The Effect of the WIC Program on the Health of Newborns

E. Michael Foster, Miao Jiang, and Christina M. Gibson-Davis

Objective. To determine the effect of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) on birth outcomes.

Data Source. The Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID). The PSID provides extensive data on the income and well-being of a representative sample of U.S. families from 1968 to present. The CDS collects information on the children in PSID families ranging from cognitive, behavioral, and health status to their family and neighborhood environment. The first two waves of the CDS were conducted in 1997 and 2002, respectively. We use information on 3,181 children and their mothers.

Study Design. We use propensity score matching with multiple imputations to examine whether WIC program influences birth outcomes: birth weight, prematurity, maternal report of the infant's health, small for gestational age, and placement in the neonatal intensive care unit. Furthermore, we use a fixed-effects model to examine the above outcomes controlling for mother-specific unobservables.

Principal Findings. After using propensity scores to adjust for confounding factors, WIC shows no statistically significant effects for any of six outcomes. Fixed-effects models, however, reveal some effects that are statistically significant and fairly substantial in size. These involve preterm birth and birth weight.

Conclusions. Overall, the WIC program had moderate effects, but findings were sensitive to the estimation method used.

Key Words. WIC, birth outcomes, propensity score, multiple imputations, fixed-effects model

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) is the third largest food program in the United States, reaching nearly 8 million women and children at a cost of U.S.\$5 billion (USDA 2007a). The program forms a key part of the safety net for poor families: nearly 60 percent of poor children under the age of 4 receive WIC, nearly twice as many as those who receive food stamps (Zedlewski and Rader 2005).

Given the program's scope and prominence, the efficacy of WIC has been studied extensively. However, findings have been inconsistent, and the

program remains an active focus of policy evaluation (Ludwig and Miller 2005). The methodological challenge facing researchers is that many characteristics are likely correlated with both WIC use and child health. This correlation introduces bias, leading one to either overstate or understate WIC's effects depending on the nature of the relationships between the measured and unmeasured confounding factors, children's outcomes and WIC participation.

In this paper, we use propensity scores to examine the association between WIC participation and birth outcomes. *Under key assumptions*, propensity scores approximate a randomized experiment by creating matched groups comprising those from both the treatment and comparison groups who are comparable except for treatment status. We further estimate fixed-effects models that allow for mother-specific unobservables. In addition, we examine how strong confounding with an unobserved variable would have to be to explain the apparent effect (or noneffect) of WIC participation using approaches proposed by Rosenbaum (2002) and Imbens (2003). These analyses involve data from the Panel Study of Income Dynamics—Child Development Supplement (PSID-CDS), a study that has tracked more than 5,000 families for nearly four decades. The data included an over-sample of poor families, making it well suited for examining issues of poverty and child development in the United States.

This study makes two contributions to the existing literature. First, by considering multiple outcomes—birth weight, born preterm, low birth weight, small for gestational age (SGA), neonatal hospitalizations, and maternal report of infant health—this study provides a more complete account of WIC's potential effects than most prior research. Second, because of the scope of the data, we account for the effect of heretofore unmeasured or omitted characteristics (such as maternal IQ and family income) that are likely to confound WIC estimates.

BACKGROUND

The WIC program promotes the development of fetuses, infants, and young children by offering supplemental food packages, nutritional education, and

Address correspondence to E. Michael Foster, Ph.D., Professor, Gillings School of Global Public Health, University of North Carolina, Rosenau Hall, Campus Box# 7445, Chapel Hill, NC 27599-7445; e-mail: emfoster@unc.edu. Miao Jiang, M.S., is with the Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC. Christina M. Gibson-Davis, Assistant Professor, is with the Terry Sanford Institute of Public Policy, Duke University, Durham, NC.

referrals to health care and other social service providers. Since its inception in the early 1970s, the program has grown tremendously in size and scope. In 1974, the program had 88,000 participants; by 2004, the program had nearly 8 million participants, served by 2,200 agencies throughout the country (USDA 2006). Half of all children <1 year and one-quarter of children between ages 1 and 4 years participate (Oliveira et al. 2002). One in every U.S.\$10 spent on food assistance in the United States is devoted to WIC (USDA 2007a).

The WIC program offers three benefits to eligible participants. The first is a food package, containing foods that are high in certain nutrients, such as protein, iron, calcium, vitamin A, and vitamin C (USDA 2007b). Infants up to a year also can receive free and reduced-cost formula. In 2006, the value of the average monthly food package per person was U.S.\$39.03, but it was higher for families receiving free and reduced cost formula (Food and Nutrition Service 2007). The second benefit is nutritional counseling, which consists of two mandatory education classes during each 6-month certification period. The final benefit is referrals to health care and other social services.

To qualify for WIC, recipients must meet three requirements. First, they must be a member of a categorically eligible group: pregnant women, postpartum women (eligible for up to 1 year if breastfeeding, 6 months if not breastfeeding), and children up to 5 years old. Second, they must be income eligible, either because their incomes are at or below 185 percent of the federal poverty line or because they participate in the Temporary Aid to Needy Families, Medicaid, or Food Stamp programs. Third, a health care professional must declare them at "nutritional risk," either because they have a medical-based risk (e.g., anemia, underweight, and poor pregnancy outcomes) or because they have a diet-based risk (e.g., an inadequate diet). In practice, nearly all mothers and children meet the nutritional risk criterion (Currie 2003).

