral responses of parents and hospitals (e.g., keeping the twins together).

Our fourth and final outcome measure is a proxy for hospital costs. Specifically, we use data on the total charges incurred by the hospital for the delivery and initial care of the infants, derived from hospital inpatient discharge abstracts. Strictly speaking, the data provide information on hospital charges which are claimed for billing purposes. Thus, while recognizing that they are conceptually different, we use the terms "costs" and "charges" interchangeably throughout the paper. When using hospital charge data in an analysis of twins, it is important to consider three potential problems. First, to the extent that hospitals "lump together" and "split" charges for twins, even if one twin receives different medical procedures, there will be little within-pair variation in charges to explain, thus biasing any result toward zero. As Table I shows, however, the contribution of intrapair variation to the overall cross-sectional variation in hospital charges is about (2.958/14.410) 21 percent; this is comparable in magnitude to the analogous figures for one-day mortality, five-minute APGAR score, ventilator use, and birth weight.[19]

Second, there could be sample selection based on whether the charges were assigned to the mother—whom we generally cannot link to the infants. This is problematic insofar as this potential sample selection is correlated with birth weights. We did find, however, that mean birth weights for the twins in the discharge data are very similar to those in our vital statistics data, which does not suffer from this selection bias. We also find that the pooled cross-sectional and fixed-effects estimates of the relation-

19. Since the discharge data do not contain gestation length information, we cannot compare the intrapair variation in hospital charges to the variation at given gestation lengths.

ship between birth weight and early-life mortality are similar in the two data sets.

A third concern is that infants within the same twin-pair could be given treatments that are charged to different payers, who may have different payment schedules.[20] For example, if one of the twin's charges are billed to the Medicaid program while those of the other are not, and this is systematically correlated with birth weight, this could cause a bias in the fixed-effects estimates. To assess the importance of this problem, we regressed 1) an indicator of whether the expected payer is the Medicaid program, and 2) whether the charge will be "self-paid" on birth weight, with mother fixed effects. In neither of these regressions are the charges significantly related to birth weight.

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

II.E. Functional Form Issues

There are a number of important functional form issues involved in estimating equations (3), (4), and (5). First, the relationships between birth weight and various health and cost outcomes may be nonlinear. Thus, we additionally estimate fixed-effects models that include a set of dummy variables for discrete birth weight categories, given by

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

where D_{ij}^k is an indicator variable that is equal to one if the jth infant of the ith mother is in the kth (out of 200) quantile of the overall birth weight distribution.[21] 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.

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

20. Approximately 4 percent of the matched twin pairs have different expected payers.

21. For the pooled singletons' plots, we used 655 quantiles of the conditional distribution of birth weights below 3000 grams.

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]. Specifically, we estimate a logit model that predicts the probability of maternal smoking during pregnancy based on all of the "pretreatment" 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 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. 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

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.

III.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 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 main analysis of twins takes advantage of a useful aspect of the infant birth/death data: beginning with the 1989 file, records for twin pairs are located next to each other. This fact, combined with detailed parental and pregnancy characteristics, allowed us to easily construct twin-pairs. The precise algorithm is described in the Data Appendix. An alternative approach to matching twins (also described in the Data Appendix) is to match infants with identical parental covariates among the subsample of records identified to be a multiple-birth event. It was necessary to use this procedure for the earlier, 1983–1985 data, since twin records were not adjacent in those data. Subsequent to our initial analysis in Almond, Chay, and Lee [2002], the NCHS released the 1995–1997 Matched Multiple Birth Data Set, which uses a "covariate-matching" algorithm that is similar to the procedure that

we used for the 1983–1985 data.[22] We have investigated the sensitivity of our results to our initial (record-adjacency) algorithm, our own covariate-matching algorithm, and the algorithm used by the NCHS: all algorithms yield nearly identical results. Finally, it is important to note that any false matching of infants to twin-pairs will tend to bias the within-twin estimate toward the pooled cross-sectional estimate; thus, any difference between the fixed-effects and cross-sectional estimates will, if anything, be an understatement of their true difference.

III.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, eighteen 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 comprise almost one-quarter of all available discharge records in 2000. We also examined the discharge data from Maryland, which provided nearly identical results (available

22. These data are used in Conley, Strully, and Bennett [2003a, 2003b].

from the authors).[23] We use a covariate-matching approach to construct a data set of 44,410 (non-Hispanic White and Black) newborn twins in New Jersey and New York between 1995 and 2000. The details of the matching algorithm are described in the Data Appendix.

III.C. Twins Compared with Singletons

Although twins represent a small subpopulation of all infants born in the United States, they are of great interest in the health and biomedical literatures for several reasons. First, while constituting 3 percent of all births in the United States 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].[24] Further, Kogan et al. find that multiple births have accounted for an increasing share of all LBW infants over the past twenty 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.[25] 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 II 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 populations. Mothers of singleton and twin infants have comparable demographic characteristics, along with 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,

23. As more than half of the Arizona discharge records are missing information on birth weight, Arizona is excluded from the analysis, along with Colorado, which has the smallest sample size (15 percent of the 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.

24. 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 United States.

25. For example, Behrman and Rosenzweig [2001] calculate that twinning results in a 12 percent reduction in lifetime earnings when compared with 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.

TABLE II
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	
Mother's demographic characteristics			
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
Mother's risk factors			
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
Characteristics of birth			
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
Infant deaths (per 1000 live births)			
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
Sample size	2,655,977	58,132	
Infant birth weight (grams)			
Mean	3,369	2,417	339.6
	(591)	(670)	
Median	3,402	2,495	
25th percentile	3,060	2,070	
10th percentile	2,693	1,503	
5th percentile	2,410	1,080	
1st percentile	1,430	501	
Fraction low birth weight (<2500 g)	0.061	0.504	

TABLE II
CONTINUED

	Sample means		*T*-ratio for diff in sample means
	Singletons	Twins	
Gestation length (in weeks)			
Mean	39.3	36.0	201.2
	(2.65)	(3.85)	
Median	40	37	
25th percentile	38	34	
10th percentile	37	31	
5th percentile	35	28	
1st percentile	29	23	
1-minute APGAR score (0–10)			
Mean	8.02	7.30	79.9
	(1.35)	(1.97)	
Median	8	8	
25th percentile	8	7	
10th percentile	7	4	
5th percentile	5	3	
1st percentile	2	1	
5-minute APGAR score (0–10)			
Mean	8.97	8.52	70.2
	(0.81)	(1.40)	
Median	9	9	
25th percentile	9	8	
10th percentile	8	7	
5th percentile	8	6	
1st percentile	6	1	
Fraction of dead with birth weight <2500 g			
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	

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.

more educated, and married. In addition, they are more likely to be anemic and have pregnancy-associated hypertension.

There are several, important differences between the two infant populations. First, twins are much more likely to die within one day, one week, one month, and one year of birth than singletons. In addition, twins tend to be lighter, with mean birth weights about 950 grams lower than singleton newborns. As Table II 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. These two facts have been previously cited as evidence that birth weight differences may indeed be a cause of the twin-singleton gap in infant mortality.

However, the table also shows substantive differences in the gestation lengths of twin and singleton births. The median twin is born prematurely, and the difference in the distributions of gestation length is similar to that for birth weight. There are also large differences in APGAR scores, the incidence of a breech birth, and the incidences of abnormal conditions and congenital anomalies at birth. Thus, along many measures, twins tend to be less healthy at birth than singletons.

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 (as shown in Figure I). Further, the cross-sectional relation between infant mortality and gestation length differs for twins and singletons.

