

# Breastfeeding and the Child Cognitive Outcomes: A Propensity Score Matching Approach

Miao Jiang · E. Michael Foster ·  
Christina M. Gibson-Davis

Published online: 17 September 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** To estimate the effect of breastfeeding initiation and duration on child development outcomes. 3,271 children and their mothers participating in the Child Development Supplement of the Panel Study of Income Dynamics provide data for these analyses. Main outcomes include Woodcock Johnson Psycho-Educational Battery-Revised (WJ-R) test score (letter word, passage comprehension, applied problem, and broad reading), and Wechsler Intelligence Scale for Children-Revised (WISC-R) test score at the 2002 survey. Controlled variables include family, maternal, and child characteristics, many of which can be traced back to the year the child was born. The analytic technique is propensity score matching with multiple imputations. After using propensity scores to adjust for confounding factors, breastfeeding initiation showed statistically significant effects but the practical scale remains small. Breastfeeding duration showed a non-linear effect on those outcomes and most of the effects are not significant. The effects of breastfeeding on child's cognitive outcomes are modest in practical terms. The non-linear effects suggest that selection into breastfeeding may account for the increased score of children who are breastfed.

**Keywords** Breastfeeding · Child cognitive outcomes · Propensity score · Multiple imputations · Confounding · Effect size

---

M. Jiang (✉) · E. M. Foster  
Department of Maternal and Child Health, University of North Carolina-Chapel Hill, Rosenau Hall, CB# 7445, Chapel Hill, NC 27599-7445, USA  
e-mail: mjiang@unc.edu

C. M. Gibson-Davis  
Terry Sanford Institute of Public Policy, Duke University,  
PO Box 90245, Durham, NC 27708, USA

## Introduction

While breastfeeding is widely regarded as providing numerous benefits for child and maternal health [1], its effect on children's cognitive development remains a matter of debate. Breastfeeding is thought to directly affect child cognition through the superior nutrients found in breast milk; breast milk contains long-chain polyunsaturated fatty acids (LCPUFAs) [2–7]. LCPUFAs, particularly  $n - 3$  docosahexaenoic (DHA) and  $n - 6$  arachidonic acid (AA), are necessary for retinal and neural development [8–11], and animal studies indicate that severe LCPUFA deprivation harms cognitive and behavioral performance [12]. Evidence showing a causal role between LCPUFAs and enhanced human neural functioning, however, is inconclusive [10, 12–14].

Likewise, the association between breastfeeding and cognition also has been mixed. Some find positive effects of breastfeeding [4, 15–20], and a meta-analysis of 20 studies find that on average, breastfeeding increases cognition by five points for low birth weight infants and three points for normal weight infants [21]. (These estimates imply small effect sizes.) Two recent reviews [22, 23] conclude that, based on evidence from more rigorous studies, breastfeeding does improve cognition, particularly for children who are low birth weight. Others find that these effects are spurious, explained by confounding factors, such as genetic and socioenvironmental factors [24–26]. Furthermore, researchers disagree on the effects of duration of breastfeeding on child outcomes—some find a significant relationship [19, 21, 27, 28]; others, a non-linear dose–response relationship [29]; and still others, no significant relationship after adjusting for confounders [30].

Much of the ambiguity in prior research reflects difficult challenges, especially those of identifying the causal effect

of any risk factor outside of random assignment. In observational analyses, researchers face difficult choices involving how to condition on observed differences between women who do and do not breastfeed and how to assess the possibility of unobserved confounding. For example, recent research indicates that maternal IQ and the home stimulation environment predict breastfeeding status better than even poverty or education [24, 25], yet few studies include these two measures as covariates. In addition, many studies have relied on clinical based samples [7, 19, 24] or have only considered low birth weight infants [16, 31]. Such samples suffer from a range of problems involving both internal and external validity [32]. For example, conditioning on low birth weight likely creates a spurious negative correlation between breastfeeding and (unobserved) positive determinants of the health of women and children, thus exaggerating the effect of breastfeeding. (For details, see Greenland [33] or Pearl [34].) Other issues include non-standardized measures of breastfeeding, small sample sizes, and retrospective collection of data.