WIC promotes healthy birth outcomes through the provision of nutritious foods that aid in fetal development and/or the promotion of advantageous maternal behaviors through its counseling sessions and referral to other social agencies (Rossi 1998). However, various factors may block these pathways. First, while situations of extreme nutritional deprivation have strong negative effects on birth outcomes (Stein and Susser 1975), the effects of moderate nutritional deficits are much less pronounced (Goldenberg and Rouse 1998; Iams 1998). Nutrition is linked to poor birth outcomes insofar as the mother has a low prepregnancy body mass index or has inadequate gestational weight gain (Kramer 2003). WIC recipients, though, are more likely to

be obese before pregnancy and do not gain substantially more weight during pregnancy than do non-WIC recipients (Bitler and Currie 2005).

Second, the counseling and referral components of WIC are limited, and evidence as to whether they are effective in changing behaviors is inconsistent. WIC offers two voluntary half-hour counseling and educational sessions per 6-month enrollment period, during which WIC workers are required to provide information on nutrition, drug, and alcohol services, and other available social programs (GAO 2004). Some studies have found that these sessions increase prenatal care and decrease smoking and tobacco use, while other work has found either no association or a negative one (Rush et al. 1988; Kahler et al. 1992; Bitler and Currie 2005; Joyce, Gibson, and Colman 2005). The one maternal behavior where the evidence appears unequivocal is breastfeeding, as research has repeatedly demonstrated that WIC reduces breastfeeding (Chatterji and Brooks-Gunn 2004; Jacknowitz, Novillo, and Tiehen 2007), presumably because of its provision of free formula.

Despite the mixed evidence supporting the pathways between WIC and birth outcomes, numerous studies have found that prenatal WIC participation is associated with increases in mean birth weights (Lazariu-Bauer et al. 2004; Bitler and Currie 2005), decreases in the number of low and very low birth weight babies, fewer preterm deliveries, and reductions in neonatal and fetal mortality rates (Kennedy and Kotelchuck 1984; Stockbauer 1987; Rush et al. 1988; Devaney 1992; Brown, Watkins, and Hiatt 1996; Moss and Carver 1998; May et al. 2001). The magnitude of the program's impact differs depending on the degree to which studies address the problem of selection bias, suggesting that endogeneity is a key problem in WIC evaluation (Ludwig and Miller 2005). For example, using data from Pregnancy Nutrition Surveillance System, Joyce et al. (2005) adjusting for the timing of WIC enrollment eliminated the association between WIC and preterm birth. The authors reasoned that women with longer pregnancies had more opportunity to enroll in WIC and hence estimates of WIC on preterm birth were confounded by gestational age bias. In another study, Joyce et al. (2005) use data on births in New York city and find that the advantage of WIC to twins, who are at particular risk of fetal growth restriction, to be minimal.

The literature also demonstrates the advantages and disadvantages of alternative data sources. Administrative data generally have more reliable information on program participation than self-report data, but they have limited information on the range of outcomes and the potential confounders. Lazariu Bauer and colleagues, for example, examine a large administrative database on WIC participation in New York state. The authors do consider 29

potential confounders—these variables are drawn from birth certificate data and program information. However, key characteristics are unavailable, such as mother's cognitive ability or family income. (The study does include rates of child poverty in the geographic region.) As discussed below, these are key predictors of both children's outcomes and WIC participation.

Perhaps most striking is that the data include only WIC participants. The authors suggest that early and late participants are more comparable to each other than to nonparticipants, but that assertion is untestable with these data. Researchers consider dosage an indicator of the program impact, but in an observational study, such assessments suffer from the same difficulties as comparisons of participants with nonparticipants. For the above reasons, we rely on survey data and then address how recall bias might influence our findings.

METHODS

Participants

The study analyzes data from the CDS-PSID. The PSID includes extensive data on the income and well-being of a representative sample of U.S. families for nearly four decades. The study covers a range of topics, including family composition change, food and housing expenditures, employment, income, health, and welfare. Data were collected annually from 1968 to 1997 and biannually after 1997.

The CDS is one component of the PSID. The CDS provides comprehensive, nationally representative, and longitudinal data on children and their caregivers (McGonagle and Schoeni 2006). In 1997, PSID randomly selected 3,563 children aged 0–12 from 2,394 PSID families and collected information on their well-being, ranging from cognitive, behavioral, and health outcomes to family and neighborhood environments. While not shown in the tables, we compared the birth weight distributions of the CDS sample with those reported in the national vital statistics during the same period (1985–1997) and found they were comparable. For example, the percentage of low birth weight babies among non-Hispanic white mothers ranged from 5.6 to 6.5 from the vital statistics, and it was 5.36 for our sample; that for non-Hispanic black mothers ranged from 12.6 to 13.1 from the vital statistics, and it was 13.75 for our sample.

Our study use information on 3,181 children and their families. Three hundred and eighty-two children were excluded because (1) they did not live with their biological mother at the time of interview ($n = 271$); (2) they

were missing information on WIC participation ($n = 101$), on race and ethnicity ($n = 5$), and on date of birth ($n = 5$).

Measures

The main variable of interest is a mother's prenatal WIC participation. The child's primary care givers (usually their biological mother) reported this information during the 1997 interview.

Our outcome measures include six child health outcomes: the child's birth weight (in grams), whether the child was born low birth weight (e.g., weighed $< 2,500$ g), whether the child was born prematurely (e.g., gestational age is < 37 weeks), whether the child was born SGA (e.g., weighing less than a specified percentile of birth weight for a given gestational age; we used a gender-specific SGA measure from Alexander et al. 1996), the mother's rating of the child's health at birth as compared with other babies (1 = worse, 2 = same, 3 = better), and whether the child was placed in the neonatal intensive care unit (NICU).

