Title	From the cradle to the labor market? The effect of birth weight on adult outcomes
Authors(s)	Devereux, Paul J.; Black, Sandra E.; Salvanes, Kjell G.
Publication date	2007-02
Publication information	Quarterly Journal of Economics, 122 (1): 409-439
Publisher	MIT Press Journals, Massachusets Institute
Link to online version	http://dx.doi.org/10.1162/qjec.122.1.409
Item record/more information	http://hdl.handle.net/10197/316
Publisher's version (DOI)	10.1162/qjec.122.1.409

FROM THE CRADLE TO THE LABOR MARKET? THE EFFECT OF BIRTH WEIGHT ON ADULT OUTCOMES*

Sandra E. Black
Paul J. Devereux
Kjell G. Salvanes

Lower birth weight babies have worse outcomes, both short-run in terms of one-year mortality rates and longer run in terms of educational attainment and earnings. However, recent research has called into question whether birth weight itself is important or whether it simply reflects other hard-to-measure characteristics. By applying within twin techniques using an unusually rich dataset from Norway, we examine the effects of birth weight on both short-run and long-run outcomes for the same cohorts. We find that birth weight does matter; despite short-run twin fixed effects estimates that are much smaller than OLS estimates, the effects on longer-run outcomes such as adult height, IQ, earnings, and education are significant and similar in magnitude to OLS estimates.

I. Introduction

Lower birth weight babies have worse outcomes, both short-run in terms of one-year mortality rates and longer run in terms of educational attainment and earnings. But is this relationship causal? Recent research has provided conflicting evidence, leaving us wondering whether birth weight itself is important or whether it simply reflects other hard-to-measure characteristics.

Understanding both the short-run and long-run effects of birth weight is important from a number of perspectives. On the policy side, governments have implemented a number of policies to improve the health of babies and, hence, their later outcomes. Consider, for example, the Women, Infants, and Children Program (WIC) in the United States, a federally funded program that provides nutrition counseling and supplemental food for pregnant

* Black and Devereux gratefully acknowledge financial support from the National Science Foundation and the California Center for Population Research. Salvanes thanks the Norwegian Research council for financial support. We are grateful to the Medical Birth Registry of Norway for providing the birth records data. We thank Marianne Bitler, Amitabh Chandra, Anne Daltveit, Susan Dynarski, Kanika Kapur, Linda Loury, Per Magnus, and Ole Martin Sundet for very useful discussions and seminar participants at the Federal Reserve Bank of New York, Boston College, Brown University, Princeton University, UCD Dublin, Tufts University, CEPR Upsala, Stockholm University, Socialforskningsinstituttet Copenhagen, Warwick, Essex, Maynooth College, Trinity College Dublin, Queens University Belfast, University College London, London School of Economics, University of Amsterdam, University of Bergen, and COST Paris for helpful comments. This research was conducted while Black was on leave at the Industrial Relations Section at Princeton University.

The Quarterly Journal of Economics, February 2007

women, new mothers, infants, and children under age five in order to improve child health and aid long-term health, growth, and development. A key presumption underlying this type of policy is that, by affecting birth weight through improved prenatal nutritional intake, it will in turn affect the health and ultimate success of the children.[1] Recent evidence suggesting little effect of birth weight on short-run outcomes may understate the true impact of these policies if there are significant longer-run effects.

Until recently, analysis of birth weight effects has relied primarily on cross-sectional variation and has established a relationship between low birth weight and poor health, cognitive deficits, and behavioral problems among young children. It has also provided evidence that this relationship persists for longer-term outcomes such as health status, educational attainment, employment, and earnings [for example, Barker 1995, Currie and Hyson 1999, Case et al. 2004].[2] However, it is possible that there are no underlying causal relationships, as low birth weight may be correlated with many difficult-to-measure socio-economic background and genetic variables.

Most recently, the literature has moved to within-twin variation to identify the effects of birth weight.[3] Both Conley et al. [2006] and Almond et al. [ACL 2005] use U.S. data to identify the effects of birth weight on short-run health outcomes, including mortality. Almond et al. conclude that the effects of low birth weight are substantially smaller than originally thought, and Conley et al. have estimates of similar magnitudes. However, neither of these studies is able to look beyond short-run health outcomes.

In contrast, Behrman and Rosenzweig [BR 2004] use a subset

1. Additionally, birth weight is very commonly used as the outcome variable of interest in studies of the effects of policy interventions such as welfare reform, health insurance, and food stamps on infant welfare (for example, Currie and Gruber [1996]), and in analyses of the impact of maternal behavior on infant health. (For example, Currie and Moretti [2003] show that increased maternal education leads to a lesser incidence of low birth weight (LBW).

2. Typically, medical studies have limited data on longer-run outcomes and small sample sizes. For example, Hack et al. [1994] finds an effect of very low birth weights on school-age outcomes using sixty-eight treatment children using across family comparisons, and Hack et al. [2002] compare 242 very low birth weight young adults to 233 normal birth weight controls and find that the educational disadvantage associated with very low birth weight persists into early adulthood. Recent work in the Norwegian medical literature also finds a positive relationship between birth weight and adult outcomes [Eide et al. 2005 and Grjibovski et al. 2005].

3. Additionally, sibling fixed effects approaches are taken by Conley and Bennett [2000], who find a negative association between LBW and timely high school graduation using U.S. panel data, and by Currie and Moretti [2005] who use birth records from California and find evidence of significant effects of own birth weight on income at time of childbirth as well as on the birth weight of the child.

of the Minnesota Twin Registry to do fixed effects using female monozygotic twins and examine the longer run effects of birth weight. They find evidence that the heavier twin goes on to be taller, have greater educational attainment, and have a higher wage, and the twin fixed effects estimates are substantially larger than the cross-sectional ones. In contrast, they find no evidence of effects on adult body mass index.

The conflicting evidence on short-run versus long-run outcomes could be real or could reflect the fact that BR rely on self-reported survey data and do not have access to comprehensive birth record data like that of ACL. As a result, their sample sizes are small (804 cases) and, because of the numerous surveys required, there is substantial attrition and item nonresponse that may not be random. Also, their use of survey data means that their outcome variables are self-reported and, unlike ACL, they cannot exclude twin pairs with congenital defects.

In this paper, we use rich administrative data on the population of Norway linked to birth records; with this, we can study both short- and long-run outcomes using large nationally representative samples that contain both administrative records of later outcomes as well as all the birth information contained in the birth register. We advance the recent literature by using twin fixed effects on a large sample of individuals to look at both short- and long-run outcomes for the same cohort of individuals. Our sample also differs from BR in that we study both men and women and analyze more recent cohorts (1967–1981 compared to their 1936–1955 cohorts). As such, the technology of birth and social conditions growing up should be more similar to those in the present day.[4]

We find that birth weight does matter. Consistent with earlier work, we find that twin fixed effects estimates of the effect of

4. In the process of completing a revision of the November 2005 version of this paper we became aware of two recently completed working papers on this topic. Oreopoulos et al. [2006] use Canadian administrative data and sibling and twin fixed effects to examine both short- and long-run effects of birth weight. Their results are similar to our own, although they are limited by small samples of twins and the outcomes they examine are different; they focus on mortality, physician visits, high school tests, grades completed by age 17, and social assistance receipt, while we examine high school completion, IQ, BMI, height, labor force participation, earnings, and intergenerational transmission. Royer [2005] uses within-twin variation and California administrative data to examine the effect of birth weight on educational attainment along with intergenerational transmission of birth weight and concludes that long-run effects are small. Her education analysis is limited by the fact that educational attainment is only observed if the woman has children in the sampling period. When we restrict our sample in a similar manner, we obtain similar results for our education variable.

birth weight on short-run outcomes such as one-year infant mortality are much smaller than their cross-sectional equivalents. However, studying only short-run outcomes may lead to incorrect inferences about the longer-run effects of birth weight; we find that birth weight has a significant effect on longer-run outcomes such as height, IQ at age 18, earnings, and education, and the fixed effects estimates are similar in size to cross-sectional ones.

When studying long-run outcomes, an important selection issue arises because twin pairs that experience infant mortality are dropped from the analysis. Because, unlike previous studies, we have information on individuals from birth to the labor market, we can investigate the potential impacts of such bias. Our investigation concludes that selection bias most likely leads to an understatement of the effects of birth weight on adult outcomes.

The paper unfolds as follows. Sections II and III discuss our methodology and data. Section IV presents our results. Section V focuses on our robustness checks, including an examination of the selection bias that might arise when studying adult outcomes. Section VI addresses issues of generalizability and Section VII concludes.