This article examines the effect of breastfeeding initiation and duration on child cognitive outcomes. Our goal is to address the methodological problems afflicting earlier studies. For example, we use propensity score matching to correct for selection on observables in estimates of the association between breastfeeding and child outcomes. This technique represents an alternative approach to ordinary regression for adjusting for observed confounding factors. Furthermore, we examine outcomes for a nationally representative sample. Thus, our results can be more easily generalized to the overall population than can results from previous studies with restricted samples. Moreover, we examine five measures of cognitive development. This range of outcomes allows us to identify an overall pattern of benefits of breastfeeding on child's cognition. The comprehensiveness of the data, including measures of maternal IQ and home cognitive stimulation, enables us to better account for the effect of observed confounders that are likely to bias our estimates.

## Methods

### Study Sample

The Panel Study of Income Dynamics (PSID) includes extensive data on the income and well-being of a representative sample of U.S. families over 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 number of families involved has grown from

around 5,000 in 1968 to 8,000 as of 2005 [35]. These data are the source of much of what is known about poverty, income instability and family well-being in the U.S. population [36]. The PSID, for example, was the first survey to document that poverty status was not always permanent.

The Child Development Supplement (CDS) is a special component of PSID. In 1997, PSID randomly selected 3,563 children from 2,394 PSID families and collected information on a broad aspect of their lives, from cognitive, behavioral, and health status to family and neighborhood environment. Those children were born between 1984 and 1997. In 2002–2003, these families were re-interviewed and data were successfully collected on 2,907 children [37]. The CDS provides researchers with comprehensive, nationally representative, and longitudinal data on children and their caregivers [35].

Our study is based on data from 3,271 CDS children and their families. Of the original 3,563 children, 292 children were excluded as follows: (1) they did not live with their biological mother at the time of the 1997 interview ( $n = 271$ ); (2) they were missing information on breastfeeding status ( $n = 11$ ); (3) they were missing information on race and ethnicity ( $n = 5$ ); and (4) they were missing information on date of birth ( $n = 5$ ). For any other information missing, we imputed those values using multiple imputations as described below.

### Variables

#### Outcomes

Woodcock Johnson Psycho-Educational Battery-Revised (WJ-R) [38] measures achievement in several domains of intellectual ability, including current developmental status, degree of mastery in reading and mathematics, and group standing (either age or grade group) [39]. CDS includes three subsets to measure reading and math skills: the letter word, passage comprehension, and applied problems sets, as well as two summation measures for broad reading and broad math.

Wechsler Intelligence Scale for Children-Revised (WISC-R) [40] tests for the child's ability to process information as it directly measures the short-term working memory in terms of how many sequential digits the respondent can repeat backward or forward. CDS includes digit-forward, digit-backward, and total score. This analysis uses the total score.

These test measures are available for both the 1997 and the 2002 surveys (except broad math, which was only administered in the 1997 survey), but we use the 2002 measures for two reasons. First, the variables indicating the stimulating environment at home were collected at the same time as the 1997 outcome measures. Therefore, to

reduce the risk of endogeneity, we rely on outcomes observed at the later time point. Second, there are less missing values in the 2002 outcomes because more children had reached either 3 or 6 years old to be eligible to take the cognitive tests.

### *Breastfeeding*

Our independent variables of primary interest are initiation and duration of breastfeeding. Both questions were asked in the 1997 CDS interview. Breastfeeding initiation is a dummy variable with “1” indicating ever breastfeeding. Breastfeeding duration is grouped into four categories: never breastfed; less than 6 months, 7–12 months, and more than 12 months. We examine alternative cutoffs under the sensitivity analysis.

### *Covariates*

Based on previous literature and the availability of the PSID and CDS data, we select the following covariates to be included in our analyses. For child covariates, we include the child’s age (in months, continuous), race and ethnicity [white, non-Hispanic (omitted category); black, non-Hispanic; and other], sex, number of siblings, whether the child was first born to the mother, whether the child was born preterm (e.g., gestational age is less than 37 weeks), whether the child was born small for gestational age (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 and colleagues [41]), the mother-rated child’s health at birth as compared to other babies (1: worse; 2: same; 3: better; treated as continuous variable), and the Home Observation for Measurement of the Environment (HOME) scale [42]. Maternal characteristics include IQ (taken from the passage completion test of Woodcock Johnson test-Revised administered to primary care givers in the 1997 CDS interview), education [high school dropout; completed high school; attended college(omitted category)], her age at the time of the child’s birth, her self-rated health status (from 1 to 5: poor, fair, good, very good, and excellent, treated as continuous variable), whether she was enrolled in Women, Children, and Infant (WIC) program or Medicaid during pregnancy, and whether she was working, married, or the head of the household (e.g., the only adult) the year of the child’s birth. Family level characteristics include the household’s income from the year of the child’s birth.