Figure V in Appendix 2 plots the average birth weight and infant mortality rate for the twin and singleton populations in 1989, by infant gestation length. At gestation lengths below 27 weeks, there is a statistically significant, twin-singleton gap in infant mortality. This gap is similar for those born at 20–23 weeks compared with those born at 24–26 weeks, even though the birth weight difference grows from less than 100 grams to about 250 grams. In addition, the twin-singleton difference in infant mortality is insignificant at 28 to 38 weeks of gestation (and changes sign for 31–37 weeks), even as the birth weight difference grows to well over 500 grams (over one pound). Taken literally, these results suggest that either: 1) birth weight per se is not the cause of the twin-singleton infant mortality gap, or 2) the effects of birth weight on infant mortality only exist for infants born at less than 27 weeks of gestation or born with a birth weight below 1000 grams.[26]

In view of these differences, we do two things when reporting our results. First, we compare the fixed-effects estimates to the

26. Two-thirds of the estimated association between the twin-singleton difference in birth weight and the infant mortality difference can be explained by twin-singleton differences in gestation length, measured in weeks (results available from the authors). Loos et al. [2005] conclude that the factors underlying twin-singleton differences in intrauterine growth include peripheral umbilical cord insertion, fusion of placentas, chorionicity, and limited uterine size of the mother.

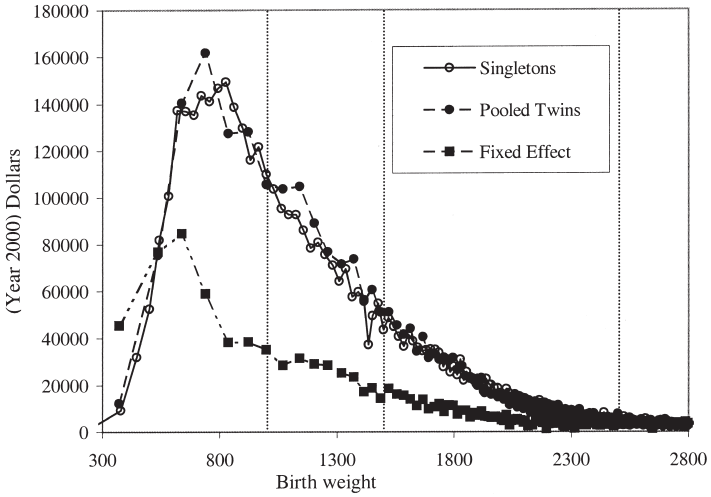

FIGURE Ia
Hospital Costs and Birth Weight
Note: 1995–2000 NY/NJ Hospital Discharge Microdata.

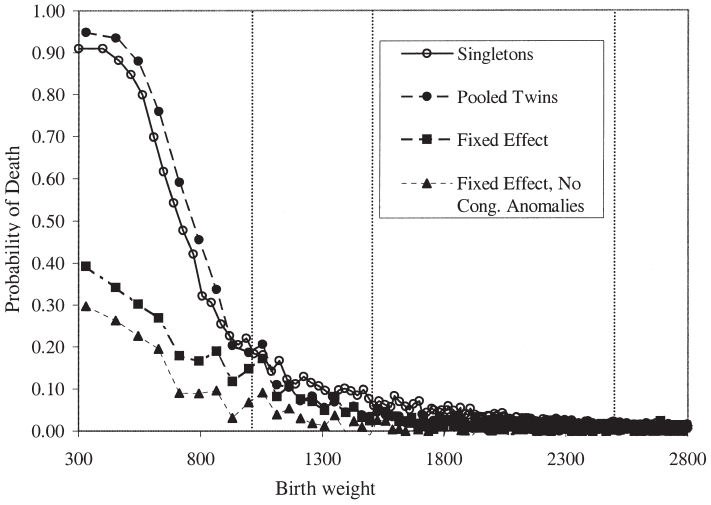

FIGURE Ib
Infant Mortality (1-year) and Birth Weight
Note: Linked Birth-Death certificate data, 1989.

pooled cross-sectional 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-sectional estimates for the twins with that for the singleton population. The extent to which they are similar can suggest that twins are subject to the same LBW-outcome relation (equation (1)) and to the same kinds of omitted variable biases (equation (2)) as singletons.

Finally, in our analysis we present results that both include and exclude pairs where at least one twin has a congenital anomaly, which is a defect 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.[27] Thus, in an effort to make our analysis more relevant for evaluating policies that focus on influencing birth weight per se, we present both sets of results. As we show below, the exclusion of pairs of infants with congenital anomalies reduces the magnitudes of the birth weight coefficients further. For cost-benefit calculations, we use the coefficients for the entire sample to give a conservative, understated view of the importance of omitted variables.

IV. Empirical Results

IV.A. Main Twins Results

Table III reports the coefficients from linear regressions of several different outcomes on birth weight (in grams) for different samples of twin pairs. 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. The cross-sectional OLS estimates in the first column indicate that a one standard deviation increase in birth weight (667 grams or one pound and seven ounces) is associated with a decrease in hospital costs of about $19,500 (in $2000), or about 0.51 of a standard deviation. When mother fixed

27. For all congenital malformations, there is a higher concordance among monozygotic than dizygotic twins (e.g., five times greater for cleft palate), suggesting that malformation differences within twin pairs are partially due to genetic differences [Bryan 1992]. However, for many malformations, MZ twins are more likely than not to be discordant.

TABLE III
POOLED OLS AND TWINS FIXED EFFECTS ESTIMATES OF THE EFFECT OF BIRTH WEIGHT

Birth weight coefficient	Including congenital anomalies		Excluding congenital anomalies	
	Pooled OLS	Fixed effects	Pooled OLS	Fixed effects
Hospital costs	−29.95	−4.93	—	—
(in 2000 dollars)	(0.84)	(0.44)	—	—
	[−0.506]	[−0.083]	—	—
Adj. R^2	0.256	0.796	—	—
Sample size	44,410	44,410	—	—
Mortality, 1-year	−0.1168	−0.0222	−0.1069	−0.0082
(per 1000 births)	(0.0016)	(0.0016)	(0.0017)	(0.0012)
	[−0.412]	[−0.078]	[−0.377]	[−0.029]
Adj. R^2	0.169	0.585	0.164	0.629
Sample size	189,036	189,036	183,727	183,727
Mortality, 1-day	−0.0739	−0.0071	−0.0675	−0.0003
(per 1000 births)	(0.0015)	(0.0010)	(0.0015)	(0.0006)
	[−0.357]	[−0.034]	[−0.326]	[−0.001]
Adj. R^2	0.132	0.752	0.127	0.809
Sample size	189,036	189,036	183,727	183,727
Mortality, neonatal	−0.105	−0.0154	−0.0962	−0.0041
(per 1000 births)	(0.0016)	(0.0013)	(0.0016)	(0.0008)
	[−0.415]	[−0.061]	[−0.38]	[−0.016]
Adj. R^2	0.173	0.683	0.169	0.745
Sample size	189,036	189,036	183,727	183,727
5-min. APGAR score	0.1053	0.0117	0.1009	0.0069
(0–10 scale,	(0.0011)	(0.0012)	(0.0011)	(0.0011)
divided by 100)	[0.506]	[0.056]	[0.485]	[0.033]
Adj. R^2	0.255	0.663	0.248	0.673
Sample size	159,070	159,070	154,449	154,449
Ventilator incidence	−0.0837	−0.0039	−0.081	−0.002
(per 1000 births)	(0.0015)	(0.0017)	(0.0015)	(0.0016)
	[−0.228]	[−0.011]	[−0.221]	[−0.005]
Adj. R^2	0.052	0.706	0.05	0.716
Sample size	189,036	189,036	183,727	183,727
Ventilator ≥30 min.	−0.0724	0.0006	−0.0701	0.0016
(per 1000 births)	(0.0013)	(0.0013)	(0.0014)	(0.0012)
	[−0.252]	[0.002]	[−0.244]	[0.006]
Adj. R^2	0.063	0.724	0.062	0.739
Sample size	189,036	189,036	183,727	183,727

See notes to Tables I and II. 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 do 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 APGAR score, the coefficients are scaled up by 100. Numbers in square brackets indicate effect size in terms of standard deviations of the outcome per one standard deviation in birth weight (667 grams). There are no other variables included in the regressions.

effects are included in the second column, the costs estimate falls by a factor of 6, implying a $3200 decrease in hospital costs.