II. Conceptual Framework

Following ACL, let

$$y_{ijk} = \alpha + \beta bw_{ijk} + x_{jk}'\gamma + f_{jk} + \varepsilon_{ijk} \tag{1}$$

where subscript i refers to the child, j refers to the mother, and k refers to birth. y_{ijk} is then the outcome of child i born to mother j in birth k, bw_{ijk} is birth weight, x_{jk} is a vector of mother- and birth-specific variables (for example, mother's education, the year of birth), f_{jk} refers to unobservables that are mother- and birth-specific (for example, the quality of prenatal care, genetic factors), and ε_{ijk} is an idiosyncratic error term assumed independent of all other terms in the equation.

Cross-sectional estimation of equation (1) by OLS will generally lead to biased estimates of β because of the presence of elements of f_{jk} that influence both birth weight and child outcomes (for example, family background variables). Therefore, we take a twin fixed effect approach to estimation. That is, our sample is composed of twin pairs and we included dummy variables for each birth in the regression. Denoting the first-born twin as "1" and the second-born as "2," this can be written in differences as follows:

$$y_{1jk} - y_{2jk} = \beta(bw_{1jk} - bw_{2jk}) + (\varepsilon_{1jk} - \varepsilon_{2jk}) \quad (2)$$

Given the assumption that within-twin differences in ε_{ijk} are independent of within-twin differences in bw_{ijk}, the twin fixed effects estimator of β is consistent. This assumption is more likely to hold in the case of monozygotic twins (who are genetically identical) than with fraternal twins (who on average share about 50 percent of genes). Our full sample contains both monozygotic and fraternal twins. The medical literature suggests that adult health outcomes among fraternal twins are similar to those among identical twins [Christensen et al. 1995, Duffy and David 1993]. Consistent with this finding, we cannot reject the hypothesis that the relationship between birth weight and adult outcomes is the same for both types of twins when we examine a subset of twins for whom we have information on zygosity (see Section V).

The control variables we use in the OLS estimation are year- and month-of-birth dummies, indicators for mother's education (one for each year), indicators for birth order (which is known to be correlated with birth weight and is also a strong predictor of outcomes in Norway, see Black, Devereux, and Salvanes [2005a]), indicators for mother's year of birth (one for each year to allow for the fact that age of mother at birth may have independent effects on child outcomes), and an indicator for the sex of the child. With twin fixed effects, all controls are differenced out except the indicators for sex and birth order (either first born or second born twin).

II.A. Why Does Birth Weight Differ?

Low birth weight can arise either because of short gestational length (preterm delivery) or because of low fetal growth rate, commonly known as intrauterine growth retardation (IUGR). When we look within twin pairs, gestation length is the same and differences in birth weight arise solely due to differences in fetal growth rates.[5]

Given that gestation is the same among twins, evidence suggests that much of the difference in birth weight is due to differences in nutritional intake.[6] In the case where there are two

5. While there are rare cases of twins who are not born at the same time, these twins are not included in our sample. We also drop twin pairs for which gestation length is unknown (about 4 percent of cases)

6. Because twins have the same gestation, we cannot examine the effect of being preterm (gestation less than thirty-seven weeks) on outcomes. We did, however, verify that there were no significant differences in the effects of birth

placentas (called dichorionic, including all fraternal twins and about 30 percent of identical twins), nutritional differences can arise because one twin is better positioned in the womb. Among single-placenta (monochorionic) twins, nutritional differences have been related to the location of the attachment of the two umbilical cords to the placenta [Bryan 1992, Phillips 1993]. Hence, since there are no genetic differences, birth weight differences within monozygotic twin pairs appear to come primarily from differences in nutritional intake.[7]

As emphasized by ACL, differences between cross-section and twin fixed effects estimates can support two different interpretations. One is that there is a homogenous birth weight effect and the cross-sectional estimate is inconsistent. Another is that different sources of variation in birth weight have different effects on child outcomes. That is, birth weight is not in itself a policy variable, and different policies that affect birth weight may have very different effects on other outcomes.

III. Data

Our primary data source is the birth records for all Norwegian births over the period 1967–1997 obtained from the Medical Birth Registry of Norway. All births, including those born outside of a hospital, are included as long as the gestation period was at least sixteen weeks.[8] The birth records contain information on year and month of birth, birth weight, gestational length, age of mother, and a range of variables describing infant health at birth. In these data, we are also able to identify twin births and the birth order of twins but cannot distinguish between fraternal and monozygotic twins. We drop twin pairs where either twin was born with a congenital defect (approximately 2.1 percent), as this suggests an underlying difference between the twins.

Using unique personal identifiers, we match these birth files to the Norwegian Registry Data, a linked administrative dataset that covers the population of Norwegians aged 16–74 in the

weight on later outcomes between preterm and full-term babies. For one-year mortality, birth weight is more important for preterm twin pairs. About 35 percent of twins are born pre-term in our sample.

7. There is an extensive medical literature examining the determinants of birth weight differences (called discordance) among twins. See Blickstein and Kalish [2003] for a summary.

8. The data also include stillbirths, which constitute approximately fifteen per 1,000 births. We exclude these from the sample.

1986–2002 period and is a collection of different administrative registers such as the education register, family register, and the tax and earnings register. These data are maintained by Statistics Norway and provide information about educational attainment, labor market status, earnings, and a set of demographic variables (age, gender) as well as information on families.[9]

Another source of data is the Norwegian military records from 1984 to 2005, which contain information on height, weight, and IQ. In Norway, military service is compulsory for every able young man. Before entering the service, their medical and psychological suitability is assessed; this occurs for the great majority between their eighteenth and twentieth birthday.[10] We match these data with our other data files and use the height, BMI, and test score data as outcome variables for men.[11]

Our final dataset is a survey of twins born from 1967 through 1979 that contains information on zygosity and can be matched to the administrative data. The survey includes information on twin pairs that were intact at age three and was collected in two waves, one in 1992 and one in 1998. This is the only survey we use that is based on voluntarily self-reported information. As a result, we only have zygosity information for surviving twin pairs who completed the survey questionnaire (approximately 64 percent of those contacted).[12]

In the literature, different variants of birth weight have been used as the primary variable of interest. These include birth weight, log(birth weight), fetal growth (defined as birth weight divided by weeks gestation), and an indicator for low birth weight (<2,500 grams). Given that there is no obvious choice a priori, we have examined the explanatory power of these variables in the

9. Our measure of child educational attainment is reported by the educational establishment directly to Statistics Norway, thereby minimizing any measurement error due to misreporting. This educational register started in 1970. See Møen, Salvanes, and Sørensen [2003] for a description of these data.

10. Of the men in the 1967–1987 cohorts, 1.2 percent died before one year and 0.9 percent died between one year of age and registering with the military at about age 18. About 1 percent of the sample of eligible men had emigrated before age eighteen, and 1.4 percent of the men were exempted because they were permanently disabled. An additional 6.2 percent are missing for a variety of reasons including foreign citizenship and missing observations. See Eide et al. [2005] for more details.

11. There is an extensive literature suggesting that height is a useful indicator of health, both in developed as well as developing nations. See Strauss and Thomas [1998] for references.

12. Zygosity assignment is based on questionnaire items about co-twin similarity during childhood. These classification techniques are considered to have a very high rate of correct classification (greater than 96 percent). See Harris, Magnus, and Tambs [2002] for more details.

twin fixed effects regressions. They indicate that log(birth weight) provides the best fit for all outcome variables. Thus, we use this variable in our analysis. Estimates are very similar when either of the other two continuous measures are used [Black, Devereux, and Salvanes 2005c].[13]

The outcomes we study are as follows:

Infant Mortality. This comes from the birth records and is defined as mortality within the first year of life. We have this variable for the full 1967–1997 period.

Five-Minute APGAR Score. APGAR scores are a composite index of a child's health at birth and take into account Activity (and muscle tone), Pulse (heart rate), Grimace (reflex irritability), Appearance (skin coloration), and Respiration (breathing rate and effort). Each component is worth up to two points for a maximum of ten. This measure comes from the birth register and is available beginning in 1977.

Height, BMI, and Ability. For the cohorts of men born from 1967 up to 1987, we have information from the military records on height, weight, and Body Mass Index (BMI, defined as kilograms divided by meters squared), all of which were measured as part of the medical examination. We also have a composite score from three speeded IQ tests—arithmetic, word similarities, and figures (see Sundet et al. [2004 2005] and Thrane [1977] for details). The composite IQ test score is an unweighted mean of the three subtests. The IQ score is reported in stanine (Standard Nine) units.[14]

Education. For the cohorts born between 1967 and 1981 (and who are therefore at least twenty-one in 2002), we create a binary

13. It is interesting to note that the LBW indicator fits most poorly for all outcomes. This suggests that using cutoffs such as <2,500 grams as the variable of interest may not be appropriate for this type of analysis. We have also tried including both ln(birth weight) and an indicator for LBW (<2,500 grams) in the same specifications. The continuous measure dominates for all outcomes and the effect of LBW is always statistically insignificant and often has the wrong sign. In the same vein, we have also tried including both ln(birth weight) and birth length; with the exception of height at age eighteen, birth length is always dominated by ln(birth weight).