Our covariates include a broad array of child, maternal, and household characteristics collected over the period of the PSID study. For child covariates, we included the child's age (in months), the child's race and ethnicity (white, non-Hispanic; black, non-Hispanic, other), the child's sex, whether the child was first born to the mother, and the number of siblings. Maternal characteristics included IQ,¹ education (high school dropout [omitted category]; completed high school, attended college), her age at the time of the child's birth, and whether she was working, married, or was the head of the household (e.g., the only adult) during the year of the child's birth. Family-level characteristics included the household's income from the year of the child's birth. This list is more extensive than that in prior research.

Matching Procedure

For a brief review of propensity score methodology, see the methodological appendix (Appendix SA2). We used PSMATCH2 (Leuven and Sianesi 2003) for *Stata 10 SE* (StataCorp. 2007). PSMATCH2 implements a variety of propensity score matching methods to adjust for preobservable differences between a treatment and a control group. The program calculates approximate standard errors on the treatment effects assuming independence among observations, fixed weights, homoskedasticity of the outcome variable within the treated and within the control groups, and that the variance of the outcome does not depend on the propensity score (Leuven and Sianesi 2003).

First, we calculated a propensity score for each individual, the predicted probability of WIC receipt. This prediction was generated using the results of a probit model for which WIC participation was the outcome. The predictors included the variables discussed above. As we mentioned earlier, the propensity score represents a “balancing score”—when matching cases on the propensity score, the distribution of the covariates between those who participated WIC and who did not should be the same (Rosenbaum and Rubin 1983b). This property, however, holds only as sample sizes approach infinity and so is only approximate in any finite sample. As a result, good practice involves comparing the distribution of covariates between groups. If the covariates do not balance, then one should modify the propensity score equation (Morgan and Winship 2007) and reassess balance. While over-fitting the data in this manner would not be desirable for many purposes (e.g., such as testing hypotheses about the determinants of WIC participation), modifying the propensity score in this way ensures that the propensity score better captures between-group differences in the covariates. Preliminary analysis suggests that running separate probit model with additional nonlinear terms for four subgroups defined by child’s sex and race improved the balance of covariates.

Having been calculated, propensity scores can be used in a variety of ways, including weighting schemes, matching, or as a covariate (Rosenbaum and Rubin 1983a). The key issue with estimating the effect of WIC is calculating a counterfactual outcome—the outcome expected had the individual not received WIC. We do this using a kernel estimator that essentially creates this estimate for a given case using a weighted average of comparison cases with a similar propensity score. The actual estimated counterfactual is a weighted average of those cases; the closer the propensity score, the greater the weight.² The econometrics literature suggests that this method is especially effective in estimating the counterfactual (Heckman and Navarro-Lozano 2004).

Handling of Missing Data

Because data were collected at multiple interviews, the amount of missing data is relatively high. Table 1 provides information on item missing data. For key variables such as maternal IQ or family income, the amount of missing is 23 percent and 14 percent, respectively. Using the ICE and MICOMBINE commands (Royston 2005) in *Stata 10 SE* (StataCorp. 2007), we multiply imputed missing values under an assumption of missing at random, analyzed each dataset, and combined the resulting parameter estimates using Rubin’s

Table 1: Descriptive Statistics of Variables (before Imputation)

Variable	All Cases			Received WIC			Did Not Receive WIC		
	Observation	Mean/ Proportion	Standard Deviation	Observation	Mean/ Proportion	Standard Deviation	Observation	Mean/ Proportion	Standard Deviation
Received WIC	3,181	0.43	0.50	1,373	1	0	1,808	0	0
Outcomes									
Mother's rating of child's health at birth*	3,170	2.20	0.57	1,364	2.16	0.55	1,806	2.24	0.59
Birth weight (g)	3,108	3,317.0	635.5	1,339	3,226.3	644.3	1,769	3,385.7	620.2
Low birth weight	3,108	0.09	0.29	1,339	0.11	0.31	1,769	0.07	0.26
Small for gestational age (SGA)	3,181	0.15	0.36	1,373	0.21	0.40	1,808	0.11	0.31
Preterm birth	3,180	0.08	0.28	1,373	0.09	0.29	1,807	0.08	0.26
Child in NICU	3,172	0.12	0.33	1,368	0.13	0.33	1,804	0.12	0.33
Covariates									
Child (at 1997 survey)									
Female	3,181	0.49	0.50	1,373	0.49	0.50	1,808	0.49	0.50
Age (months)	3,181	88.8	46.1	1,373	83.8	45.7	1,808	92.6	46.0
Number of siblings	3,136	1.4	1.2	1,353	1.5	1.4	1,783	1.3	1.1
First born	3,119	0.40	0.49	1,340	0.39	0.49	1,779	0.42	0.49
Black	3,181	0.40	0.49	1,373	0.63	0.48	1,808	0.23	0.42
White	3,181	0.48	0.50	1,373	0.22	0.41	1,808	0.68	0.47
Other	3,181	0.11	0.32	1,373	0.15	0.36	1,808	0.09	0.28

Mother (at 1997 survey)									
Age 20 (years)									
IQ	3,180	7.35	5.77	1,372	5.11	5.58	1,808	9.06	5.30
	2,464	30.69	5.2 ^a	1,060	28.38	5.2	1,404	32.44	4.75
Education									
High-school dropout	3,181	0.15	0.35	1,373	0.26	0.44	1,808	0.06	0.24
High-school graduate	3,181	0.36	0.48	1,373	0.42	0.49	1,808	0.31	0.46
College or more	3,181	0.39	0.49	1,373	0.18	0.38	1,808	0.54	0.50
Missing education	3,181	0.11	0.31	1,373	0.14	0.35	1,808	0.09	0.28
Mother (year of birth)									
Family income (U.S.\$)	2,726	35,108	31,679	1,136	20,353	18,115	1,590	45,650	34,926
Married	3,181	0.63	0.48	1,373	0.37	0.48	1,808	0.82	0.39
Employed	2,511	0.47	0.50	996	0.32	0.47	1,515	0.57	0.50
Head of household [†]	2,585	0.15	0.36	1,046	0.28	0.45	1,539	0.07	0.25

*1, worse; 2, same; 3, better as compared with other babies.