Sampling weights are available in the data. These weights are not included in the propensity score equation, as the equation has already included most of the covariates that are used to calculate the weights (e.g., race). In addition, preliminary analysis reveals that including the survey

weight in the propensity score equation does not make any real difference in predicting the breastfeeding status. The 2002 survey weight is, however, used in the multiple imputations as a covariate to provide more information for filling the missing values of the data.

### *Statistical Analysis*

#### *Propensity Score Matching*

A propensity score is the conditional probability of assignment to treatment given a set of covariates. Rosenbaum and Rubin [43] demonstrate that given the propensity score, treatment assignment and the observed covariates are conditionally independent. They further demonstrate that if the treatment is ignorable given the covariate vector  $X$ , then it is also ignorable given the propensity score. This assumption of “ignorability” is the key in propensity score matching (as well as in ordinary least square regression). It means conditional on observed characteristics, treatment assignment is effectively randomized. Under this assumption, one can estimate the treatment effect by conditioning on a balanced propensity score. Therefore, propensity score summarizes the multi-dimension covariate vector into a single measure.

Propensity score matching has several advantages relative to ordinary regression. First, regression generally assumes linearity between the covariates and the outcome of interest. When this assumption does not hold, the model may be mis-specified, leading to over- or understate of the program effect. (Model mis-specification creates a form of omitted variable bias.) Propensity score matching relaxes the linearity assumption, thus is less sensitive to the functional form between the covariates and the outcome.

Second, when the distributions of the covariates among the treated and the comparison cases do not overlap, regression uses extrapolation to form the predicted counterfactual values for those cases unique to one group or the other. This procedure can lead to misleading estimates of the treatment effect if the model’s functional form is not correct [44]. Propensity score matching, on the other hand, identifies treated cases (in our case, breast-feeding mothers) that have no comparable comparison cases (they have a propensity score close to 1). These mismatched cases can then be excluded from the analysis.

Third, when the treatment effects varies across subgroups of individuals (i.e., moderated by observed covariates), regression produces an estimated effect that is not equal to the effect of treatment on the treated; neither can it be converted to the effect of treatment on the untreated, or the effect of treatment in the population. For more discussion, see [45].

Note however, like regression, propensity score matching relies on the key assumption of no unobserved

confounders [46–49]. We conduct the sensitivity analysis examining this issue.

#### *The Propensity Score Model for the Breastfeeding Initiation*

First, we calculate the propensity score—the predicted probability of breastfeeding initiation for each individual—using a probit model. The next step is to check the covariates balance. The propensity score represents a “balancing score”—when matching cases on the propensity score, the distribution of the covariates between those who did and did not breastfeed should be the same [43]. This property, however, holds only as sample sizes approach infinity and so is only approximate in any given sample. As a result, a good practice involves comparing the distribution of covariates between groups adjusted for propensity score. If the covariates do not balance, then one should modify the propensity score equation [50] and re-match the cases. While over-fitting the data in this manner would not be desirable for many purposes (e.g., such as testing the hypotheses about the determinants of breast-feeding), modifying the propensity score in this way ensures that the propensity score better captures the between-group differences in the covariates.

To improve the covariates balance, we ran race/ethnicity-specific propensity score models defined by black, non-Hispanic ( $n = 1299$ ) and non-black (white, non-Hispanic ( $n = 1,547$ ) and other ( $n = 425$ )). Because the number is not big enough to run separate model for the “other” group, we still include a dummy variable for race = other in the model.

The covariates balance was further improved by including square terms for some of the continuous predictors such as age, and family income. Note each child still only has one propensity score, which is calculated using the prediction model corresponding to his or her specific gender-race subgroup.

With the propensity score, one can then estimate the effect of breastfeeding. For each treatment case, we first obtain a counterfactual outcome by calculating a weighted average outcome of all non-breastfeeding children. The closer a non-breastfeeding child’s propensity score is to the treatment case, the greater the weight. The econometrics literature suggests that this method is especially effective in estimating the counterfactual [51]. Then the mean difference between the actual and the counterfactual outcome of the treatment cases is the average effect of breastfeeding for a breastfeeding child.