A similar dramatic drop in the magnitudes occurs for the other outcomes. The first and second columns show that the coefficients for mortality within one year, one day, one month fall—when including mother fixed effects—by factors of 5, 10, and 7, respectively. The parameter estimate for five-minute APGAR score similarly falls by a factor of 9. The magnitude for ventilator use falls by a factor of 20, and the point estimate for ventilator use for more than 30 minutes is statistically insignificant.

The third and fourth columns illustrate the sensitivity of the fixed effects estimates to the inclusion of twin pairs with congenital anomalies. Overall, the coefficients in the third column fall by no more than 10 percent, compared with those in the first column. On the other hand, the exclusion of these cases (which make up about 3 percent of the sample) cause the fixed-effects estimates to fall by additional factors of 2 or 3 (compared with those in the second column). After excluding these pairs, neither of the coefficients in the ventilator use regressions is statistically significant. The fixed effects estimates for the full sample are 5 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.[28]

We conclude that a substantial portion of the remaining within-twin correlation between birth weight and the outcomes is driven by one of the twins having a congenital abnormality. Since many birth defects (for example, Down's syndrome) are clearly not the result of low birth weight, we believe that cases that can be identified as congenital anomalies at birth should be excluded from the analysis. Nevertheless, since we do not have the congenital anomaly variable in the HCUP data, and to be more conservative in our conclusions, our cost calculations use the estimates from the sample that includes the pairs with congenital abnormalities.

Figures I and II 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 coeffi-

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

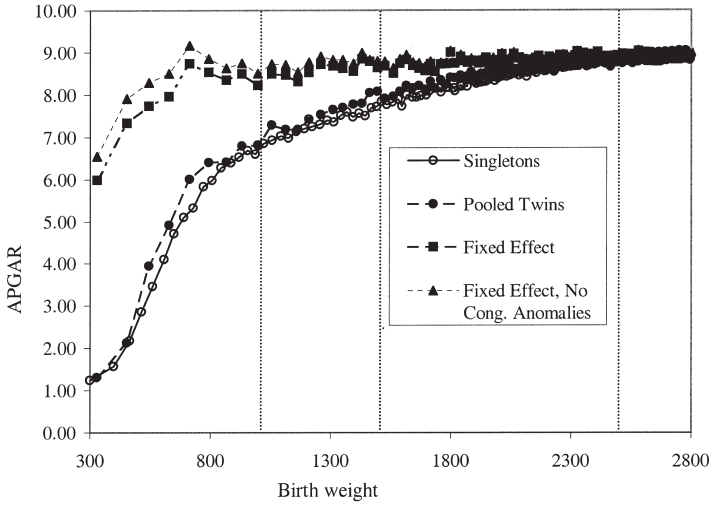

FIGURE IIa
Five-minute APGAR Score and Birth Weight
Note: Linked Birth-Death certificate data, 1989.

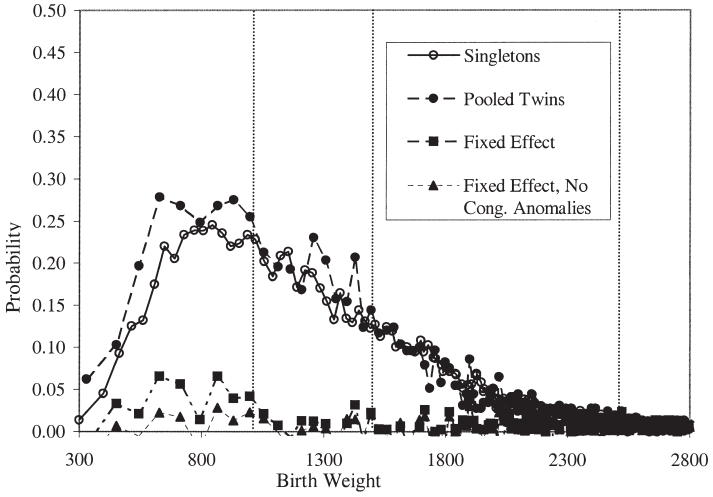

FIGURE IIb
Assisted Ventilation (30 minutes or more) and Birth Weight
Note: Linked Birth-Death certificate data, 1989.

cients β_k from the specification in equation (6).[29] When data on congenital anomalies are available, the solid triangles are the fixed-effects estimates when infants with those anomalies (and their cotwins) are excluded from the sample.

In Figure Ia 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.[30] The fixed effects estimates reveal a substantially smaller effect throughout the 800 to 2000 gram range of birth weights. The contrast between the cross-sectional and fixed-effects associations is more pronounced in the 2000 to 2500 gram range. For the cross section the decrease in costs continues to be substantial until the 2500 gram mark. By comparison, the fixed-effects estimates are relatively flat throughout that range of birth weights.

Figure Ib shows qualitatively similar results for mortality rates within one year of birth.[31] In the cross section the relationship between birth weight and mortality is steepest in the 300 to 1000 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 1000 to 1500 gram range.

The results for the five-minute APGAR score in Figure IIa show significant differences between the cross-sectional and fixed-effects "response functions" throughout the birth weight distribution. In the cross section the APGAR-birth weight relation is strongly positive even at birth weights well above 1000

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

30. 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.

31. Even though IUGR is emphasized in the literature, studies that have computed cost implications have not used LBW variation conditional on gestation length. To be consistent with this convention, our cross-sectional estimates do not include controls for gestation length. For completeness, however, we also derived cross-sectional estimates and figures that are adjusted for a full set of gestation-length (at the week level) dummies; the HCUP data do not include gestation length information. The results for infant mortality, five-minute APGAR score, and assisted ventilation are qualitatively similar to those presented here—controlling for mother fixed effects significantly reduces the implied "impact" of birth weight relative to cross-sectional estimates that are adjusted for gestation length (results available from the authors).

grams. The fixed-effects estimates reveal a somewhat positive relation in the 300–800 gram range, but little relation at birth weights above 800 grams.[32]

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

Figures Ib, IIa, and IIb 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 I and II 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 consistent with the notion that the cross-sectional relationships for singletons are subject to omitted variables biases of the same magnitude as for the twins population. Strictly speaking, however, we cannot rule out the possibility that singletons and twins have very different "structural" relations between birth weight and health outcomes, and also have different omitted variables that offset each other to yield a similar cross-sectional reduced-form relationship. This is one of the reasons why it is instructive to examine the singleton population by itself, using a completely different source of birth weight variation, as we do with maternal smoking in subsection IV.C.

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. Almond, Chay, and Lee [2004] document very similar results for overall incidence of ventilator use.

IV.B. Magnitudes of the Twins Effects

Table IV 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 IV 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 a $5 cost reduction. Table IV 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 cross-section" estimates. For the birth weight range of 800 to 2500 grams, the fixed-effects coefficients are 3.1 to 5.7 times smaller than the pooled coefficients. In the 2500–3000 gram range, the pooled coefficient implies a statistically significant $5 reduction in costs for every added gram in weight, but the corresponding fixed effects estimate implies a point estimate of no cost savings.

The next sets of columns in Table IV show qualitatively similar results for any ventilator use and ventilator use for more than 30 minutes. 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-sectional coefficients.