[†]Mother was the only adult in the household.

NICU, neonatal intensive care unit; WIC, women, infants, and children.

rules. Appendix SA2 also provides more detail on our multiple imputation strategy.

In addition to all the outcomes and covariates used in the previous propensity score analysis, the latest child level weights and two other IQ measures³ were included in the imputation.

RESULTS

Except for Table 1, all results reflect pooled estimates across the imputations using Rubin's rule. Table 1 provides descriptive statistics for the actual data analyzed (before the imputation), disaggregated by whether the mother received WIC.

Among our sample, 43 percent of them reported receiving WIC prenatally. Nine percent of children were born low birth weight, and 8 percent had been born prematurely. One in six mothers did not complete high school, and one-third was unmarried when the child was born.

WIC recipients compare *unfavorably* to other women in terms of their sociodemographic characteristics and some of their children's outcomes. At birth, children of WIC recipients weighed less (3,226 versus 3,386 g), were more likely to be born low birth weight (11 percent versus 7 percent), or small for gestational age (21 percent versus 11 percent). As for demographic characteristics, mothers who received WIC had lower IQs, less educational attainment, and lower household incomes the year the child was born. Those mothers also were younger, reported larger family size, and were disproportionately African American.

Table 2 presents the results of the model used to generate the propensity scores. The table presents the probit coefficients for each of the four subgroups defined by race and the child's gender, pooling across imputations using Rubin's rules. These analyses essentially identify those characteristics that distinguish WIC participants from other women. This multivariate assessment is appropriate to understanding the overall potential of the covariates to confound unadjusted comparisons of recipients and other women. Comparing across subgroups, the effect of some predictors did vary. In some instances, a characteristic predicted WIC participation only for one gender. For example, for reasons that are not entirely clear, the number of siblings predicted WIC participation only for female children. On the other hand, married mothers were less likely to use WIC than other women, and this relationship is consistent across the four subgroups.

Table 2: Propensity Score Models (Imputed)

	<i>Female</i>		<i>Male</i>	
	<i>Black</i>	<i>White/Other</i>	<i>Black</i>	<i>White/Other</i>
Covariates				
Child (at 1997 survey)				
Age (in years)	-0.04	-0.07	0.00	-0.07
	0.069	0.065	0.071	0.066
Number of siblings	0.22***	0.19***	0.04	0.09
	0.059	0.052	0.045	0.061
First born	0.22	-0.10	-0.32*	-0.25
	0.167	0.142	0.146	0.140
Mother (at 1997 survey)				
Mom's age giving birth-20	-0.04	-0.08*	-0.03	-0.10**
	0.033	0.037	0.029	0.032
Mom's IQ	-0.01	-0.03	-0.02	-0.02
	0.016	0.018	0.013	0.015
Education				
High-school graduate	-0.02	-0.17	0.04	0.01
	0.088	0.102	0.089	0.089
High-school dropout (omitted)				
College or more	-0.21***	-0.23**	-0.03	-0.10
	0.068	0.080	0.069	0.075
Mother (year of birth)				
Family income (logged)	0.61**	0.61	0.21	0.50
	0.243	0.337	0.258	0.338
Married	-0.39**	-0.35*	-0.50***	-0.46**
	0.160	0.159	0.159	0.183
Employed	-0.08	-0.16	-0.26	-0.14
	0.143	0.145	0.135	0.165
Head of household [†]	-0.10	-0.12	0.09	-0.09
	0.203	0.298	0.190	0.360
Additional terms for nonlinearity				
Mother's age (squared) [‡]	0.01	0.11	-0.13	0.28
	0.177	0.218	0.155	0.165
Child's age (squared) [‡]	-0.11	-0.06	-0.49	-0.20
	0.440	0.416	0.435	0.431
Income (logged then squared)	-0.18***	-0.19***	-0.10	-0.21***
	0.048	0.061	0.051	0.063
Child level weight (2002)	0.00	0.00	0.00	0.00
	0.010	0.005	0.010	0.005
Race (other)		0.83***		0.59***
		0.150		0.150
Intercept	1.46**	1.88**	2.14***	2.00**
	0.545	0.671	0.547	0.744
Observations	609	959	678	935

Coefficients are in the first row of each variable; standard errors are right below.

*Significant at 5%.

**Significant at 1%.

***Significant at 0.1%.

[†]Rescaled as [mom's age at giving birth-20] ^2/100

[‡]Rescaled as child's age (in years) ^2/100

As we mentioned earlier, a good practice involves checking the covariate balance after the matching. Propensity-score-based comparisons of WIC participants and nonparticipants revealed no substantial differences in the means of the covariates. For more details on covariate balance, see Appendix SA3.

Table 3 presents the estimated effects of WIC participation on the birth outcomes. Each panel represents a different comparison. For each, we present the estimated effect, the standard error, and the *p*-value. For continuous outcomes, the effect is the mean difference for the outcome for a WIC participant and the estimated counterfactual. For the dichotomous outcomes (like preterm), the effect is the difference in the predicted probability of the outcome.