In addition, we calculate the effect size, which is the standardized mean difference for the treatment effects regardless of the sample size [52]. This measure reflects the

magnitude of the estimated effects of breastfeeding on child cognitive outcomes more accurately because we have multiple outcomes and the statistical significance subjects to the issue of multiple testing. (We adjusted the  $P$  value using Bonferroni correction.) Cohen has defined, with caution, an effect size of 0.2 as “small”, 0.5 as “medium”, and 0.8 as “large” [52]. We adopt these conventions here.

#### *The Propensity Score Model for the Breastfeeding Duration*

Until recently, propensity score methods have been limited to the binary treatment situation (e.g. single treatment and comparison group). Imbens, however, extends the method to the multi-group situation where treatment can have several levels (ordered or not) [53]. His work builds on the fact that propensity score matching essentially involves a weighting scheme [54]. Like the Horvitz-Thompson estimator [55], we can get an average effect by weighing observations by the inverse of the probability of treatment level (extent of breastfeeding) actually observed [53].

We follow Imbens’ procedure to examine the relationship between duration of breastfeeding and child outcomes. The propensity score for the four categories of duration is calculated from a multinomial logit model. The resulting coefficient estimates are used to calculate the propensity score for each duration category. Then we run ordinary regression that included three dummy variables on breastfeeding duration (the “never breastfeeding” category serving as the omitted group), with the inverse of the propensity score for each category as the weights.

#### *Handling of Missing Data*

The analysis relies on data collected at multiple interviews. This feature is a strength of the study—data are available on family income in the actual year the birth occurred. However, the rate of missing data is relatively high. Therefore, our analyses rely on multiple imputations (MI), which essentially created multiple datasets with filled-in estimates for the missing values. To maximize the use of information available, we include child level weights of 2002 and two other IQ measures collected from the 1972 PSID interview [56] as additional covariates in the imputation. We estimated four imputations and combined the estimates using Rubin’s rules [57].

The analyses were conducted using PSMATCH2 [58] as well as ICE and MICOMBINE commands [59] embedded in Stata SE 10 [60].

Except the descriptive statistics, all other results are pooled estimates across the four imputations using Rubin’s rule.

## Results

### Descriptive Statistics

Table 1 provides descriptive statistics for our study sample. The number of observations represents what is available in the data. It varies across the rows of the table—some variables were missing much more often than other. (Descriptive statistics for the data set after imputations are available from the authors.)

At baseline (the 1997 survey), the children were on average 89 months old (7.5 years old), the youngest being 4 months old, and the oldest 15 years old. Half were girls; 40% were African Americans, and 41% were the first child of the family (see Table 1). 8% of children were born prematurely, and 15% were born small for gestational age.

Also at baseline, one in seven mothers did not complete high school, and one-third were unmarried when the child was born. Mothers were on average 27 years old at the time of giving birth. A little more than one-third of them received Medicaid during the pregnancy and 43% of them participated in the WIC program. 46% of the mothers reported having ever breastfed their babies.

Unadjusted comparisons suggest that in terms of the outcomes and the sociodemographic characteristics, children who were breastfed have advantages over those who were not. Breastfed children had higher scores on WJ-R and WISC-R tests, were disproportionately white, less likely to be born early (6% vs. 10%) or born small for gestational age (10% vs. 20%), and had a more cognitively stimulating home environment. Furthermore, the mothers of breastfed children tend to be older, with higher IQs and higher levels of educational achievement. Additionally, mothers who breastfed, in the year of the child's birth, had higher annual family incomes (\$43,695 vs. \$28,357), were more likely to be married (77% vs. 50%), and less likely to be enrolled in the WIC or Medicaid program.

### Checking the Balance of Covariates

Figure 1 illustrates how the covariate balance is examined using one outcome, the letter word test. We plot the  $t$  statistics for comparing the group-means on every covariate. Any points outside of the region between the two vertical dashed lines are significant at  $P = 0.05$  levels.

The unadjusted comparison of the covariates shows a number of covariates were significantly different between the two groups (see points for “Unmatched”). These covariates include, for example, family income at the year of child's birth, mother's education, WIC enrollment, and the proportion of mothers who were married at the time of the birth. On the other hand, after matching, all the  $t$  statistics fell within the non-significant region, indicating

that all the covariates are now balanced. Covariate balance checking for other outcomes is available from the authors.