14. The arithmetic test is quite similar to the Wechsler Adult Intelligence Scale (WAIS) [Sundet et al. 2005; Cronbach 1964]. The word test is similar to the vocabulary test in WAIS, and the figures test is similar to the Raven Progressive Matrix test [Cronbach 1964]. Stanine units are a method of standardizing raw scores into a nine point standard scale with a normal distribution, a mean of five, and a standard deviation of two.

indicator for whether the person has at least twelve years of education.[15]

Labor Market Outcomes. We look at attachment to the labor force by studying whether individuals who are aged greater than or equal to twenty-five are full-time, full-year workers in 2002 (the last year of our panel). To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995, and in the fourth quarter thereafter).[16] We label these individuals as full-time workers. This includes about 62 percent of men and 43 percent of women in our sample.

We also study the earnings of full-time full-year employees, measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We use the most recent year of earnings in which we observe earnings for both twins, provided the twins are aged at least twenty-five in that year. Because of the age restrictions, the labor market variables are for 1967–1977 cohorts.

Birth Weight of First Child. Finally, we also examine whether or not there is evidence of intergenerational transmission of birth weight. This sample consists of women born between 1967 and 1988 whose first births occurred by 2004.[17] If the first birth is a twin birth, the woman is dropped from the sample. The outcome variable is the birth weight of the first born child.

III.A. Summary Statistics

Table I presents summary statistics for our sample. Statistics are broken down into twin and singleton samples in Columns 1 and 2 and the twin sample is reduced to same-sex twin pairs by sex in Columns 3 and 4. Figure I shows the substantial variation in birth weight within twin pairs; 21 percent of the variation in

15. While we describe this as high school completion, in Norway many individuals with twelve years of education obtain vocational rather than academic qualifications. We also tried using a more continuous measure of educational attainment as our dependent variable. However, because this necessitated restricting the sample further (aged twenty-five or older, at a minimum), our standard errors became quite large and we were not able to draw any real conclusions.

16. An individual is labeled as employed if currently working with a firm, on temporary layoff, on up to two weeks of sickness absence, or on maternity leave.

17. To get information on births up to 2004, we used a more recent birth register that has information on births between 1998 and 2004.

TABLE I
SUMMARY STATISTICS

	Singletons	Twins	Same-sex male twins	Same-sex female twins
Child's characteristics (1967–1997)				
Infant birth weight	3528 (558)	2598 (612)	2594 (639)	2540 (599)
Fraction low birth weight (<2500 grams)	.03 (.18)	.39 (.49)	.39 (.49)	.42 (.49)
Gestation (weeks)	39.83 (2.17)	36.90 (3.18)	36.62 (3.30)	37.02 (3.20)
Fetal growth	88.46 (13.07)	69.84 (13.81)	70.14 (14.38)	68.06 (13.47)
Fraction female	.49 (.50)	.50 (.50)	0	1
Fraction with complications	.31 (.46)	.49 (.50)	.49 (.50)	.49 (.50)
Mother's characteristics (1967–1997)				
Education	11.25 (2.64)	11.53 (2.61)	11.54 (2.60)	11.52 (2.62)
Age	26.64 (5.23)	28.06 (5.12)	27.81 (5.11)	27.74 (5.18)
N	1,595,203	33,366	11,528	11,284
Short-run outcomes				
1 year mortality rate (per 1000 births) (1967–1997)	6.23 (78.68)	31.11 (173.62)	41.20 (198.77)	28.00 (164.99)
Five-minute APGAR score (1977–1997)	9.29 (.75)	9.01 (1.10)	8.95 (1.19)	9.02 (1.10)
Education (1967–1981)				
High school completion	.73 (.44)	.74 (.44)	.74 (.44)	.75 (.43)
Military data (male sample 1967–1987)				
Height (centimeters)	179.96 (6.51)	—	179.34 (6.57)	—
BMI (kilograms/meters^2)	22.50 (3.38)	—	21.84 (2.90)	—
IQ (stanines)	5.20 (1.79)	—	5.08 (1.82)	—

TABLE 1
(CONTINUED)

	Singletons	Twins	Same-sex male twins	Same-sex female twins
Earnings data (1967–1977)				
Fraction working full time	.55 (.50)	.53 (.50)	.62 (.49)	.43 (.50)
Earnings for full time workers	297,834 (161,914)	295,962 (127,805)	337,888 (137,214)	250,619 (95,357)
Intergenerational transmission (female sample 1967–1984)				
Birth weight of first child	3,465 (619)	—	—	3,490 (614)

Standard deviations are given in parentheses. Fetal growth is calculated as birth weight divided by weeks gestation. Years in parentheses indicate the cohorts for which data is available. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests-arithmetic, word similarities, and figures [see Sundet et al. 2004 2005, Thrane 1977 for details]. The arithmetic test is quite similar to the Wechsler Adult Intelligence Scale (WAIS) [Sundet et al. 2005, Cronbach 1964]. The word test is similar to the vocabulary test in WAIS, and the figures test is similar to the Raven Progressive Matrix test [Cronbach 1964]. The composite IQ test score is an unweighted mean of the three subtests. The IQ score is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995 and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, the sample consists of women born between 1967 and 1988 whose first births occurred by 2004. If the first birth is a twin birth, the woman is discarded from the sample.

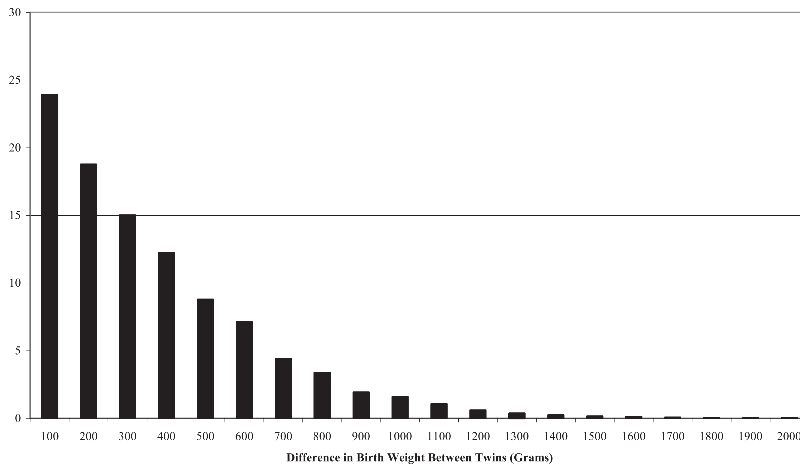

FIGURE I
Distribution of Differences in Birth Weight of Twins
Each bar represents the percentage of twins whose birth weight difference falls within the specified range. The first bar is 0–100 gram differences, the second bar is 101–200, etc. The mean birth weight difference among twins in our sample is 320 grams. The sample includes all twins born between 1967 and 1997 in Norway.

birth weight is within-twin. Table II reports sample averages for heavier and lighter same-sex twins. It is clear that heavier twins have better outcomes on average than lighter twins.

IV. RESULTS

As discussed earlier, different outcome variables are available for different cohorts. Initially, we maximize precision by using all available cohorts for each measure. Later, we show results when we study different outcomes using the exact same cohorts and even the exact same observations.

We first examine the sample of all twins and compare the results when we use pooled OLS versus a twins fixed-effect estimation strategy. Table III presents these estimates. Each coefficient represents the estimate from a separate regression. We present the results in approximate chronological order so that outcomes measured earlier in the life-cycle come first.