For infant mortality in the final two columns, the differences between the cross-sectional and fixed-effects estimates are particularly large at birth weights below 1000 grams. For example, in the 600 to 800 gram range, the cross-sectional estimate is -2.11 per 1000 live births, while the fixed-effects estimate is -0.42. On the other hand, in the 1000–1500 gram range the cross-sectional 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 Ic. For example, consider a 200 gram increase in birth weight. In the 600–800 gram range, the

TABLE IV

PIECEWISE 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 per gram		Usage per 1000 births per gram		Usage per 1000 births per gram		Score times 1000 per gram		Deaths per 1000 births per gram	
Birth weight segment	Percent	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.	Pooled	F.E.
<600 g	0.18	428.09	247.79	0.7995	0.2718	0.7434	0.2361	12.26	4.41	−0.5754	−0.4400
		(39.56)	(99.35)	(0.0640)	(0.1012)	(0.0578)	(0.0969)	(0.39)	(1.02)	(0.0640)	(0.1560)
600–800 g	0.16	186.59	−270.30	0.3027	−0.0052	0.2369	0.0261	9.38	2.19	−2.1122	−0.4197
		(62.03)	(133.79)	(0.1022)	(0.0914)	(0.0970)	(0.0891)	(0.58)	(1.06)	(0.0913)	(0.2139)
800–1000 g	0.19	−212.77	−49.75	−0.2164	−0.0772	−0.2136	−0.0549	3.04	0.16	−1.1302	−0.4332
		(47.89)	(67.31)	(0.0863)	(0.0797)	(0.0815)	(0.0792)	(0.40)	(0.81)	(0.0752)	(0.1669)
1000–1500 g	0.55	−121.20	−36.88	−0.2216	−0.0005	−0.2124	0.0168	2.12	0.60	−0.2258	−0.1693
		(7.57)	(9.89)	(0.0218)	(0.0243)	(0.0201)	(0.0232)	(0.09)	(0.22)	(0.0144)	(0.0336)
1500–2000 g	1.11	−74.27	−24.07	−0.2121	−0.0063	−0.1891	−0.0018	0.98	0.18	−0.0279	−0.0391
		(2.50)	(2.95)	(0.0110)	(0.0137)	(0.0096)	(0.0118)	(0.04)	(0.09)	(0.0051)	(0.0123)
2000–2500 g	3.95	−18.19	−3.21	−0.0671	−0.0091	−0.0469	0.0008	0.51	0.14	−0.0146	−0.0113
		(0.84)	(0.89)	(0.0052)	(0.0067)	(0.0038)	(0.0050)	(0.02)	(0.04)	(0.0022)	(0.0051)
2500–3000 g	15.54	−5.19	0.58	−0.0206	−0.0009	−0.0128	0.0014	0.27	0.05	−0.0076	−0.0043
		(1.45)	(0.56)	(0.0034)	(0.0050)	(0.0020)	(0.0032)	(0.02)	(0.03)	(0.0020)	(0.0034)
>3000 g	78.31	6.66	1.37	−0.0035	−0.0012	−0.0027	−0.0033	−0.02	−0.02	0.0120	−0.0019
		(3.31)	(0.83)	(0.0033)	(0.0053)	(0.0016)	(0.0031)	(0.02)	(0.03)	(0.0038)	(0.0036)
Observations		44,410	44,410	189,036	189,036	189,036	189,036	159,070	159,070	189,036	189,036

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. There are no other variables included in the regressions.

TABLE V
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
<600 g	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

The entries represent the average reduced hospital costs associated with increasing an infant's birth weight from the given birth weight category to above 2500 grams. These are calculated using the estimated spline coefficients in Table VI. 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.

cross-sectional 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-sectional and fixed-effects estimates, respectively.

Table V 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 2500 grams in the NCHS birth certificate files (1989–1991), we calculated how much money would be "saved" by raising their weight to 2500 grams. That is, we imputed the hospital costs in excess of the costs associated with 2500 grams for every infant, using the slope coefficients in the first two columns of Table IV. Table V 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

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 Ia and Table IV.

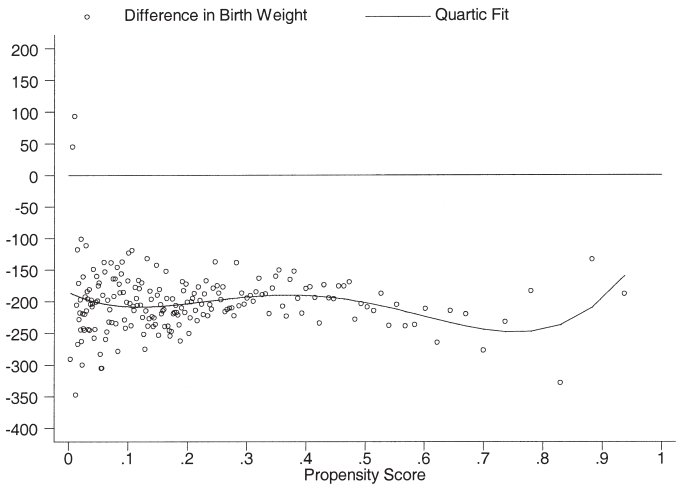

FIGURE IIIa
Difference between Smokers and Nonsmokers in Average Birth Weight by the Propensity Score, Pennsylvania Singletons 1989–1991

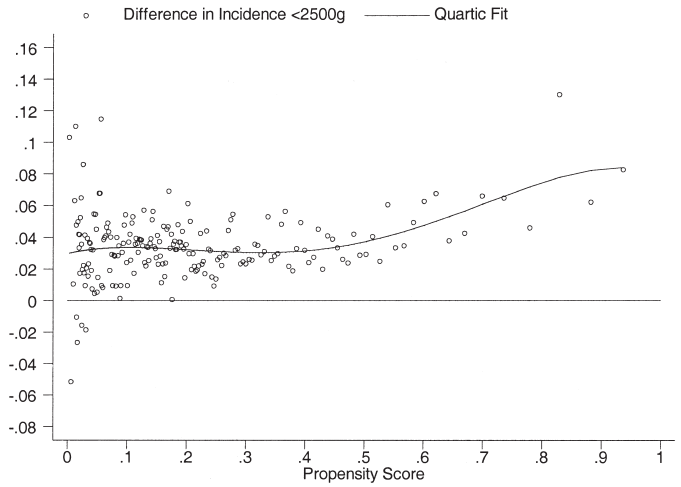

FIGURE IIIb
Difference between Smokers and Nonsmokers in Low Birth Weight (< 2500 grams) Incidence, Pennsylvania Singletons 1989–1991

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.

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 V to impute the aggregate excess costs for all infants born below 2500 grams. For twins born in the United States between 1989 and 1991, the cross-sectional estimates imply aggregate costs of LBW of $2.22 billion, while the fixed-effects estimates imply aggregate costs of $691 million. Table V 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-sectional estimates imply aggregate costs of LBW of $3.40 billion, while the fixed-effects estimates imply aggregate costs of $1.07 billion.

IV.C. Maternal Smoking Results for Singletons

Figure III 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 "pretreatment" 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

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. Chay, Flores, and Torelli [2005] find similar results for the other states in which smoking reporting was comprehensive—i.e., Florida, Georgia, Illinois, Michigan, North Carolina, and Ohio—and in other years of data.

36. The variables used are mother's age, education, race, ethnicity, marital status, foreign-born status; father's age, education, race, and ethnicity; dummies for trimester of first prenatal care visit, number of prenatal visits, and adequacy of care; controls for alcohol use and number of drinks per week; pregnancy history variables (parity indicator, live birth-order, number of previous births where newborn died, interval since last live birth, indicators for previous birth over 4000 grams and previous birth preterm or small-for-gestational-age); maternal medical risk factors that are not believed to be affected by smoking during pregnancy (anemia, cardiac disease, lung disease, diabetes, genital herpes, hyrdamnios/oligohydramnios, hemoglobinopathy, chronic hypertension, eclampsia, incompetent cervix, renal disease, Rh sensitization, uterine bleeding); month of birth and county of residence indicators. For the details of the specification, see Chay, Flores, and Torelli [2005].

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.

Figure IIIa plots the 200 smoking-nonsmoking differences in birth weight, which appear to be tightly scattered at around −200 grams, as the quartic polynomial fit illustrates. This difference is consistent with the regression-adjusted estimate of the birth weight effect of maternal smoking reported in the first row of Table VI (column 2).[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]).

Figure IIIb is analogous to Figure IIIa, but substitutes a binary indicator for birth weight below 2500 grams as the dependent variable. 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 VI. The LBW differential appears to be slightly larger at propensity scores above 0.7. A similar analysis finds no evidence of a systematic relation between maternal smoking and the incidence of birth weights below 1000 grams (see Almond, Chay, and Lee [2004]). Table VI shows that while there are significant raw differences in the incidences of birth weights below 1500 and 1000 grams (column 1), these differences disappear after regression adjustment for maternal and background characteristics (column 2)—maternal smoking only appears to affect birth weights for infants with weights above 1500 grams. In addition, the next rows of Table VI reveal 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.