First, consider the unmatched samples in the leftmost panel. Results indicated that children born to WIC-recipients had lower scores on the maternal health rating of the child, were more likely to be born low birth weight, born small for gestational age, and had lower birth weights. Given the obvious confounding, one could argue that these figures are of little or no use, but they do represent a baseline against which adjusted figures can be assessed. (One also can compare these figures to other datasets to assess across-study differences in sample composition.)

For the matched sample (second panel), the between-group differences were no longer significant. The differences between the matched and unmatched sample were quite striking: for example, in the unmatched sample, the difference in birth weight between the two groups was 159 g (or around 5.6 ounces). In the matched sample, the difference was only 18 g (or 0.7 ounces) and was no longer statistically significant at 0.05 level. Likewise, the difference in the number of babies born low birth weight changed from 4 percent higher in the unmatched sample to a 1 percent decrease in the matched sample. Most of the outcomes were pointing to the anticipated direction (improved outcomes) and, if not, matching reduced the magnitude of negative effects.

Sensitivity Analyses

We addressed several limitations of our analyses and methodology. First, one potential problem with our data is the length of recall required for the women with older children in 1997. We limited the analyses to the 1,627 children ages 7 and younger.

Second, the analyses reported above include nonrecipients who may not have been eligible for WIC and so are not comparable to those using WIC. We reduced this heterogeneity by limiting the analysis to those women

Table 3: Effects of WIC Program on Birth Outcomes

Outcomes	Unmatched			Propensity Score Matched with Multiple Imputations						Fixed Effects		
				Full Sample		Kid < 7		Medicaid Sample				
	Effect [†]	SE	p [‡]	Effect	SE	Effect	SE	Effect	SE	Effect	SE	p [‡]
				N = 3,181		N = 1,627		N = 1,107		N = 1,742 pairs		
Mother's rating of child's health at birth [§]	-0.08	0.02	***	-0.04	0.03	0.03	0.06	-0.07	0.06	0.08	0.07	
Birth weight (grams)	-159.4	22.8	***	-18.2	37.9	-69.7	59.3	-19.4	73.3	176.5	67.1	**
Low birth weight	0.04	0.01	***	-0.01	0.02	-0.01	0.03	-0.02	0.04	-0.09	0.03	**
Small for gestational age (SGA)	0.09	0.01	***	0.03	0.02	0.07	0.03	0.03	0.04	-0.03	0.04	
Preterm birth	0.02	0.01		-0.02	0.02	-0.02	0.03	-0.02	0.04	-0.07	0.03	*
Child in NICU	0.01	0.01		0.00	0.02	-0.06	0.03	-0.04	0.04	-0.03	0.04	

*Significant at 5%.
**Significant at 1%.
***Significant at 0.1%.
[†]p-values refer to the null hypothesis that the mean for WIC is equivalent to that for the “nonWIC” category.
[‡]Effect size, which represents the difference in means divided by the standard deviation for the nonWIC participants (the reference group).
[§]1, worse; 2, same; 3, better as compared with other babies.
NICU, neonatal intensive care unit; SE, standard error; WIC, women, infants, and children.

enrolled in the Medicaid program at the time of the child's birth. Doing so makes our study more comparable to prior research such as Bitler and Currie (2005), who limit their analyses to women whose delivery was paid for by Medicaid. In both instances, while statistical significance varies somewhat, the fundamental results of the earlier analyses are unchanged.

As noted before, propensity score methodology does not allow for between-group differences in *unobserved* factors affecting both participation and child outcomes. We addressed this problem in two ways. First, we estimated fixed-effects models that allowed for mother-specific unobservables. In essence, these analyses involved sibling, within-family comparisons and have been used to examine the impact on child outcomes of a range of maternal and family characteristics such as breastfeeding, mother's age, and neighborhood conditions (Rosenzweig and Wolpin 1995; Aaronson 1997; Der, Batty, and Deary 2006; Holmlund, 2005). In this case, fixed-effect estimation involves comparisons of children born to mothers when receiving WIC with their siblings born to the same mother when she was not using WIC. These results are reported in the last two panels of Table 3. These analyses shed two insights. First, comparisons of the fixed- and random-effects estimates indicate that the two sets of estimates differ statistically. (At the request of a reviewer, we do not present the latter here, but they are available from the authors.) In general, the fixed-effects estimates are larger.

Even within fixed-effect estimation, time-varying unobserved differences may confound comparisons of those children who did and did not receive WIC. Such factors would vary within child and may reflect changes in the mother's circumstances or broader developments in the community. For that reason, we conducted sensitivity analyses like those proposed by Rosenbaum (2002) and Imbens (2003). Such sensitivity analyses assess how strongly confounding with an unobserved variable would have to be to explain the apparent effect (or noneffect) of a treatment (WIC participation). Following Rosenbaum (2002), we began by grouping the data into pairs based on the propensity score. Each pair included a WIC participant and a matching nonparticipant drawn without replacement.⁴

First, we examined the pattern of matched pairs assuming no *unobserved* confounding. Among these pairs, one can calculate a nonparametric test (the sign test) of the relationship between WIC and the outcome using discordant pairs. Taking low birth weight as an example, our matching produced 1,373 pairs. Of these, 245 pairs were "discordant"—one birth in the pair was low birth weight and one was not, and the low birth weight baby was born to the WIC participant in 141 (58 percent) pairs. If WIC users and nonusers were

perfectly matched, any deviation from 50 percent (the null hypothesis) would reflect the effect of WIC. One can use binomial distribution to calculate the corresponding p -value on the observed effect.