IV.A. *Short Run Outcomes: Mortality and Five-Minute APGAR Score*

For mortality, the pooled OLS coefficient of −280 implies that a 10 percent increase in birth weight would reduce one-year

TABLE II
SUMMARY STATISTICS: SAME-SEX TWINS

	Heavier	Lighter	T-statistics (difference in means)
Infant birth weight			
Mean	2726 (611)	2415 (586)	
Median	2800	2490	
Twenty-fifth percentile	2400	2080	
Tenth percentile	1940	1640	
Fifth percentile	1570	1310	
First percentile	860	730	
Fraction low birth weight (<2500 grams)	.30 (.46)	.51 (.50)	
Fetal growth	73.35 (13.39)	64.97 (13.22)	
Fraction with complications	.48 (.50)	.50 (.50)	
ln(birth weight)	.97 (.28)	.84 (.30)	
N	22,366		
Outcomes			
1 year mortality rate (per 1,000 births) (N = 22,366)	32.55 (177.46)	34.96 (183.70)	1.48
Five-minute APGAR score (N = 14410)	9.01 (1.11)	8.96 (1.16)	3.19
Height (males only) (N = 5264)	179.67 (6.57)	178.99 (6.56)	7.34
BMI (males only) (N = 5254)	21.90 (2.89)	21.77 (2.86)	2.62
IQ (males only) (N = 4804)	5.10 (1.81)	5.04 (1.82)	1.99
High school graduation rate (N = 8832)	.76 (.43)	.73 (.44)	3.08
Percentage working full time (N = 6446)	.52 (.50)	.52 (.50)	.03
ln(earnings) for full time workers (N = 4020)	12.52 (.45)	12.52 (.48)	.48
ln(birth weight of first child) (N = 1832)	8.15 (.23)	8.12 (.28)	2.05

Standard deviations are given in parentheses. N indicates the number of twins. T-statistics for the difference in means between heavier and lighter twins have been adjusted to reflect the covariance between the samples. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures [see Sundet et al. 2004 2005, Thrane 1977 for details]. The arithmetic test is quite similar to the Wechsler Adult Intelligence Scale (WAIS) [Sundet et al. 2005, Cronbach 1964]. The word test is similar to the vocabulary test in WAIS, and the figures test is similar to the Raven Progressive Matrix test [Cronbach 1964]. The composite IQ test score is an unweighted mean of the three subtests. The IQ score is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995 and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, the sample consists of women born between 1967 and 1988 whose first births occurred by 2004. If the first birth is a twin birth, the woman is discarded from the sample.

TABLE III
REGRESSION RESULTS: TWINS SAMPLE COEFFICIENT ON LN (BIRTH WEIGHT)

Dependent variable	Singleton sample OLS	Singleton sample Family fixed effects	Twins sample OLS	Twins sample Twin fixed effects
One-year mortality	−123.46** (1.71)	−186.71** (.69)	−279.64** (9.12)	−41.10** (7.64)
N	1,253,546		33,366	
Five minute APGAR score	.73** (.01)	1.08** (.01)	1.46** (.06)	.35** (.07)
N	674,577		21,580	
Height (males only)	11.03** (.11)	7.33** (.12)	7.48** (.55)	5.68** (.56)
N	203,741		5,382	
BMI (males only)	−6.19 (7.67)	−22.22 (15.23)	.56** (.23)	1.12** (.30)
N	203,378		5,372	
Underweight	−.09** (.004)	−.07** (.01)	−.07** (.02)	−.11** (.04)
N	203,378		5,372	
Overweight	.08** (.01)	.08** (.01)	.03 (.02)	.09** (.04)
N	203,378		5,372	
IQ (males only)	.91** (.03)	.58** (.04)	.48** (.14)	.62** (.18)
N	184,045		4,920	
High school completion	.16** (.01)	.04** (.01)	.07** (.02)	.09** (.04)
N	536,020		13,106	
Full-time work	.17** (.004)	.21** (.01)	.29** (.02)	.03 (.05)
N	368,582		10,388	
ln(earnings) FT	.09** (.01)	.08** (.01)	.09** (.03)	.12** (.06)
N	239,906		5,952	
ln(birth weight of first child)	.25** (.01)	.13** (.01)	.18** (.04)	.15** (.06)
N	63,842		1,862	

Standard errors are in parentheses. The control variables we use in the OLS estimation are year- and month-of-birth dummies, indicators for mother's education (one for each year), indicators for birth order, indicators for mother's year of birth, and an indicator for the sex of the child. Family fixed effects regressions include all of the above minus mother's education and mother's year of birth. Twin fixed effects regressions include indicators for sex and birth order of the twin (either first born or second born twin). Both cross-sectional and fixed effects regressions for height, BMI, and IQ also include indicator variables for the year the boy was tested by the military. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures—and is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986-1995 and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, the sample consists of women born between 1967 and 1988 whose first births occurred by 2004. If the first birth is a twin birth, the woman is dropped from the sample.

** Denotes statistically significant at the 5 percent level.

* Denotes statistically significant at the 10 percent level.

mortality by approximately twenty-eight deaths per 1,000 births. The twin fixed effects coefficient of −41 is statistically significant but only one sixth the size of the OLS coefficient. Similarly, when we look at five-minute APGAR scores as our outcome, we find a large OLS estimate but a much smaller twin fixed effects estimate. When we use linear measures of birth weight, our estimates are almost identical to the estimates of Almond et al. for the U.S., suggesting that the infant health production function may be similar in the U.S. and Norway.[18] For example, our twin fixed effects mortality estimate using birth weight is −10 (3) while theirs is −11 (.1).

IV.B. Height, BMI, and IQ at Age Eighteen–Twenty for Men

We next turn to male outcomes measured between ages eighteen and twenty.[19] Height is measured in centimeters so the OLS estimate suggests that a 10 percent increase in birth weight translates into about .75 extra centimeters of height at around age eighteen, and an increase in BMI of around .06. Twin fixed effects estimates are quite similar, with a 10% increase in birth weight leading to a .57 centimeter increase in height and a .11 increase in BMI. Our IQ measure is on a scale from one to nine; the estimated twin fixed effects coefficient of .62 suggests that an increase in birth weight by 10 percent will increase the score by .06 (about one twentieth of a stanine). For all three variables, fixed effects estimates are similar in magnitude to cross-sectional ones.

Given that BMI is an ambiguous health measure, as health may be adversely affected if BMI is too high (so men are overweight) or BMI is too low (so men are underweight), we have used the Center for Disease Control (CDC) cutoffs for overweight (BMI greater than or equal to 25—11 percent of the twins sample) and underweight (BMI less than 18.5—8 percent of the twins sample) to analyze the effect of birth weight on the probability of being in either of these two groups. The twin fixed effects estimates show

18. Because infant mortality is a rare outcome, estimated derivatives may be sensitive to functional form. When we assume other functional forms and estimate logit or probit equations instead of linear probability models, we get very different marginal effects (smaller by a factor of six) in the pooled estimation. Marginal effects from a fixed effects conditional logit model are also very different from the linear twin fixed effects estimates (not very surprising, given the selection problem induced by the fact that the logit only includes cases in which one twin lives and one twin dies).

19. To take account of the fact that men enter the military and take the test in different years and at different ages, we add dummies for the test year to the controls used earlier.

that increased birth weight significantly increases the probability of being overweight and significantly decreases the probability of being underweight.[20]

IV.C. High School Completion

We find that the within-twin estimates of the effect of birth weight on high school completion are similar in magnitude to the OLS estimates and statistically significant. The magnitude implies that an increase in birth weight of 10 percent increases the probability of high school completion by a bit less than 1 percentage point.[21]

IV.D. Labor Market Outcomes

In terms of labor market participation, the twin fixed effects estimates provide no evidence that increased birth weight increases the probability of working full time, despite relatively large OLS estimates (which suggest that a 10 percent increase in birth weight increases the probability of working full time by about .03). In contrast, both OLS and twin fixed effects estimates suggest that a 10 percent increase in birth weight raises full-time earnings by about 1 percent.[22] Given the return to education in Norway has been estimated to be about 4 percent for men [Black, Devereux, and Salvanes 2005b], this suggests that 10 percent more birth weight is about as valuable in the labor market as a quarter of a year of education.[23]

IV.E. Birth Weight of First Child

When we examine the subsample of female twins who both have children in our sample, we find that OLS and twin fixed effect estimates of the effects of birth weight on child's birth

20. Compared to Behrman and Rosenzweig [2004], we find smaller effects of birth weight on height and larger effects of birth weight on BMI. The estimates are not directly comparable, however, as theirs are for middle-aged women while ours are for young men.

21. Unlike with infant mortality, logit and probit marginal effects for high school graduation are very close to those from the linear probability model. However, fixed effects logit marginal effects are larger than the fixed effects linear probability model estimates.

22. Despite our finding of birth weight effects on education and earnings, controlling for birth weight has a negligible impact on the return to education estimated using twin difference models with our data. This is largely because birth weight explains very little of the education differences between twins (the within R^2 is .002 in the twins fixed effects regression of high school graduation on birth weight).

23. In contrast, Behrman and Rosenzweig [2004] find very large effects of fetal growth on female earnings (twin fixed effects estimates are about 6 times as large as their OLS estimates).

weight are quite similar, with a 10 percent increase in mother's birth weight leading to a 1.5 percent increase in the birth weight of their first child. This implies that our estimate is about twice that of Royer [2005]. Interestingly, we have found no evidence of an effect of birth weight on the weight of later-born children. Royer finds a small effect of birth weight using a sample of both first and second born children.