Figure IVa is the analogue to Figures IIIa and IIIb, except that the dependent variable is now infant mortality within one year of birth. The graph reveals no systematic impact of maternal smoking on infant mortality throughout the propensity score distribution. The bottom two rows of Table VI show that while

37. The regression analysis adjusts for the 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.

TABLE VI
OLS ESTIMATES OF EFFECT OF MATERNAL SMOKING DURING PREGNANCY, PENNSYLVANIA SINGLETONS 1989–1991

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

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 "pretreatment" variables included in the propensity score model in Figures III and IV (see footnote 36 for the list of variables). The standard errors are in parentheses and are corrected for heteroskedasticity.

there is a significant raw difference in the infant and neonatal mortality rates of smokers and nonsmokers, this difference disappears after regression adjustment.[39]

39. The results are insensitive to dropping the maternal medical risk factors from the analysis and to dropping all higher-order and interaction terms from the regression specification. Also, we find similar results when analyzing the direct association between maternal smoking and the outcomes of twins, who are born at

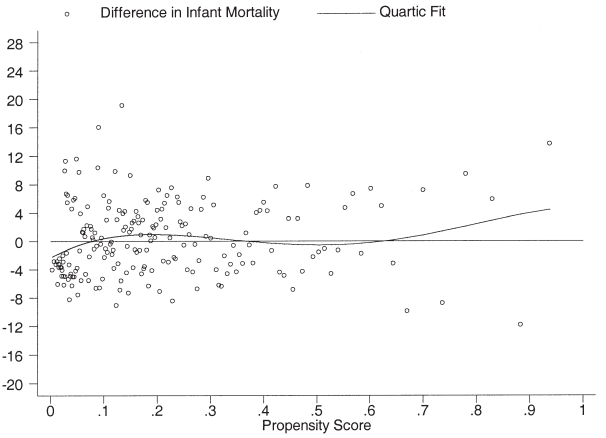

FIGURE IVa
Difference between Smokers and Nonsmokers in Infant Deaths within One Year of Birth (per 1000 Live Births) by the Propensity Score, Pennsylvania Singletons 1989–1991

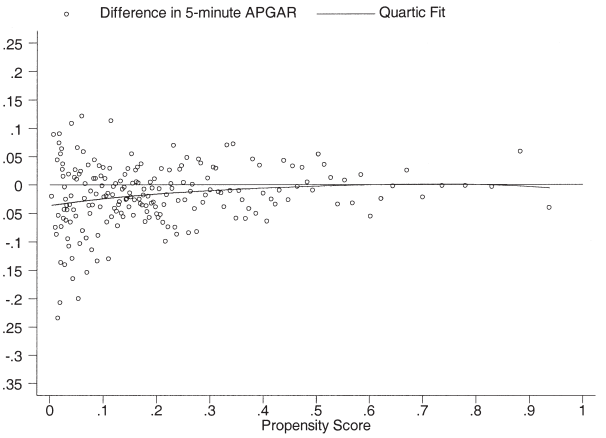

FIGURE IVb
Difference between Smokers and Nonsmokers in the Five-Minute APGAR Score, Pennsylvania Singletons 1989–1991

In addition, Figure IVb shows no impact of maternal smoking on the five-minute APGAR score. Panel c reveals no impact of maternal smoking on the probability that the newborn requires

much lower weight than singletons. We found no evidence of an interaction effect between birth weight and gestation length; e.g., the birth weight effect was equally small for term and preterm infants.

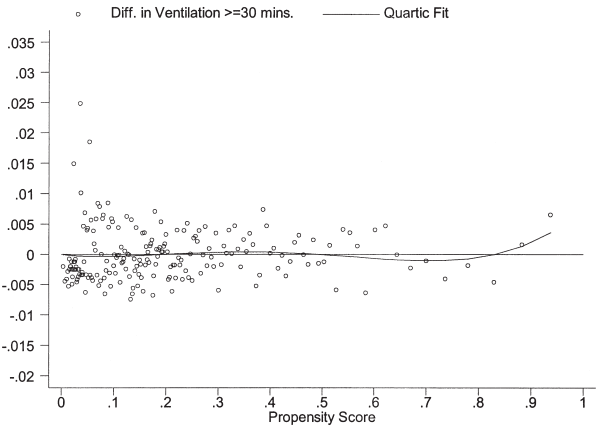

FIGURE IVc
Difference between Smokers and Nonsmokers in Incidence of Assisted Ventilation ≧ 30 Minutes, Pennsylvania Singletons 1989–1991

the use of a ventilator for more than 30 minutes. Again, the regression-adjusted estimates in Table VI are quite consistent with the figures.

Birth weight stands out as the *only* variable that has a strong empirical association with maternal smoking, once mother's characteristics are included in the analysis. 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 is 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 [Chay, Flores, and Torelli 2005]. This suggests that sample selection bias is probably not a first-order issue in analyses of maternal smoking that focus on the population of live births.

It is useful to compare the findings from our nonexperimental analysis with those of the medical literature on the effects of maternal smoking on infant birth weight and health. In particular, according to a Cochrane review [Lumley, Oliver, Chamberlain, and Oakley 2004] of 64 randomized trials of the impact of smoking cessation interventions on health outcomes of fetuses

and infants, the studies show, on average, a 33-gram increase in birth weight in response to programs that encourage pregnant women to cease smoking. On the other hand, there is no evidence of a statistically significant impact on the incidence of neonatal mortality—consistent with the results of our analysis—although due to relatively small sample sizes, standard errors were substantial, and even very large effects could not be ruled out.

Sexton and Hebel [1984] is one of the more well-known of these studies (and one of three which specifically examines neonatal deaths). First, it shows that the smoking cessation intervention led to a 20 percentage point reduction (from 83 percent for the control, to 63 percent for the treated) in late pregnancy smoking. They find a corresponding 92 gram effect in birth weights with a confidence interval of 15.4 to 168.61 grams. Inflating the effects by a factor of 1/0.2, this yields a point estimate of 460 grams with a confidence interval of 77 to 843 grams. Thus, our 200 gram estimate falls below the point estimate but well within the confidence interval range given by this study. Also, in our sample mothers who smoke during pregnancy consume, on average, 11 cigarettes-per-day, which implies a cigarette-per-day effect on birth weight of 18 grams. Sexton and Hebel find that the treated group consumes 6.4 cigarettes-per-day less than the control group in late pregnancy, implying a cigarette-per-day effect of 14.4 grams.[40]

Further, there is little-to-no association between treatment status and the incidence of birth weight less than 1500 grams, gestation length, and one- and five-minute APGAR scores. As for neonatal mortality, 2 out of 461 live births in the control group died, while 5 out of 454 in the treatment group died. Thus, mortality rates were actually smaller in the control group but the confidence interval for the relative (treatment to control) risk of mortality ranges from 0.50 to 13.02—too large to rule out very large impacts in either direction. Overall, it appears that the presumption that impacts of smoking on birth weight necessarily translate into effects of birth weight on mortality are not supported either by our (potentially biased, but statistically precise) nonexperimental statistical analysis or by the existing (unbiased, but statistically imprecise) experimental evidence.

40. It should be noted that the samples are not directly comparable, since the Sexton and Hebel study recruited smoking women, suggesting that the preintervention smoking intensities were probably particularly high.

Finally, we use the estimates in Table IV and Table VI 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 VI. The costs-birth weight gradients in Table IV then 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 IV 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 VI. Again, presume that maternal smoking reduces birth weight by 200 grams. The estimated cost savings fall by a factor of 17 to $53 for each woman who is prevented from smoking during pregnancy.[41] In Section VI we discuss the implications of this finding for cost-benefit analyses of smoking intervention programs targeted at pregnant women.

V. Robustness of the Findings

In this section we briefly summarize additional analyses—reported in greater detail in Almond, Chay, and Lee [2004]—that address important issues with our twins analysis. The first major concern is the possibility that twins are treated more or less identically, irrespective of their health status. That is, in principle, it is possible that if only one twin requires immediate medical attention and treatment (e.g., the need for assisted ventilation), the other receives the treatment as well. This would naturally lead to small fixed-effects estimates.