### The Effects of Breastfeeding Initiation

Table 2 presents the estimated effects of breastfeeding initiation on child outcomes. The left panel is a simple unadjusted between-group comparison. The right panel compares children who were breastfed to their propensity-score-matched controls. For each model, the table presents the effect, the standard error, and the significance of the  $P$  value. Chance findings with five outcomes are possible. Therefore, we only report significance levels at 0.01 or 0.001 level.

First consider the unmatched samples (the first column for each of the outcomes). As we have seen in the descriptive statistics, children who were breastfed, as compared to children who were not, had higher scores on intellectual tests when assessed in the 2002 survey. All these coefficients indicate a beneficial effect of breastfeeding, and they are all statistically significant. The differences, when expressed as effect sizes, are quite large: children who were breastfed had the applied problem scores that were .57 of a standard deviation higher than children who were not breastfed. All of the other scores, with the exception of the WISC, were at least a third of a standard deviation higher for breastfed children. To put these effects in context, these differences are larger than those reported, on the same measures of achievement, between poor and non-poor children [61]. Effect sizes of this magnitude strike us as implausibly large.

On the other hand, in the propensity-score-matched comparison, many of the effects shown in the unadjusted comparison are greatly reduced. For example, the effect on the passage comprehension test score is reduced from 6.83 to 2.18, that for the broad reading score being reduced from 8.26 to 3.58. After adjusting for multiple testing, three out of the five effects remain significant, the effect sizes are smaller and more plausible, with only one effect size (on applied problem) being larger than one-fifth of a standard deviation. All of these effect sizes are considered small under Cohen's rules [52].

### The Effect of Breastfeeding Duration

Table 3 describes the effects of breastfeeding duration. The table presents the mean outcome for each category, the  $P$ -value for comparing each of the other three categories with the reference category (never breastfeeding), and the effect size.

These results show a non-linear relationship between the outcomes and the breastfeeding duration. Compared to children who were never breastfed, children who were breastfed for 1–6 months had significantly higher scores on

**Table 1** Descriptive statistics for the study sample

	All cases			Breastfeeding			No breastfeeding		
	Obs <sup>a</sup>	Mean/ proportion	SD	Obs	Mean/ proportion	SD	Obs	Mean/ proportion	SD
Ever breastfeeding	3271	0.46	0.50	1512	1.00	0.00	1759	0.00	0.00
Length of breastfeeding (in months)	3271	3.7	6.21	1512	7.9	7.07	1759	0.0	0.00
<i>Outcomes (measured in 2002)</i>									
Woodcock Johnson test-revised									
Letter word	2444	103.9	19.0	1136	108.5	18.6	1308	99.9	18.4
Passage comprehension	2361	103.5	17.4	1099	107.2	17.8	1262	100.3	16.4
Applied problem	2436	102.3	17.0	1133	107.2	16.8	1303	98.0	16.0
Broadreading	2452	99.6	27.1	1139	104.1	27.8	1313	95.8	25.9
WISC <sup>b</sup> digit span test	2434	14.5	4.5	1131	14.7	4.6	1303	14.3	4.3
<i>Covariates</i>									
Child (at 1997 survey)									
Female	3271	0.49	0.50	1512	0.50	0.50	1759	0.49	0.50
Age (months)	3271	89.3	46.0	1512	86.5	46.3	1759	91.8	45.5
Number of siblings	3225	1.4	1.2	1499	1.4	1.2	1726	1.4	1.2
First born	3198	0.41	0.49	1477	0.42	0.49	1721	0.39	0.49
Black	3271	0.40	0.49	1512	0.20	0.40	1759	0.57	0.50
White	3271	0.47	0.50	1512	0.62	0.48	1759	0.34	0.47
Other race/ethnicity group	3271	0.13	0.34	1512	0.18	0.38	1759	0.09	0.28
HOME measure <sup>c</sup>	3271	17.8	3.1	1512	18.2	3.1	1759	17.4	3.0
Born preterm	3270	0.08	0.28	1511	0.06	0.24	1759	0.10	0.30
Born small for gestational age	3271	0.15	0.36	1512	0.10	0.30	1759	0.20	0.40
Child's health status at birth <sup>d</sup>	3260	2.21	0.57	1508	2.29	0.59	1752	2.13	0.55
Mother (at 1997 survey)									
WIC participation	3180	0.43	0.50	1442	0.29	0.45	1738	0.55	0.50
Mom's age giving birth (years)-20	3270	7.3	5.8	1512	8.5	5.6	1758	6.4	5.8
IQ	2523	30.7	5.4	1195	32.5	5.0	1328	29.0	5.3
Education									
High-school dropout	3271	0.14	0.35	1512	0.08	0.26	1759	0.20	0.40
High school graduate	3271	0.35	0.48	1512	0.27	0.44	1759	0.42	0.49
College or more	3271	0.38	0.49	1512	0.49	0.50	1759	0.28	0.45
Missing education	3271	0.13	0.33	1512	0.17	0.37	1759	0.09	0.29
Mother (year of child birth)									
Family income (\$)	2751	35176	31671	1223	43695	35789	1528	28357	26014
Married	3271	0.63	0.48	1512	0.77	0.42	1759	0.50	0.50
Employed	2527	0.47	0.50	1150	0.52	0.50	1377	0.43	0.50
Head of household <sup>e</sup>	2603	0.15	0.36	1179	0.08	0.27	1424	0.21	0.41
Self-rated health status <sup>f</sup>	2522	4.1	0.9	1164	4.2	0.9	1358	4.0	1.0
Received medicaid during pregnancy	3175	0.35	0.48	1439	0.24	0.43	1736	0.44	0.50