IV.F. Possible Mechanisms for Birth Weight Effects

There are many possible channels through which birth weight can affect adult outcomes. Plausible candidates include both biological and behavioral explanations. For example, nutrition in utero can affect brain development [Mogane et al 1993], which is consistent with our IQ findings. Also, size itself may matter for children's longer-run outcomes; certainly our height and BMI results suggest size advantages persist into adulthood for men.

Other explanations involve how parental and societal investments interact with birth weight. For example, if parents perceive that the return to investment is higher for the bigger twin, they may invest more in him/her and this may lead to better long-run outcomes. On the other hand, parents may engage in compensatory investment behavior that would attenuate birth weight effects.[24] While we cannot observe investments, one might expect that such behavior might differ depending on family resources. To examine this, we tried breaking our sample on a number of dimensions, including by mother's education (less than twelve years and twelve or more years), by family income, and by birth order of the children. Also, effects might differ based on local attitudes towards social support. As a proxy for this, we broke the sample by the voting behavior of the municipality of the mother (the proportion voting for the labor party in 1961). In none of these cases did we find statistically significant differences in the effects of birth weight.[25] However, in the absence of information

24. Recent work by Rosenzweig and Zhang [2006] provides some evidence of reinforcing parental investments using twins data from China. Also, Datar et al. [2005] find evidence that is generally suggestive of reinforcing parental investments among singletons using U.S. data.

25. Royer [2005] reports a similar finding from U.S. data. In other literature, there is some evidence that family resources affect the degree of differential investment; this appears in the cross-sectional and sibling fixed effects context, but not controlling for twin fixed effects. See Loughran et al. [2004] for one example.

on investments, we cannot come to any definitive conclusions about the importance of these mechanisms.

V. Robustness Checks

V.A. Singletons versus Twins

In Section VI, we address the question of the generalizability of our results. However, to facilitate comparison with other literature, we have included both cross-section and mother fixed effects estimates for singletons in Table III.[26] The singleton fixed effects provide an interesting contrast to twin fixed effects as they are robust to any factors that are mother-specific and unchanging but are not robust to any omitted factors that are correlated with particular pregnancies. The fixed effects estimates provide further support for birth weight having important long-run effects. However, they differ from the twin fixed effects models in that birth weight is also seen to have large short-run impacts.[27]

V.B. Sample Consistency

In Table III, we used all available observations for each outcome to maximize precision. However, a key feature of our paper is that we can study both short- and long-run outcomes for the same cohorts of individuals. Table IV presents these results. The first set of results (Columns 1 and 2) are for male same-sex twins born between 1978 and 1986. For this group, we can use exactly the same twin pairs to study APGAR, height, BMI, and IQ. We also include infant mortality for these cohorts even though it obviously cannot be restricted to exactly the same twin pairs. The second set of results is for same-sex female twins born between 1967 and 1977. For these cohorts we present estimates for infant mortality, high school graduation, whether they work full-time, full-time earnings, and birth weight of first child.[28] Because of the selection problems resulting from infant mortality, work

26. We omit cases where there are fewer than two children in the family. However, the OLS results are essentially unchanged if we use the sample of all singletons.

27. Our family fixed effects estimates are quite consistent with estimates from North America. For example, Currie and Moretti [2005] estimate the intergenerational correlation in ln(birth weight) in California to be about .17 (.004) while we get an estimate of .13 (.01). Similar to our own findings, Oreopoulos et al. [2006] use Canadian data and find that the sibling fixed effects estimates for infant mortality are more negative than the corresponding OLS estimates.

28. Because of a variety of age restrictions and data availability, we selected these samples to maximize our ability to compare outcomes over time while maintaining the same sample of individuals.

TABLE IV
REGRESSION RESULTS CONSTANT SAMPLE: COEFFICIENT ON LN (BIRTH WEIGHT)

Dependent variable	Male same-sex twins 1978–1986 OLS	Male same-sex twins 1978–1986 FE	Female same-sex twins 1967–1977 OLS	Female same-sex twins 1967–1977 FE
1-year mortality	−299.24** (31.74)	−33.20 (22.92)	−390.56** (27.62)	5.84 (23.57)
N	2760		3804	
Five-minute APGAR score	.90** (.15)	.37* (.21)	—	—
Height (males only)	7.55** (.89)	8.56** (.85)	—	—
BMI (males only)	.20 (.40)	2.24** (.54)	—	—
Underweight	−.06 (.03)	−.05 (.05)	—	—
Overweight	.03 (.04)	.17** (.07)	—	—
IQ (males only)	.29 (.20)	1.09** (.28)	—	—
N	1894			
High school completion	—	—	.03 (.04)	.11* (.06)
N			3466	
Full-time work	—	—	.19** (.03)	−.03 (.08)
N			3574	
ln(earnings) FT	—	—	.17** (.07)	.14 (.10)
N			1732	
ln(birth weight of first child)			.19** (.03)	.18** (.06)
N			1722	

Standard errors are given in parentheses. The control variables we use in the OLS estimation are year- and month-of-birth dummies, indicators for mother's education (one for each year), indicators for birth order, indicators for mother's year of birth, and an indicator for the sex of the child. Twin fixed effects regressions include indicators for sex and birth order of the twin (either first born or second born twin). Both cross-sectional and fixed effects regressions for height, BMI, and IQ also include indicator variables for the year the boy was tested by the military. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures—and is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995, and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, the sample consists of women born between 1967 and 1977 whose first births occurred by 2004. If the first birth is a twin birth, the woman is discarded from the sample.

** Denotes statistically significant at the 5 percent level.

* Denotes statistically significant at the 10 percent level.

decisions, and fertility decisions, we do not attempt to study a common set of twin pairs. Instead, the estimates are all for the same cohorts. For both cuts of the data, our results are similar to before: while the fixed effects estimates for short-run outcomes are smaller than OLS, the equivalent estimates for long-run outcomes are generally about the same magnitude or larger than OLS.[29]

V.C. The Role of Zygosity

Because our twin sample includes both fraternal and monozygotic twins, estimates could in part reflect genetic differences between twins. To investigate this issue, we first restrict our sample to same-sex twin pairs. While this sample is not limited to monozygotic twins, by eliminating opposite-sex twin pairs (which are clearly not monozygotic), the sample now contains a larger fraction of identical twin births. Table V reports fixed effects estimates for all twins (from Table III) and all same-sex twin pairs for comparison. The estimates are very similar in both samples, suggesting no large differences in estimates by zygosity.[30]

While we don't observe zygosity for all twins in our sample, we do observe it for a subset of the twins born between 1967–1979 who completed the twins questionnaire described in Section III. We can thereby see how our results differ when we isolate monozygotic twins from all same-sex twins. These results are in further columns of Table V. Because the twins who complete the questionnaire are a selected sample, we present results for (1) all same-sex twin pairs in the 1967–1979 cohorts, (2) all same-sex twin pairs who complete the survey, and (3) all monozygotic twin pairs known from the survey.[31] It is clear from the last two columns of Table V that estimates for monozygotic twins are almost identical to those for all same-sex twins who complete the survey, suggesting that genetic factors are not confounding our earlier estimates. It is also interesting to note that our results are somewhat different from the results when we use our full administrative sample (that does not rely on any information being obtained from the individual), suggesting that there is selection as to who chooses to complete these twin surveys.[32] Given that

29. Our high school graduation estimate for women is a little larger than, but not statistically different from, the equivalent estimate from Royer [2005] of .05 (.05) for childbearing women. Also, we find that this estimate falls and becomes very close to Royer's if we restrict our sample to women who give birth in our sample period.

30. There are no statistically significant differences between estimates for same-sex twins and mixed-sex twins. We also tried breaking the sample by gender and did not find significant differences between men and women.

31. We are unable to look at 1-year mortality because questionnaires were mailed only to twin pairs that were intact at age 3, and we exclude APGAR as it does not become available until 1977.

32. Comparing estimates for all same-sex twins, and same-sex twins for the 1967–1979 period, it also appears that the effects of birth weight on height, BMI, and IQ get larger over the sample period. We examine this issue more thoroughly in Table VI.