The sensitivity analyses then considered whether this pattern could be explained by hypothetical, unobserved dichotomous characteristic (“healthy” versus “unhealthy”). The link between “healthy” status and WIC participation would have to be relatively strong to explain the null findings. For birth weight, for example, 64 percent of women enrolling in WIC would have to be drawn from the “unhealthy” class. This relationship would have had to exist *over and above* the balancing of observed characteristics. Our sensitivity analyses for the other outcomes were similar and are available from the first author.

DISCUSSION

Using propensity scores, this study has examined the association between WIC and the health of newborns. In general, we find modest effects of the WIC program, and our findings seem robust to changes in model specification. We also considered the possibility of unobserved confounding, estimating fixed-effects estimates and Rosenbaum bounds. Neither of these supplemental analyses suggests that our findings for maternal rating of the child’s health, for spending time in the NICU, or being small for gestational age are affected by WIC.

While prior studies are far from uniform on the benefits of WIC, these effects are generally smaller than those found in earlier research. We believe several explanations are possible. First, we are able to include more and better covariates in our analyses that adjust for the confounding caused by such characteristics. As a result, earlier studies may exaggerate the effect of WIC.

Second, the handling of Medicaid status actually may inflate the estimated effect of the program. Medicaid participation presumably will be positively associated with both children’s health and with WIC participation. The latter may involve a direct effect—WIC may enroll women in Medicaid. According to the standard formula for omitted variable bias, failing to include Medicaid status as a covariate may inflate the estimated effect of WIC (Greene 2008). However, that framework assumes the Medicaid *itself* is not confounded by unobservables (Sobel 2008). If such confounding exists, then Medicaid is a “collider,” and conditioning on it establishes a spurious correlation between unobserved determinants of Medicaid enrollment and WIC (Pearl 2000, 2005; Greenland and Pearl 2008). If women who are of poorer

health enroll in Medicaid (which seems likely if providers enroll women in the program), the effect of WIC is inflated. As shown above, when we limit our analyses to Medicaid enrollees, the estimated effects are larger (though still not statistically significant).

However, some results are encouraging from a program perspective. We do find fairly substantial effects of WIC for some outcomes in the fixed-effect analyses. These effects involve two outcomes related to birth weight and pre-term status, and all three are fairly substantial. They appear, however, only in the fixed-effect analysis. As a result, the policy implications drawn by the reader likely depend on how much confidence he or she has in the ignorability assumption. The most straightforward argument is that WIC reflects time-varying unobserved determinants of health, and that the fixed-effect estimates are preferred. From that perspective, the bottom line for the paper is much more positive.

However, the difference between the propensity score-based and fixed-effect estimates may be explained by relatively subtle forces. One reason the fixed- and propensity-score estimates differ is that even the latter were larger in the sample of families with two or more children. One explanation is that some predictors of family size—either observed or unobserved—may moderate the effect of WIC. In that case, the fixed-effect estimates indicate not that the propensity score estimates are wrong but are estimating a different effect. If larger families are more disadvantaged, then perhaps the program's effect is larger among those families.

Another possibility is methodological in nature. It is also possible that controlling for the fixed effects actually increased other biases, such as that involving time-varying unobservables and WIC participation. This pattern would imply that time-varying and time-invariant unobservables bias the findings in different directions. Whether that is the case is unknown. Our sensitivity analyses, however, show that the net effect of these two biases would have to be fairly strong to explain the null findings.

Other limitations of our own analyses remain. First, data on WIC were based on maternal recall. The rate of WIC participation is somewhat lower than one would anticipate. However, for recall to influence the estimated impact of WIC, the relationship between actual and reported WIC status would have to be moderated by the outcomes of interest. For example, if healthy and unhealthy women underreport their WIC participation to the same degree, the effect of WIC would not be affected. We have no evidence that this effect exists, and resolving the issue would require administrative data on participation.

Second, we might have adjusted for additional covariates, and doing so might reveal benefits of the program otherwise obscured. Indeed, many more variables are included in the data that might be used to calculate the propensity score. However, as we performed these analyses, as more covariates were added, the estimate of the program effect moved closer and closer to zero.

Third, this study has examined prenatal WIC participation, but pregnant women constitute only 11 percent of the WIC caseload (USDA 2006). Therefore, our results do not address the effectiveness of the program for postpartum women and children.

Finally, our study suffers from the limitations inherent to observational studies. However, the putative gold standard of evaluation—a randomized trial—is not possible or even desirable for a nearly universal program like WIC. A body of research considers the circumstances under which observational analyses produces findings similar to a randomized experiment (Cook, Shadish, and Wong, 2008; Steiner et al. in press). That literature suggests that a well-done observational study can produce findings quite similar to those of a randomized trial (where both exist). The key issue seems to be an understanding of the process shaping program participation and the availability of appropriate covariates that capture the key, shared determinants of program participation and the outcomes of interest.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was not funded by any governmental or nongovernmental agency.

Disclosures: None.

Disclaimers: None.

NOTES

1. This measure is taken from the passage completion test of Woodcock Johnson test-revised administered to primary caregivers in a 1997 CDS interview.
2. Even though we run race-, gender-specific propensity score model, every child only has one propensity score. In calculating the kernel density estimate, we did not limit eligible cases to those of the same race and gender.
3. They are grandmother's and mother's IQ, both taken from a short form IQ test with 13 sentence completion questions (taken from the Lorge-Thorndike intelligence test) administered in the 1972 interview (Veroff, McClelland, and Marquis, 1971).