This hypothesis is rejected by the data—at least for the utilization and cost measures we use in our study. For example,

41. 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.

as Table I shows, about 27 percent (0.0102/0.037) of the pooled cross-sectional variation in the use of assisted ventilation (more than 30 minutes) is driven by the within-pair variation, which should be zero if twins are treated identically. Similarly, about 21 percent of the cross-sectional variation in costs is driven by within-twin pair variation. Both of these proportions are greater than the 17 percent of cross-sectional birth weight variation that is explained by within-pair differences.

Furthermore, if twins are treated the same independent of health status, then within-twin pair differences should be uncorrelated with within-pair differences in treatment. As documented in Almond, Chay, and Lee [2004], the fixed-effects estimate of the effect of the five-minute APGAR score on ventilator use is as large in magnitude as the pooled cross-sectional estimate and highly significant. In addition, the APGAR score performs notably better than birth weight in predicting within-twin pair differences in both one-year and one-day mortality.

A second important concern is that our analysis—while accounting for genetic differences across mothers—cannot control for genetic differences between fraternal twins of the same mother, since the data do not allow us to distinguish between monozygotic (identical) and dizygotic (fraternal) twins. Thus, some of our twins estimates may be driven by genetic differences between dizygotic twins of the same mother. Almond, Chay, and Lee [2002, 2004] provide a partial assessment of the differences in effects by zygosity, by examining the necessarily dizygotic infants of boy-girl twin pairs. Overall, the fixed-effects estimates across twin-types (boy-boy, girl-girl, boy-girl) are similar; more importantly, the significant fall in the estimates, when moving from the OLS to the fixed-effects specification, is of the same magnitude across types.[42]

In Almond, Chay, and Lee [2004] we also demonstrate that our fixed effects-estimates and qualitative conclusions are similar across: 1) different years of analysis (we examine the 1983–1985, 1989–1991, and 1995–1997 periods); 2) functional form assumptions (we examine the results in a logit/conditional-logit framework, as well as examine the appropriateness of the spline speci-

42. Note that the stability of the results across twin-types rules out the possibility that the small coefficients are caused by accidental "switching" of twins at birth, since such switching will not occur for boy-girl pairs. Conley, Strully, and Bennett [2003a] argue that among identical twins sharing a placenta, the fixed-effects estimate will *overstate* the importance of size at birth.

fication); 3) alternative matching algorithms for constructing twin pairs (we compare a "record-location" matching algorithm with a simple covariate-matching algorithm, and also examine the NCHS algorithm); 4) whether the first-born in a twin pair is heavier or lighter than the second-born (we test the equality and symmetry of the effects by twin birth order); and 5) subgroups defined by the race and education of the mother.

Finally, we note that the findings here and originally reported in Almond, Chay, and Lee [2002] are consistent with those of Conley, Strully, and Bennett [2003a], who use the NCHS 1995–1997 data to obtain within-pair estimates. For example, their main result is an 18.1 percentage increase in the probability of dying within the first year of life associated with a 1 pound (454 gram) decrease in birth weight. Eighteen point one percent, multiplied by the mean mortality rate of 38.71 deaths per 1000 live births, yields a difference of $0.181*38.71 = 7.01$ deaths per 1000 live births. Based on the linear regression, fixed-effects estimate for infant mortality in Table III (column 2), a 454-gram decrease leads to $(-454)*(-0.0222) = 10.07$ more deaths per 1000 live births. Thus, their estimate is slightly *smaller* in magnitude than our simple linear fixed-effects estimate. But as emphasized above, and shown in Table III and in Figure Ib, the *comparison* of the magnitudes between the cross-sectional relation and the fixed-effects estimates, as well as the sensitivity of the results to the inclusion of twins with congenital anomalies, suggests that the return to reducing LBW is many times smaller than typically presumed.

VI. Interpretation and Implications

VI.A. Implications for Policy

In two different settings—corresponding to birth weight variation driven by intra-twin-pair differences and variation driven by 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, Phibbs, and Glantz [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.[43] 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 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.[44]

The case of maternal smoking also highlights the policy relevance of the highly nonlinear relations between birth weight and infant outcomes that we document. For example, the second column of Table VI suggests that maternal smoking does not affect the incidence of births below 1500 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 sizable 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 nonexperimental, it may not identify

43. 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].

44. Our estimate of the excess costs due to LBW is the average of the rows in the second column of Table V (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, Phibbs, and Glantz [1999] exclude these births from their analysis.

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 dependent on which environmental factor is influencing birth weight. Cross-sectional birth weight variation is directly or indirectly influenced by immutable factors (genetics), socioeconomic factors (education, income), 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 birth weight variability due to variable intrauterine growth rates—holds constant all mother- and pregnancy-specific factors, and is partly generated by differences in fetal nutritional intake. Our smoking analysis, on the other hand, examines birth weight variation due to a maternal behavior that is the leading modifiable cause of LBW in the United States.

Our findings suggest that while some interventions may indeed succeed in both raising birth weights and improving health outcomes, others may only be effective in raising birth weights, with little or no 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.

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 Ib 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, as only one-half of 1 percent of all babies in the United States are born with a weight below 1000 grams.

VI.B. Prenatal versus Postnatal Investments

Several, recent studies have documented a relation between birth weight and longer run outcomes such as cognitive development, educational attainment, labor market outcomes, and adult

health.[45] Our analysis, which focuses on early infant outcomes, may provide some indirect evidence on the role of prenatal investments by the parents. However, it provides no evidence on the interactions between birth weight, overall health, and postnatal investments in the child, which is an important area of research in the economics literature (e.g., Becker and Tomes [1976] and Behrman, Rosenzweig, and Taubman [1994]). The relatively small effects of birth weight that we find for early infant health status suggest that previous studies' findings of significant impacts of birth weight on later adult outcomes is driven more by an interaction between postnatal investments and initial birth weight than by the effects of birth weight (or prenatal investments) per se.

Some evidence on this question can be marshaled from the medical and epidemiology literatures on the effects of parental smoking on child health and development. MacArthur et al. [2001], who analyze a randomized intervention, document no effect of maternal smoking in pregnancy on the child's IQ and height at age nine. The handful of nonexperimental studies that have longitudinal data on maternal (and paternal) smoking behaviors during and after pregnancy find that *postnatal* smoking is the dominating predictor (over prenatal smoking) for developmental and behavioral problems, cognitive performance, lower respiratory illness, and allergies.[46]

Finally, Chay, Flores, and Torelli [2005], find that Sudden Infant Death Syndrome (SIDS) is the only cause of infant death that has a significant correlation with maternal smoking after regression adjustment for mother's characteristics. However, whether this association is due to prenatal or postnatal smoke

45. See Strauss [2000], Matte et al. [2001], Jefferis, Power, and Hertzman [2002], Durousseau and Chavez [2003], Morley et al. [2004], and Behrman and Rosenzweig [2004]. Behrman and Rosenzweig [2001], based on a sample of 404 MZ female twin pairs, find intrapair birth weight effects on educational attainment and wages that are larger at normal weights than at weights below 2500 grams. Boomsma and Ball [1998], using samples of MZ and DZ twin pairs, find evidence that genetic differences between siblings accounts for much of the estimated effect of birth weight on IQ. 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. The 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, Stehouwer, and Boomsma [2000], Zhang et al. [2001], and Hübinnette et al. [2001]).

46. See DiFranza, Aligne, and Weitzman [2004], Ernst, Moolchan, and Robinson [2001], Eskenazi and Castorina [1999], Bauman, Flewelling, and Laprelle [1991], Yolton et al. [2005], Strachan and Cook [1997], Cook and Strachan [1999], and Kulig et al. [1999].

exposure is a topic of ongoing debate in the medical literature.[47] In addition, experimental animal studies have established significant effects of postnatal, environmental tobacco smoke (ETS) exposure on body and brain function, but little to no effect of prenatal exposure [Gospe, Zhou, and Pinkerton 1996; Slotkin et al. 2001]. The relative importance of prenatal and postnatal smoke exposure is of obvious health policy significance since women who quit smoking during pregnancy have high relapse rates following delivery [Fingerhut, Kleinman, and Kendrick 1990; Kahn, Certain, and Whitaker 2002].