TABLE V
FIXED EFFECTS RESULTS FOR LN (BIRTH WEIGHT)

	All twins	All same-sex twins	All same-sex twins 1967–1979	Same-sex twins in survey	Monozygotic twins
One-year mortality	−41.10** (7.64) [33,366]	−40.15** (9.44) [22,812]	14.03 (16.48) [9,120]	—	—
Five-minute APGAR score	.35** (.07) [21,580]	.38** (.08) [14,684]			
Height (males only)	—	5.68** (.56) [5,382]	4.20** (.71) [3,558]	3.79** (.83) [2,700]	4.14** (.64) [1,376]
BMI (males only)	—	1.12** (.30) [5,372]	.59* (.35) [3,552]	.37 (.39) [2,696]	.43 (.37) [1,372]
IQ (males only)	—	.62** (.18) [4,920]	.37 (.23) [3,332]	.20 (.27) [2,538]	.22 (.30) [1,312]
High school completion	.09** (.04) [13,122]	.10** (.04) [9,002]	.10** (.04) [8,186]	.08* (.05) [6,638]	.09* (.06) [3,468]
Full-time work	.03 (.05) [10,388]	−.02 (.06) [7,124]	−.02 (.06) [7,124]	.01 (.07) [5,434]	.01 (.09) [2,794]
ln(earnings) FT	.12** (.06) [5,952]	.13** (.06) [4,098]	.13** (.06) [4,098]	.08 (.06) [3,614]	.09 (.09) [1,904]
ln(birth weight of first child)	—	.15** (.06) [1,862]	.14** (.07) [1,812]	.12* (.07) [1,636]	.18** (.09) [858]

Standard errors are given in parentheses. The control variables we use include indicators for sex and birth order of the twin (either first born or second born twin). Regressions for height, BMI, and IQ also include indicator variables for the year the boy was tested by the military. The twins survey is available for cohorts born between 1967–1979. Twin pairs that experienced infant mortality are not in the survey universe. APGAR scores are only available from 1977 so estimates from the 1967–1979 period are not reported. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures—and is reported in stanine (Standard Nine) units. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. We restrict attention to individuals aged at least twenty-five. Working full-time indicates whether individuals are full-time, full-year workers. To identify this group, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995, and in the fourth quarter thereafter). We label these individuals as full-time workers. For ln(birth weight) of child, unless otherwise specified, the sample consists of women born between 1967 and 1988 whose first births occurred by 2004. If the first birth is a twin birth, the woman is discarded from the sample. The outcome variable is the birth weight of the first born child.

** Denotes statistically significant at the 5 percent level.

* Denotes statistically significant at the 10 percent level.

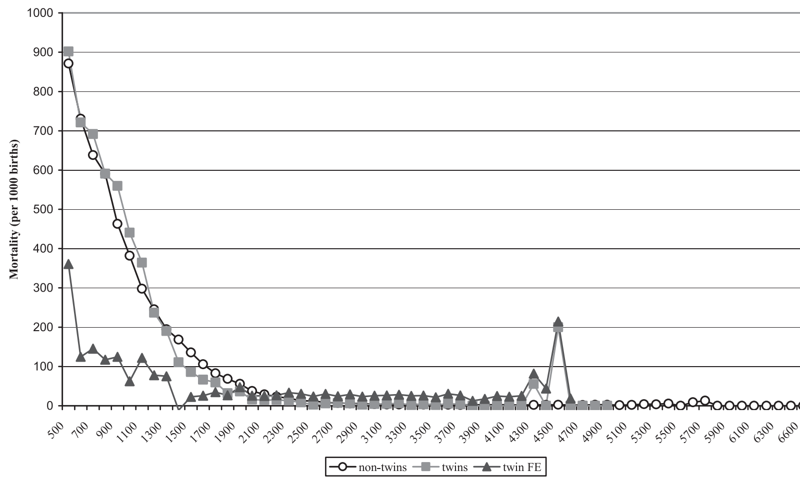

FIGURE II
Mortality Rate by Birth Weight

The calculations for the non-twins and the twins samples are simply the average infant mortality per 1,000 births in that birth weight cell. The calculations for twin fixed effects (FE) are the average mortality rate for the cell after controlling for twin fixed effects. The sample is based on all Norwegian individuals born between 1967 and 1997.

the results for monozygotic twins are so similar to those for all same-sex twins, we will continue to stress the results using the twins samples from the administrative data.

V.D. Heterogeneous Effects across the Birth Weight Distribution

While using the natural log of birth weight does allow for nonlinear effects, it is possible to allow the effects of birth weight to be more flexible. Figures II–V do this graphically. Figure II illustrates the differences between the OLS estimates for mortality and those with the twin fixed effects across the birth weight distribution by presenting the average one-year mortality rate (per thousand births) by birth weight, both with and without twin fixed effects. It is clear that not only are the twin fixed effects estimates much smaller than the OLS, but there is also evidence of significant nonlinearities, with increased body mass having a negative effect on mortality at low birth weights but little discernable effect at weights above 2,000 grams (in fact, the twin mortality rate falls from sixteen per thousand to one per thousand as one moves from 2,000 to 3,500 grams). This is also true of the 5 minute APGAR score (see Black, Devereux and Salvanes [2005c]).

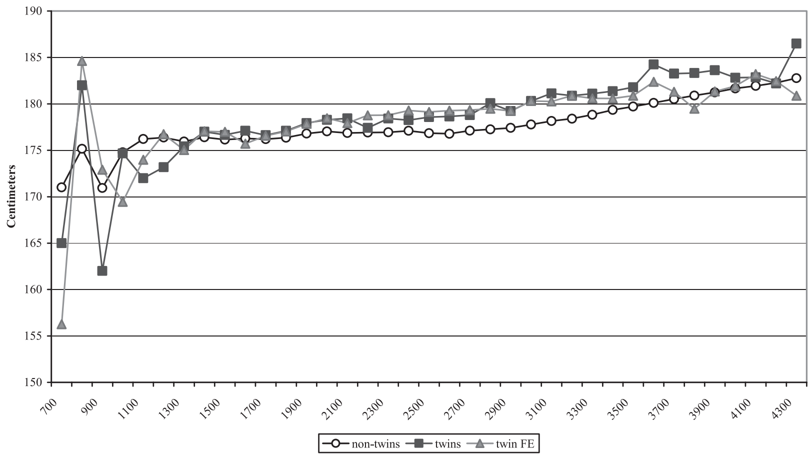

FIGURE III
Height by Birth Weight (Males Only)

The calculations for the non-twins and the twins samples are simply the average male height measured between ages eighteen and twenty in that birth weight cell. The calculations for twin fixed effects (FE) are the average height for the cell after controlling for twin fixed effects. The sample is based on all Norwegian males who registered for mandatory military service in Norway and who were born between 1967 and 1987.

Unlike the case with mortality, Figure III shows that OLS and twin fixed effects estimates for height are very similar to each other. Once again, there is some evidence of a non-linear relationship, with the positive relationship between birth weight and height flattening out after about 1,500 grams. The equivalent figures for IQ (Figure IV), and full-time earnings (Figure V) show once again that OLS and fixed effects estimates are very similar across the distribution and provide little evidence of strong non-linearities.[33] In all figures, the estimates are noisy at very low and very high birth weights, reflecting the paucity of data in these regions.

V.E. Selection into the Later Outcomes Sample

When looking at the effect of birth weight on later outcomes, we are inherently including only those individuals for whom we

33. In the working paper version, we also allowed for splines in birth weight with less than 1,500, 1,500–2,500, and 2,500 or more as the cutoffs. We found substantial non-linearities in mortality and the five minute APGAR score, with a large marginal benefit for additional grams among very low birth weight babies in terms of both these outcomes. However, as is suggested by Figures II–V, we found little evidence of significant nonlinearities in later outcomes (see Black, Devereux and Salvanes [2005c]).