4. We matched the pairs using the optimal matching procedure in *MatchIt* (Ho et al. 2007), a software package in R.

REFERENCES

- Aaronson, D. 1997. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes." *The Journal of Human Resources* 33 (4): 915–46.
- Alexander, G. R., J. H. Himes, R. B. Kaufman, J. Mor, and M. Kogan. 1996. "A United States National Reference for Fetal Growth." *Obstetrics and Gynecology* 87 (2): 163–68.
- Allison, P. D. 2002. *Missing Data*. Thousand Oaks, CA: Sage.
- Angrist, J., and S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Bitler, M., and J. Currie. 2005. "Does WIC Work? The Effects of EIC on Pregnancy and Birth Outcomes." *Journal of Policy Analysis and Management* 24 (1): 73–91.
- Brown, H. L., K. Watkins, and K. A. Hiatt. 1996. "Fetus-Placenta-Newborn: The Impact of the Women, Infants, and Children Food Supplement Program on Birth Outcome." *American Journal of Obstetrics and Gynecology* 174 (4): 1279–83.
- Chatterji, P., and J. Brooks-Gunn. 2004. "WIC Participation, Breastfeeding Practices, and Well-Child Care among Unmarried, Low-Income Mothers." *American Journal of Public Health* 94 (8): 1324–27.
- Cook, T. D., W. R. Shadish, and V. C. Wong. 2008. "Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management* 27 (4): 724–50.
- Currie, J. 2003. "U.S. Food and Nutrition Programs." In *Means-Tested Transfer Programs in the United States*, edited by R. Moffitt, pp. 199–289. Chicago: The University of Chicago Press.
- Dehejia, R. H., and S. Wahba. 2002. "Propensity Score-Matching Methods for Non-experimental Causal Studies." *Review of Economics and Statistics* 84 (1): 151–61.
- Der, G., G. D. Batty, and I. J. Deary. 2006. "Effect of Breast Feeding on Intelligence in Children: Prospective Study, Sibling Pairs Analysis, and Meta-Analysis." *British Medical Journal* 333 (7575): 945–50.
- Devaney, B. L. 1992. *Very Low Birthweight among Medicaid Newborns in Five States: The Effects of Prenatal WIC Participation*. Alexandria, VA: U.S. Department of Agriculture.
- Foster, E. M., G. Y. Fang and the Conduct Problems Prevention Research Group. 2004. "Alternative Methods for Handling Attrition in Prevention Research: An Illustration Using Data from Fast Track." *Evaluation Review* 28: 434–64.
- Foster, E. M., and A. Krivelyova. 2007. Non-Ignorable Non-Response in Longitudinal Studies." In *Handbook of Longitudinal Research*, edited by S. Menard, pp. 185–196. Burlington, MA: Elsevier.

- GAO. 2004. *Nutrition Education: USDA Provides Services through Multiple Programs, But Stronger Linkages among Efforts Are Needed*. Washington, DC: Government Printing Office.
- Gibson-Davis, C. M., and E. M. Foster. 2006. "A Cautionary Tale: Using Propensity Scores to Estimate the Impact of Food Stamps on Food Insecurity." *Social Service Review* 80: 93–126.
- Goldenberg, R. L., and D. J. Rouse. 1998. "Prevention of Premature Birth." *New England Journal of Medicine* 339 (5): 313–20.
- Greene, W. H. 2008. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Greenland, S., and J. Pearl. 2008. "Causal Diagrams." In *Encyclopedia of Epidemiology*, edited by S. Boslaugh and L.-A. McNutt, pp. 149–56. Los Angeles, CA: Sage.
- Heckman, J. J., and S. Navarro-Lozano. 2004. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86 (1): 30–57.
- Heckman, J. J., and E. J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." *Handbook of Econometrics* 6: 4779–874.
- Ho, D. E., K. Imai, G. King, and E. A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.
- Holmlund, H. 2005. "Estimating Long-Term Consequences of Teenage Childbearing—An Examination of the Siblings Approach." *Journal of Human Resources* 40 (3): 716–43.
- Iams, J. 1998. "Prevention of Preterm Birth." *New England Journal of Medicine* 339 (5): 313–230.
- Imbens, G. W. 2003. "Sensitivity to Exogeneity Assumptions in Program Evaluation." *The American Economic Review* 93: 126–32.
- Jacknowitz, A., D. Novillo, and L. Tiehen. 2007. "Special Supplemental Nutrition Program for Women, Infants, and Children and Infant Feeding Practices." *Pediatrics* 119 (2): 281–89.
- Joyce, T., D. Gibson, and S. Colman. 2005. "The Changing Association between Prenatal Participation in WIC and Birth Outcomes in New York City." *Journal of Policy Analysis and Management* 24 (4): 661–85.
- Kahler, L. R., R. M. O'Shea, L. C. Duffy, and G. M. Buck. 1992. "Factors Associated with Rates of Participation in WIC by Eligible Pregnant Women." *Public Health Reports* 107 (1): 60–5.
- Kennedy, E. T., and M. Kotelchuck. 1984. "The Effect of WIC Supplemental Feeding on Birth Weight: A Case-Control Analysis." *The American Journal of Clinical Nutrition* 40: 579–85.
- Kramer, M. S. 2003. "The Epidemiology of Adverse Pregnancy Outcomes: An Overview." *The Journal of Nutrition* 133: 1592S–96S.
- Lazariu-Bauer, V., H. Stratton, R. Pruzek, and M. L. Woelfel. 2004. "A Comparative Analysis of Effects of Early versus Late Prenatal WIC Participation on Birth Weight: NYS, 1995." *Maternal and Child Health Journal* 8 (2): 77–86.