VII. Conclusion

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 empirically important sources of birth weight variation, 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.[48] 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 thereafter. Therefore, we cannot make any conclusions about the relation between LBW (or birth weight more generally) and longer run outcomes. 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

47. See Klonoff-Cohen et al. [1995], Anderson and Cook [1997], Schoendorf and Kiely [1992], and Dwyer, Ponsonby, and Couper [1999]. Almond, Chay, and Lee [2002] find a significant (though small in magnitude) association between twin differences in birth weight and twin differences in postneonatal mortality and infant death due to SIDS. Two-thirds of the overall twin differences estimate of the effect of birth weight on postneonatal mortality is attributable to deaths due to SIDS or congenital anomalies at birth.

48. 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.

developmental and 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. A more direct investigation of this hypothesis seems to be a useful direction for future research.

Appendix 1: Data

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. Although we examined the results from both approaches, the results reported in this paper use the record-location matching approach. The second approach, necessary for examining births occurring before 1989 when multiple births were not located next to each other in the "denominator-plus" file, is described in Almond, Chay, and Lee [2004].

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 education, 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 percent) 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 sparser set of covariate checks is used.

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 who 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). The covariate-matching approach yielded larger sample sizes, so we used this algorithm for our analysis of the HCUP data.

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.

APPENDIX 2

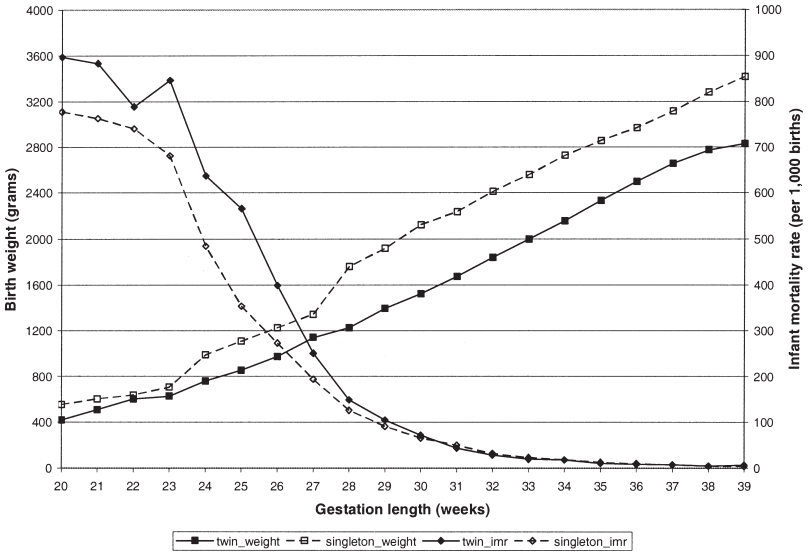

FIGURE V
Average Birth Weight and Infant Mortality Rate for Twins and Singletons, by Gestation Length
Note: Linked Birth-Death certificate data, 1989.

COLUMBIA UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF CALIFORNIA AT BERKELEY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF CALIFORNIA AT BERKELEY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

REFERENCES

Almond, Douglas, Kenneth Y. Chay, and Michael Greenstone, "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in Mississippi," mimeo, 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.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee, "The Costs of Low Birth Weight," NBER Working Paper No. 10552, June 2004.

Alexander, Duane, "Child Health: Birth Outcomes and Infant Mortality," Speeches and Reports, National Institute for Child and Human Development, 2000.

Anderson, H. Ross, and Derek G. Cook, "Passive Smoking and Sudden Infant Death Syndrome: Review of the Epidemiological Evidence," *Thorax,* LII (1997), 1003–1009.

Bauman, Karl E., Robert L. Flewelling, and John LaPrelle, "Parental Cigarette Smoking and Cognitive Performance of Children," *Health Psychology,* X (1991), 282–288.

Becker, Gary S., and Nigel Tomes, "Child Endowments, and the Quantity and Quality of Children," *Journal of Political Economy,* LXXXIV (1976), S143–S162.

Behrman, Jere R., and Mark R. Rosenzweig, "The Returns to Increasing Body Weight," PIER Working Paper No. 01-052, 2001.

Behrman, Jere R., and Mark R. Rosenzweig, "The Returns to Birth Weight," *Review of Economics and Statistics,* LXXXVI (2004), 586–601.

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, Richard, Robert M. Kliegman, and Hal B. Jenson, *Nelson Textbook of Pediatrics* (Amsterdam, The Netherlands: Elsevier Science, 2000).

Boomsma, Dorret I., and G. C. M. Ball, "Genetic Influences on Childhood IQ in 5- and 7-Year Old Dutch Twins," *Developmental Neuropsychology,* XIV (1998), 115–126.

Brooks, Anne-Marie, Robert S. Byrd, Michael Weitzman, Peggy Auinger, et al., "Impact of Low Birth Weight on Early Childhood Asthma in the United States," *Archives of Pediatric and Adolescent Medicine,* CLV (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., Carlos A. Flores, and Paul A. Torelli, "The Association between Maternal Smoking during Pregnancy and Fetal and Infant Health: New Evidence from United States Birth Records," mimeo, University of California, Berkeley, 2005.

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.

Chay, Kenneth Y., and Michael Greenstone, "Air Quality, Infant Mortality, and the Clean Air Act of 1970," NBER Working Paper No. 10053, 2003b.

Conley, Dalton, Kate W. Strully, and Neil G. Bennett, "A Pound of Flesh or Just Proxy? Using Twin Differences to Estimate the Effect of Life Chances," NBER Working Paper No. 9901, 2003a.

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

Cook, Derek G., and David P. Strachan, "Summary of Effects of Parental Smoking on the Respiratory Health of Children and Implications for Research," *Thorax,* LIV (1999), 357–366.

Corman, Hope, Theodore J. Joyce, and Michael Grossman, "Birth Outcome Production Functions in the U. S.," *Journal of Human Resources,* XXII (1987), 339–360.

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," *Quarterly Journal of Economics,* CXVIII (2003), 1495–1532.
Derom, Richard, Catherine Derom, and Robert Vlietinck, "Placentation," in Keith, et al., eds., *Multiple Pregnancy: Epidemiology, Gestation and Perinatal Outcome* (New York: The Parthenon Publishing Group, 1995).
DiFranza, Joseph R., C. Andrew Aligne, and Michael Weitzman, "Prenatal and Postnatal Environmental Tobacco Smoke Exposure and Children's Health," *Pediatrics,* CXIII (2004), 1007–1015.
Durousseau, Sharon, and Gilberto F. Chavez, "Associations of Intrauterine Growth Restriction among Term Infants and Maternal Pregnancy Intendedness, Initial Happiness about Being Pregnant, and Sense of Control," *Pediatrics,* CXI (2003), 1171–1175.
Dwyer, Terence, Anne-Louise Ponsonby, and David Couper, "Tobacco Smoke Exposure at One Month of Age and Subsequent Risk of SIDS—A Prospective Study," *American Journal of Epidemiology,* CXLIX (1999), 593–602.
Ernst, Monique, Eric T. Moolchan, and Miqun L. Robinson, "Behavioral and Neural Consequences of Prenatal Exposure to Nicotine," *Journal of the American Academy of Child and Adolescent Psychiatry,* XL (2001), 630–641.
Ershoff, Daniel H., Patricia Dolan-Mullen, and Virginia P. Quinn, "Randomized Trial of a Serialized Self-Help Smoking Cessation Program for Pregnant Women in an HMO," *American Journal of Public Health,* LXXIX (1989), 182–187.
Eskenazi, Brenda, and Rosemary Castorina, "Association of Prenatal Maternal or Postnatal Child Environmental Tobacco Smoke Exposure and Neurodevelopmental and Behavioral Problems in Children," *Environmental Health Perspectives,* CVII (1999), 991–1000.
Fingerhut, Lois A., Joel C. Kleinman, and Juliette S. Kendrick, "Smoking Before, During, and After Pregnancy," *American Journal of Public Health,* LXXX (1990), 541–544.
Fowler, Mary Glenn, Joel Kleinman, John L. Kiely, and Samuel Kessel, "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,* CCCXXXIX (1998), 313–320.
Gospe, Sidney, Shan Shan Zhou, and Kent E. Pinkerton, "Effects of Environmental Tobacco Smoke Exposure *in Utero* and/or Postnatally on Brain Development," *Pediatric Research,* XXXIX (1996), 494–498.
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 Economic Review,* LXXXVI (1996), 276–284.
Hübinnette, Anna, Sven Cnattingius, Anders Ekbom, Ulf de Faire, et al., "Birth Weight, Early Environment, and Genetics: A Study of Twins Discordant for Acute Myocardial Infarction," *The Lancet,* CCCLVII (2001), 1997–2001.
Ijzerman, Richard G., Coen Stehouwer, and Dorret Boomsma, "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, DC: National Academy Press, 1996).