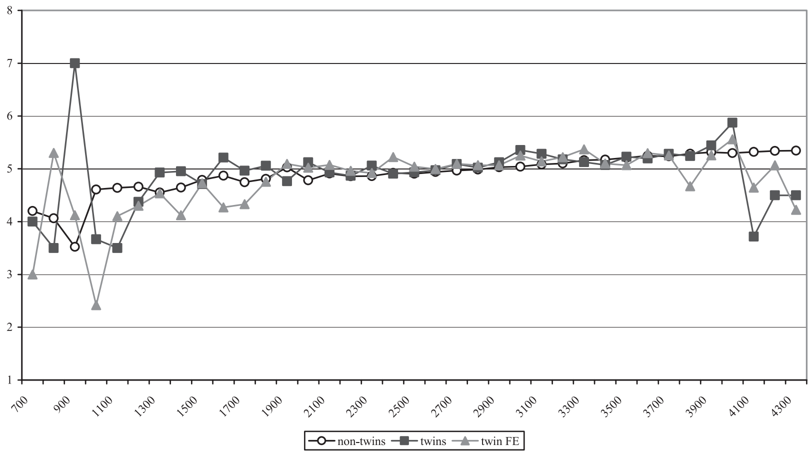

FIGURE IV
IQ by Birth Weight (Males Only)

The calculations for the non-twins and the twins samples are simply the average male IQ measured between ages eighteen and twenty in that birth weight cell. The calculations for twin fixed effects (FE) are the average IQ for the cell after controlling for twin fixed effects. The sample is based on all Norwegian males who registered for mandatory military service in Norway and who were born between 1967 and 1987. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures (see Sundet et al. [2004 2005] and Thrane [1977] for details). The arithmetic test is quite similar to the Wechsler Adult Intelligence Scale (WAIS) [Sundet et al. 2005, Cronbach 1964]. The word test is similar to the vocabulary test in WAIS and the figures test is similar to the Raven Progressive Matrix test [Cronbach 1964]. The composite IQ test score is an unweighted mean of the three subtests. The IQ score is reported in stanine (Standard Nine) units.

observe later outcomes. In particular, individuals who did not survive are not included in our sample and this may bias our estimates. Unlike previous twin studies of this nature, we observe birth characteristics (such as birth weight) of twin pairs who are subsequently impacted by infant mortality and can examine the characteristics associated with selection into the sample.

Table VI demonstrates that the twin fixed effects estimates of the impact of birth weight on later outcomes have tended to increase over time (we have omitted outcomes for which there were very few observations outside the first time period). Over this same time period, infant mortality amongst twins has declined, from about sixty-six per thousand births in 1967 to less than thirteen per thousand in 1997. While there are many possible reasons for the temporal pattern in the estimates, one possibility is that later effects are larger because the sample includes more twins who were on the margin of survival in infancy.

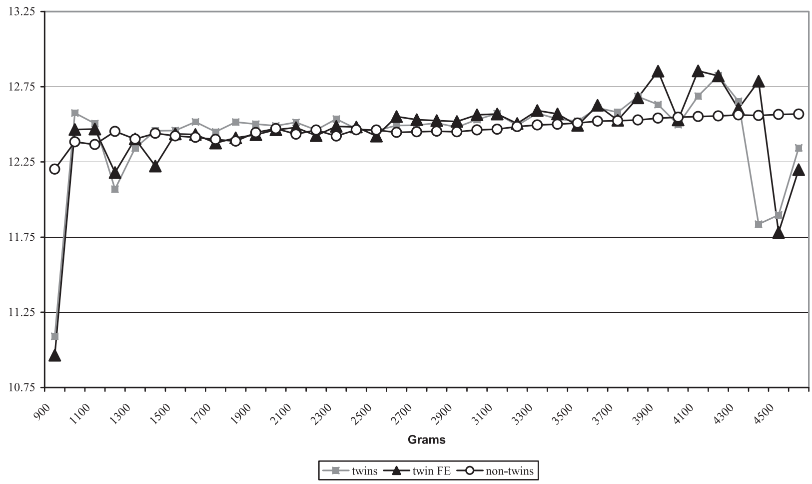

FIGURE V

ln(Full-Time Earnings) By Birth Weight

The calculations for the non-twins and the twins samples are simply the average ln(earnings) for full-time workers aged at least twenty-five in that birth weight cell. The calculations for twin fixed effects (FE) are the average ln(earnings) for full-time workers for the cell after controlling for twin fixed effects. Earnings are measured as total pension-qualifying earnings reported in the tax registry. These are not topcoded and include labor earnings, taxable sick benefits, unemployment benefits, parental leave payments, and pensions. To identify full-time workers, we use the fact that our dataset identifies individuals who are employed and working full time (30+ hours per week) at one particular point in the year (in the second quarter in the years 1986–1995, and in the fourth quarter thereafter). For the twin sample, both twins must be working full time in a given year to be included in our data. We use the most recent year of earnings available. Because of the age restriction, the cohorts included are those born between 1967 and 1977.

Though it is inherently impossible to know what the effects of birth weight would have been on the later outcomes of the individuals we do not observe, we do try to think about how this selection may be biasing our results. If there are heterogeneous effects of birth weight across twin pairs and birth weight is actually more important for twin pairs who subsequently experience mortality, we may be underestimating the effect of birth weight on later outcomes. Because we observe the five-minute APGAR score for all individuals (even those who subsequently die in infancy) beginning in 1977, we can test this theory by separately estimating the relationship between birth weight and APGAR for the full sample and the sample of twin pairs where both twins live. When we do this using twin fixed effects, we find that

TABLE VI
DIFFERENCES BY BIRTH COHORTS: RESULTS FOR LN (BIRTH WEIGHT) OLS AND TWIN FIXED EFFECTS

	1967–1976		1977–1986		1987–1997	
Year of birth	OLS	FE	OLS	FE	OLS	FE
One-year mortality	−427.00** (15.44)	7.11 (16.20)	−279.02** (18.75)	−45.17** (14.11)	−173.89** (12.91)	−78.26** (10.03)
	[10604]		[8618]		[14144]	
Five-minute APGAR score	—	—	1.84** (.11)	.62** (.12)	1.29** (.07)	.20** (.08)
			[7616]		[13964]	
Height (males only)	7.52** (.76)	3.38** (.79)	7.80** (.81)	8.24** (.78)	—	—
	[2900]		[2400]			
BMI (males only)	.82** (.29)	.37 (.38)	.22 (.37)	2.01** (.48)	—	—
	[2898]		[2392]			
IQ (males only)	.63** (.21)	.32 (.25)	.33* (.19)	1.02** (.27)	—	—
	[2700]		[2150]			
High school completion	.05** (.02)	.04 (.04)	.15** (.04)	.22** (.07)	—	—
	[9500]		[3622]			

Standard errors are given in parentheses. Samples sizes are given in brackets. The control variables we use include indicators for sex and birth order of the twin (either first born or second born twin). Regressions for height, BMI, and IQ also include indicator variables for the year the boy was tested by the military. High school completion indicates whether or not the individual has completed at least twelve years of schooling and is restricted to those twenty-one and older. The IQ measure is generated from a composite score from three speeded IQ tests—arithmetic, word similarities, and figures—and is reported in stanine (Standard Nine) units.

** Denotes statistically significant at the 5 percent level.

* Denotes statistically significant at the 10 percent level.

log birth weight has a significantly larger positive effect on the APGAR score for the full sample of twin births. The difference is large—.35 (.07) for the full sample versus .19 (.06) for the sample without mortality. If this relationship is also true of other, later outcomes, then we may be underestimating the true effect of birth weight on later outcomes by a substantial amount.[34]

VI. External Validity

While using within-twin variation allows us to credibly identify the causal effect of differences in birth weight arising from differences in access to nutrition in utero, the issue of generalizability of these results to the general population of births remains.

From Table I, we can see that there are substantial differences between twin and singleton births. Gestation is longer for singletons, with the average at 39.8 weeks versus 36.9 for twins. Five-minute APGAR scores are also higher for singletons, there is a lower fraction with complications, and the one-year mortality rate is only six per 1,000 births as opposed to thirty-one for twins. Parental education is similar for both groups but the mothers of twins tend to be older.

One of the most notable differences is that twins come disproportionately from the lower part of the birth weight distribution. Only 3 percent of singletons are classified as low birth weight (less than 2,500 grams), while 39 percent of twins are low birth weight. In addition, there are few large birth weight twins; the fifth percentile singleton is approximately the same weight as the median twin (2,640 versus 2,660 grams) and the median singleton is approximately the same weight as the 95th percentile twin (3,540 vs 3,500 grams). However, when one compares twins and singletons with the same birth weight, their outcomes are

34. A more formal approach to the missing data problem is to model the probability that a twin pair will experience mortality within the first year and hence attrit from the later outcomes sample. We have tried allowing the probability of attrition to depend on flexible functions of the birth weight of each twin as well as the gestation length of the twin pair and used these estimates to form weights equal to the inverse of the probability of not attriting due to mortality in the first year. When we do this reweighting, we again find that our estimates are likely underestimating the true effect of birth weight on later outcomes, although the differences between the weighted and unweighted estimates are not large. Also, if we carry out this exercise allowing for attrition other than infant mortality, we find similar estimates.

surprisingly similar. In Figures II–V, we have graphed the relationship between birth weight and mortality, height, IQ, and earnings for the samples of twins and singletons. The twins and nontwins actually have quite similar outcomes conditional on birth weight, suggesting that our results may be generalizable to the rest of the population.[35] This conclusion is bolstered by the earlier finding in Table 3 that sibling fixed effects estimates for later outcomes are generally quite similar to our twin fixed effects estimates.