- Lee, M. 2005. *Micro-Econometrics for Policy, Program, and Treatment Effects*. Oxford, England: Oxford University Press.
- Leuven, E., and B. Sianesi. 2003. "PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing." Available at <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Little, R. J. A. 1993. "Pattern-Mixture Models for Multivariate Incomplete Data." *Journal of the American Statistical Association* 88 (421): 125–34.
- . 1994. "A Class of Pattern-Mixture Models for Normal Incomplete Data." *Econometrika* 81 (3): 471–83.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley and Sons.
- . 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley and Sons Inc.
- Ludwig, J., and M. Miller. 2005. "Interpreting the WIC Debate." *Journal of Policy Analysis and Management* 24 (4): 691–701.
- May, R., J. Tramp, K. Bass, S. Schroeder, and M. A. Kerr. 2001. "Early Postnatal Growth of Low Birthweight Infants in the WIC Program." *American Journal of Human Biology* 13: 261–67.
- McGonagle, K. A., and R. F. Schoeni. 2006. "The Panel Study of Income Dynamics: Overview and Summary of Scientific Contributions after Nearly 40 Years" [accessed on May 21, 2007, 2006]. Available at <http://www.psidonline.isr.umich.edu/Publications/Papers/montreal.pdf>
- Morgan, S. L., and C. Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, England: Cambridge University Press.
- Moss, N. E., and K. Carver. 1998. "The Effect of WIC and Medicaid on Infant Mortality in the United States." *American Journal of Public Health* 88 (9): 1354–61.
- Oliveira, V., E. Racine, J. Olmstead, and L. M. Ghelfi. 2002. *The WIC Program: Background, Trends, and Issues*. Food Assistance and Nutrition Research Report No. 27, Food and Rural Economics Division, Economic Research Service, U.S. Department of Agriculture.
- Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press.
- . 2005. "Direct and Indirect Effects." *Proceedings of the American Statistical Association Joint Statistical Meetings* 411 (20): 1572–81.
- Rosenbaum, P., and D. Rubin. 1983a. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.
- Rosenbaum, P. R. 2002. *Observational Studies*. New York: Springer-Verlag.
- Rosenbaum, P. R., and D. B. Rubin. 1983b. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.
- Rosenzweig, M., and K. Wolpin. 1995. "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context." *Econometrica* 63 (2): 303–26.

- Rossi, P. H. 1998. *Feeding the Poor: Assessing Federal Food Aid*. Washington, DC: The American Enterprise Institute.
- Royston, P. 2005. "Multiple Imputation of Missing Values: Update of Ice." *Stata Journal* 5 (4): 527–36.
- Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.
- Rush, D., J. Leighton, N. L. Sloan, J. M. Alvir, D. G. Horvitz, W. B. Seaver, G. C. Garbowski, S. S. Johnson, R. A. Kulka, J. W. Devore, J. T. Lynch, M. B. Woodside, and D. B. Shanklin. 1988. "The National WIC Evaluation: Evaluation of the Special Supplemental Food Program for Women, Infants, and Children: Study of Infants and Children." *American Journal of Clinical Nutrition* 48: 484–511.
- Sobel, M. E. 2008. "Identification of Causal Parameters in Randomized Studies With Mediating Variables." *Journal of Educational and Behavioral Statistics* 33 (2): 230–51.
- StataCorp. 2007. *Stata Statistical Software (Release 10)*. College Station, TX: Stata-Corp.
- Stein, Z., and M. Susser. 1975. "The Dutch Famine, 1944–1945, and the Reproductive Process. I. Effects on the Six Indices at Birth." *Pediatric Research* 9: 70–6.
- Steiner, P. M., T. D. Cook, W. R. Shadish, and M. H. Clark. in press. "The Importance of Covariate Selection in Controlling for Selection Bias in Observational Studies." *Psychological Methods*.
- Stockbauer, J. W. 1987. "WIC Prenatal Participation and Its Relation to Pregnancy Outcomes in Missouri: A Second Look." *American Journal of Public Health* 77: 813–18.
- UCLA: Academic Technology Services and Statistical Consulting Group. "Multiple Imputation Using ICE" [accessed on August 11, 2009]. Available at <http://www.ats.ucla.edu/stat/Stata/library/ice.htm>
- USDA. 2006. *WIC Participant and Program Characteristics 2004*. Alexandria, VA: USDA.
- . 2007a. "The Food Assistance Landscape Fiscal Year 2006." Washington, DC: U.S. Department of Agriculture, Economic Research Service.
- . 2007b. "WIC Program: Average Monthly Benefit per Person" [accessed on May 23, 2007]. Available at [http://www.fns.usda.gov/pd/25wifyavgfd\\$.htm](http://www.fns.usda.gov/pd/25wifyavgfd$.htm)
- Verbeke, G., and G. Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Veroff, J., L. McClelland, and K. Marquis. 1971. *Measuring Intelligence and Achievement Motivation in Surveys: Final Report to HEW, OEO*. Ann Arbor, MI: University of Michigan.
- Winship, C., and S. L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25 (1): 659–707.
- Zedlewski, S. R., and K. Rader. 2005. "Feeding America's Low-Income Children." In *New Federalism, National Survey of America's Families*. Series B, No. B-65. Washington, DC: The Urban Institute.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2. Methodological Appendix.

Appendix SA3. Details on Balancing.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.