Jefferis, Barbara, Chris Power, and Clyde Hertzman, "Birth Weight, Childhood Socioeconomic Environment, and Cognitive Development in the 1958 British Birth Cohort Study," *British Medical Journal,* CCCXXV (2002), 1–6.
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.
Kahn, Robert S., Laura Certain, and Robert C. Whitaker, "A Reexamination of Smoking Before, During, and After Pregnancy," *American Journal of Public Health,* XCII (2002), 1801–1808.
Keith, Louis G., Emile Papiernik-Berkhauer, and Donald M. Keith, *Multiple Pregnancy: Epidemiology, Gestation and Perinatal Outcome* (New York: The Parthenon Publishing Group, 1995).
Klonoff-Cohen, Hillary Sandra, Sharon Edelstein, Ellen S. Lefkowitz, Indu P. Srinivasan, et al., "The Effect of Passive Smoking and Tobacco Exposure through Breast Milk on Sudden Infant Death Syndrome," *Journal of the American Medical Association,* CCLXXIII (1995), 795–798.
Kogan, Michael D., Greg R. Alexander, Milton Kotelchuck, Marian MacDorman, et al., "Trends in Twin Birth Outcomes and Prenatal Care Utilization in the United States, 1981–1997," *Journal of the American Medical Association,* CCLXXXIV (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,* XCII (2002), 799–804.
Kramer, Michael S., "Intrauterine Growth and Gestational Duration Determinants," *Pediatrics,* LXXX (1987), 502–511.
Kulig, Michael, W. Luck, S. Lau, B. Niggerman, et al., "Effect of Pre- and Postnatal Tobacco Smoke Exposure on Specific Sensitization to Food and Inhalant Allergens during the First 3 Years of Life," *Allergy,* LIV (1999), 220–228.
Lewit, Eugene M., Linda Schuurmann Baker, Hope Corman, and Patricia Shiono, "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,* CIV (1999), 1312–1320.
Loos, Ruth J., Catherine Derom, Robert Derom, and Robert Vlietinck, "Determinants of Birth Weight and Intrauterine Growth in Liveborn Twins," *Paediatric and Perinatal Epidemiology,* XIX (2005), 15–22.
Lucas, Alan, Ruth Morley, and Tim J. Cole, "Randomised Trial of Early Diet in Preterm Babies and Later Intelligence Quotient," *British Medical Journal,* CCCXVII (1998), 1481–1487.
Lumley, Judith, Sandy S. Oliver, Catherine Chamberlain, and Laura Oakley, "Interventions for Promoting Smoking Cessation during Pregnancy," *The Cochrane Database of Systematic Reviews* (2004), Issue 3.
MacArthur, Christine, E. George Knox, and Robert J. Lancashire, "Effects at Age Nine of Maternal Smoking in Pregnancy: Experimental and Observational Findings," *British Journal of Obstetrics and Gynaecology,* CVIII (2001), 67–73.
Marks, James S., Jeffrey Koplan, Carol Hogue, and Michael Dalmat, "A Cost-Benefit/Cost Effectiveness Analysis of Smoking Cessation for Pregnant Women," *American Journal of Preventative Medicine,* VI (1990), 282–289.
Matte, Thomas D., Michaeline Bresnahan, Melissa D. Begg, and Ezra Susser, "Influence of Variation in Birth Weight within Normal Range and within Sibships on IQ at Age 7 Years: Cohort Study," *British Medical Journal,* CCCXXIII (2001), 310–314.
McCormick, Marie C., Jeanne Brooks-Gunn, K. Workman-Daniels, J. Turner, et al., "The Health and Developmental Status of Very Low-Birth Weight Children at School Age," *Journal of the American Medical Association,* CCLXVII (1992), 2204–2208.
Morley, Ruth, Mary S. Fewtrell, Rebecca A. Abbott, Terence Stephenson, et al., "Neurodevelopment in Children Born Small for Gestational Age: A Randomized Trial of Nutrient-Enriched versus Standard Formula and Comparison with a Reference Breastfed Group," *Pediatrics,* CXIII (2004), 515–521.

National Institutes of Health, "NIH Guide: Low Birth Weight in Minority Populations," PA-99-045, 1999.
Nelson, Karin B., and Judith K. Grether, "Cerebral Palsy in Low-Birth Weight Infants: Etiology and Strategies for Prevention," *Mental Retardation and Developmental Disabilities Research Reviews,* III (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,* XC (2000), 395–400.
Poulter, N. R., et al., "Association between Birth Weight and Adult Blood Pressure in Twins: Historical Cohort Study," *British Medical Journal,* CCCXIX (1999), 1330–1333.
Richards, Marcus, Rebecca Hardy, Diana Kuh, and Michael Wadsworth, "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,* LXX (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.
Rosenzweig, Mark R., and Kenneth I. Wolpin, "Sisters, Siblings, and Mothers: The Effects of Teen-Age Childbearing on Birth Outcomes," *Econometrica,* LXIII (1995), 303–326.
Schoendorf, Kenneth C., and John L. Kiely, "Relationship of Sudden Infant Death Syndrome to Maternal Smoking During and After Pregnancy," *Pediatrics,* XC (1992), 905–908.
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.
Slotkin, Theodore A., Kent Pinkerton, M. C. Garofolo, J. T. Auman, et al., "Perinatal Exposure to Environmental Tobacco Smoke Induces Adenylyl Cyclase and Alters Receptor-Mediated Cell Signaling in Brain and Heart of Neonatal Rats," *Brain Research,* DCCCXCVIII (2001), 73–81.
Stein, Zena, Mervyn Susser, Gerhart Saenger, and Francis Marolla, *Famine and Human Development: The Dutch Hunger Winter of 1944–45* (New York, NY: Oxford University Press, 1975).
Strachan, David P., and Derek G. Cook, "Parental Smoking and Lower Respiratory Illness in Infancy and Early Childhood," *Thorax,* LII (1997), 905–914.
Strauss, Richard S., "Adult Functional Outcome of Those Born Small for Gestational Age: Twenty-six-Year Follow-up of the 1970 British Birth Cohort," *Journal of the American Medical Association,* CCLXXXIII (2000), 625–632.
Vogler, George P., and Lynn T. Kozlowski, "Differential Influence of Maternal Smoking on Infant Birth Weight," *Journal of the American Medical Association,* CCLXXXVII (2002), 241–242.
Wilcox, Allen J., "On the Importance—and the Unimportance—of Birth Weight," *International Journal of Epidemiology,* XXX (2001), 1233–1241.
Windsor, Richard A., John B. Lowe, Laura L. Perkins, Dianna Smith-Yoder, et al., "Health Education for Pregnant Smokers: Its Behavioral Impact and Cost Benefit," *American Journal of Public Health,* LXXXIII (1993), 201–206.
Yolton, Kimberly, et al., "Exposure to Environmental Tobacco Smoke and Cognitive Abilities among U. S. Children and Adolescents," *Environmental Health Perspectives,* CXIII (2005), 98–103.
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,* CLIII (2001), 779–782.