However, generalizability should still be viewed with caution, as different sources of variation in birth weight may have different effects on outcomes. Also, we cannot rule out the possibility that twins and singletons have very different causal relationships between birth weight and outcomes but that they are subject to different confounding factors that happen to cancel each other out so that the cross-sectional profiles are similar.

VII. Conclusions

In this paper, we have examined the effect of birth weight on adult outcomes using within-twin variation in birth weight to control for other, often unobservable, parental and environmental factors. Consistent with the recent literature, we find that OLS estimates for infant mortality and APGAR are much larger than those from twin fixed effects. However, we find significant effects of birth weight on adult outcomes, including height, BMI, IQ, education, earnings, and birth weight of the first-born child. Twin fixed effects estimates for these adult outcomes are similar in size to OLS estimates.

It is not clear why twin fixed effects estimates are so much smaller than OLS estimates for short-run but not for adult outcomes. It may be that some omitted variables that are correlated with short-run outcomes may be less correlated with long-run

35. This is consistent with findings in the medical literature that suggest that the primary cause of disparities in outcomes between twins and singletons is due to differences in size at birth. Allen [1995] notes that, in a sample of preterm births, no differences were present between twins and singletons with respect to neurodevelopmental outcomes at eighteen months from due date, after adjusting for confounding social, obstetric, and neonatal factors (including birth weight). Differences were only found when they examined preterm infants with birth weights of <800 grams, suggesting greater vulnerability of twins born at the limit of viability. See also Hoffman and Bennett [1990].

outcomes such as earnings, and so OLS results are more biased for short-run outcomes than for longer-run outcomes. For example, maternal smoking may have a significant effect on short-run health outcomes but may have little or no effect on longer-run outcomes. Another possibility is that parental investments favor the heavier twin. If it is the case that the returns to parental investment are higher for heavier twins, this would tend to increase the twins fixed effects estimates for adult outcomes relative to the pooled OLS. As discussed in Section III, our limited tests provide little evidence of this type of behavior but they are far from conclusive. While birth weight clearly affects longer run outcomes, further research is required to determine the mechanisms underlying it.

To get a sense of the magnitude of our estimates, we consider the WIC program in the United States. Earlier work by Kowaleski-Jones and Duncan [2002] estimated the effect of WIC participation by a pregnant woman to be about a 7.5 percent increase in child birth weight.[36] Using this estimate, we can translate this increased birth weight into the effect of WIC on longer run outcomes. Based on our estimates, a 7.5 percent increase in birth weight would lead to a little less than half a centimetre increase in height, a .05 stanine increase in IQ, a 1 percent increase in full-time earnings, and a 1.1 percent increase in the birth weight of their children.

An important caveat to this quantification exercise is that we are identifying off of variation related to access to nutrition in utero. Other factors affecting birth weight, such as maternal behavior (smoking, etc) and gestation length, may have different effects on children's outcomes. However, the evidence does seem to suggest that, by looking exclusively at the effect of birth weight on short run outcomes, one may miss out on sizeable effects of birth weight that manifest themselves in the longer run.

Department of Economics, University of California, Los Angeles,
National Bureau of Economic Research, and IZA
School of Economics and Geary Institute, University College Dublin,
CEPR, and IZA
Department of Economics, Norwegian School of Economics,
Statistics Norway, Center for the Economics of Education, and IZA

36. They use the National Longitudinal Survey of Youth and apply a sibling fixed effects approach, identifying off of mothers who participated in WIC during one pregnancy but not during the other one.

REFERENCES

Allen, Michael. C, "Factors Affecting Developmental Outcome" *in Multiple Pregnancy: Epidemiology, Gestation & Perinatal Outcome*, Louis Keith, Emile Papiernik, Donald Keith, and Barbara Luke, eds. (New York, NY: Parthenon, 1995).

Almond, Douglas, Kenneth Y. Chay, and David S. Lee, "The Costs of Low Birth Weight," *Quarterly Journal of Economics*, CXX (2005), 1031–1083.

Barker, David J, "Fetal Origin of Coronary Hearth Disease," *British Medical Journal*, CCCXVII (1995), 171–174.

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

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

——, "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Education," *American Economic Review*, VC (2005b), 437–449.

——, "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," NBER Working Paper No. 11796, 2005c.

Blickstein, Isaac, and Robin B. Kalish, "Birthweight Discordance in Multiple Pregnancy," *Twin Research*, VI (2003), 526–531.

Bryan, Elizabeth, *Twins and Higher Order Births: A Guide to their Nature and Nurture*, (London, UK: Edward Arnold, 1992).

Case, Anne, Angela Fertig, and Christina Paxson, "The Lasting Impact of Childhood Health and Circumstance," Princeton Center for Health and Wellbeing Working Paper, April 2004.

Christensen, Kaare, JamesW. Vaupel, Niels V. Holm, and Anatoli I. Yashin, "Mortality Among Twins After Age 6: Fetal Origins Hypothesis Versus Twin Method," *British Medical Journal*, CCCX (1995), 432–36.

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

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

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

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.

——, "Biology as Destiny? Short and Long-Run Determinants of Inter-Generational Transmission of Birth Weight," NBER Working Paper No. 11567, 2005.

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 Socio-Economic Status? The Case of Birth Weight," *American Economic Review*, LXXXIX (1999), 245–250.

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

Duffy, David L, "Twin Studies in Medical Research," *Lancet,* CCCXLI (1993), 1418–1419.

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

Grjibovski, A. M., J. H. Harris, and P. Magnus, "Birth Weight and Adult Health in a Population-Based Sample of Norwegian Twins," *Twin Results and Human Genetics*, VIII (2005), 148–155.

Hack, Maureen, Daniel J. Flannery, Mark Schluchter, Lydia Cartar, Elaine Borawski, and Nancy Klein, "Outcomes in Young Adulthood for Very-Low-Birth Weight Infants," *New England Journal of Medicine*, CCXLVI (2002), 149–157.

Hack, Maureen, H., Gerry Taylor, Nancy Klein, Robert Eiben, Christopher

Schatschneider, and Nori Mercuri-Minich, "School-Age Outcomes in Children with Birth Weights Under 750 Grams," *New England Journal of Medicine*, CCCXXXI (1994), 753–759.

Harris, Jennifer R, Per Magnus, and Kristian Tambs, "The Norwegian Institute of Public Health Twin Panel: A Description of the Sample and Program of Research," *Twin Research*, V (2005), 415–423.

Hoffman, E. L., and F. C. Bennett, "Birth Weight Less than 800 Grams: Changing Outcomes and Influences of Gender and Gestation Number," *Pediatrics*, LXXXVI (1990), 412–414.

Keith, Louis G, Emile P, Donald M. Keith, and Barbara Luke, eds. *Multiple Pregnancy: Epidemiology, Gestation, & Perinatal Outcome*, (New York, NY: Parthenon, 1995).

Kowaleski-Jones, Lori, and Greg J. Duncan, "Effects of Participation in the WIC Program on Birthweight: Evidence from the National Longitudinal Survey of Youth," *Journal of Public Health*, XLII (2002), 799–804.

Little, Roderick J. A, and Donald B. Rubin, *Statistical Analysis with Missing Data*, (Hoboken, NJ: Wiley, 1987).

Loughran, David S., Ashlesha Datar, and M. Rebecca Kilburn, "The Interactive Effect of Birth Weight and Parental Investment on Child Test Scores," RAND Labor and Population Working Paper No. WR-168, 2004.

Møen, Jarle., Kjell G. Salvanes, and Erik Ø. Sørensen, "Documentation of the Linked Empoyer-Employee Data Base at the Norwegian School of Economics," Mimeo, The Norwegian School of Economics and Business Administration, 2003.

Morgane, Peter J., Robert Austin-LaFrance, Joseph Bronzino, John Tonkiss, Sofia Diaz-Cintra, L. Cintra, Tom Kemper, and Janina R. Galler, "Prenatal Malnutrition and Development of the Brain," *Neuroscience and Biobehavioral Reviews*, XVII (1993), 81–128.

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

Phillips, David I. W, "Twin Studies in Medical Research: Can They Tell Us whether Diseases are Genetically Determined?" *Lancet,* CCCXXXIV (1993), 1008–9.

Rosenzweig, Mark R, and Junsen Zhang, "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One Child' Policy," IZA Discussion Paper No. 2082, April 2006.

Royer, Heather, Separated at Girth: Estimating the Long-Run and Intergenerational Effects of Birthweight Using Twins, Unpublished manuscript, 2005.

Strauss, John, and Duncan Thomas, "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, XXXVI (1998), 766–817.

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

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

Thrane, Vidkunn Coucheron, "Evneprøving av Utskrivingspliktige i Norge 1950–53," Arbeidsrapport nr. 26, INAS 1977.