<sup>a</sup> Number of observations with non-missing values

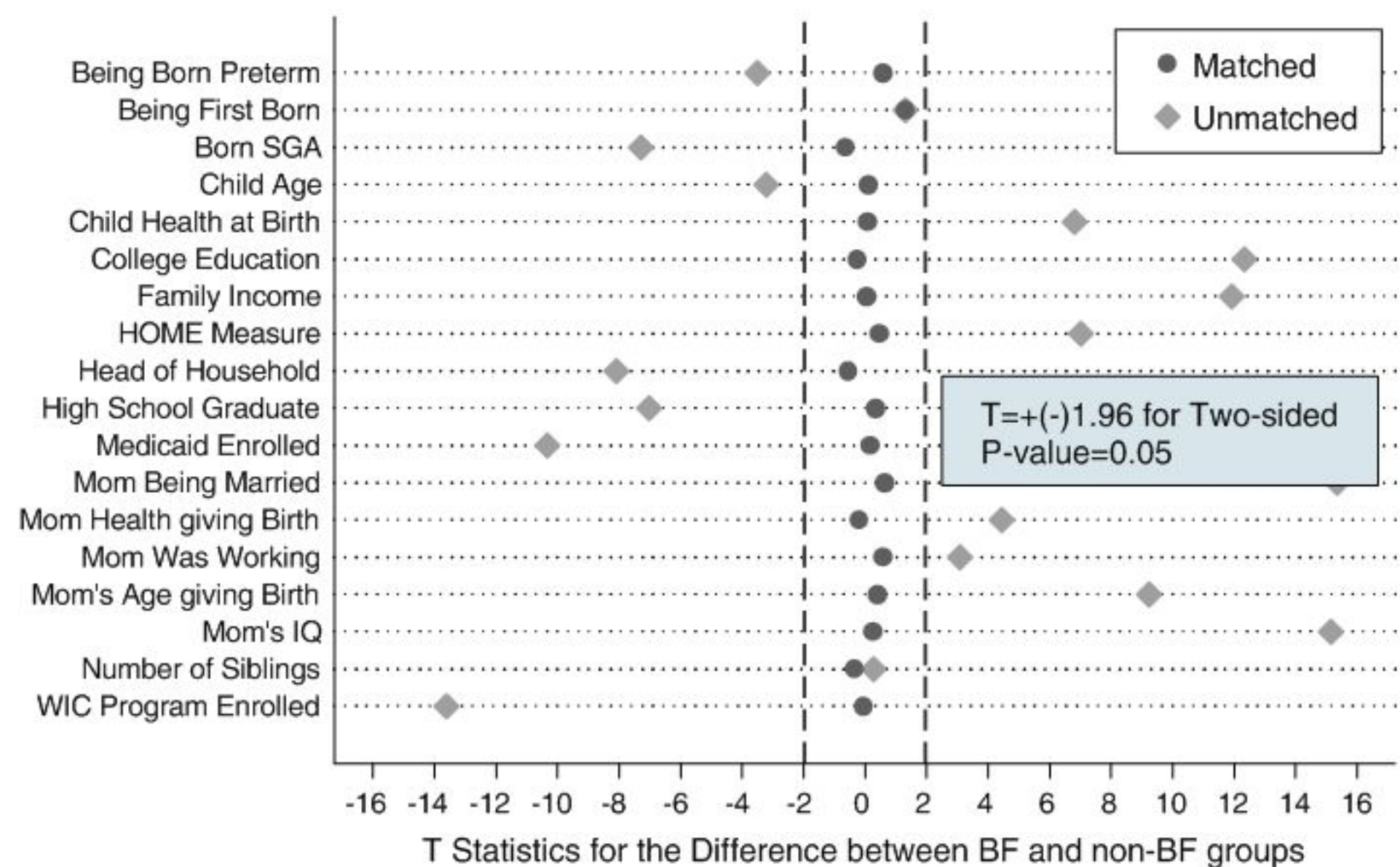
<sup>b</sup> Wechsler Intelligence Scale for Children-Revised

<sup>c</sup> Home Observation for Measurement of the Environment

<sup>d</sup> 1 Worse, 2 Same, 3 Better (as compared with other babies born at birth)

<sup>e</sup> Indicates that mother was the only adult in the household

<sup>f</sup> 1 Poor, 2 Fair, 3 Good, 4 Very good, 5 Excellent

**Fig. 1** Checking covariates balance for letter word test**Table 2** Effects of breastfeeding initiation on outcomes

Outcomes	Unmatched				Propensity score matched with imputations			
	Effect	SE <sup>a</sup>	P <sup>b</sup>	Effect size <sup>c</sup>	Effect	SE <sup>a</sup>	P <sup>b</sup>	Effect size <sup>c</sup>
Woodcock Johnson test-revised								
Letter word	8.59	0.75	***	<b>0.46</b>	3.43	1.05	***	<b>0.18</b>
Passage comprehension	6.83	0.71	***	<b>0.41</b>	2.18	0.97		<b>0.13</b>
Applied problem	9.17	0.66	***	<b>0.57</b>	3.43	0.92	***	<b>0.21</b>
Broadreading	8.26	1.09	***	<b>0.33</b>	3.58	1.52	**	<b>0.14</b>
WISC <sup>d</sup> Digit span test	0.48	0.18	**	<b>0.11</b>	0.50	0.25		<b>0.11</b>

<sup>a</sup> Standard error

<sup>b</sup> P-values refer to the null hypothesis that the mean outcome for breastfeeding is equivalent to that for the “not-breastfeeding” category.  
\* Significant at 5%; \*\* significant at 1%; \*\*\* significant at 0.1%

<sup>c</sup> Effect size represents the difference in means divided by the standard deviation for the “not-breastfeeding” group (the reference group)

<sup>d</sup> Wechsler Intelligence Scale for Children-Revised

the applied problem tests; for the other outcomes or durations, there is no significant effect found. Moreover, for both passage comprehension and broad reading, the children who were never breastfed compare favorably to those who were breastfed for more than 12 months. For example, for children who were not breastfed had average passage comprehension scores of 98.1; on the same measure, children who were breastfed for more than a year had average scores of 96.7.

These results are meaningful because they suggest that longer spells of breastfeeding are uncorrelated with increases in the measures of achievement. Therefore, for breastfeeding to be causally related to higher cognitive scores, it must be the case that the cognitive benefits of breastfeeding arise only if the child is breastfed within the first 6 months of life. While this explanation is plausible—there may be a window of opportunity when breastfeeding

can do the most good—these results are also consistent with an explanation in which positive selection into breastfeeding explains the higher scores.

### Sensitivity Analysis

We address several limitations of our analyses and methodology. First, bias is possible because of the long lag between birth and the survey. Such measurement error involving breastfeeding would bias estimates downward [62]. We limit our analysis of the effect of breastfeeding initiation to children who aged 7 or younger when the mother was interviewed in 1997. Two outcomes, the letter word and the applied problem test scores remain significant after the matching. But the effect sizes (0.24 and 0.30, respectively) are still modest. This result is available from the first author.

**Table 3** Effects of breastfeeding duration analysis

Outcomes		Propensity score adjusted with multiple imputations			
		Never	<6 months	7–12 months	12 + months
	Sample size	1759	895	349	268
Woodcock Johnson test-revised					
Letter word	Mean	102.67	105.04	103.90	105.01
	Effect size <sup>a</sup>		0.13	0.07	0.13
Passage comprehension	Mean	98.07	99.18	100.67	96.66
	Effect size		0.04	0.09	−0.05
Applied problem	Mean	100.81	103.64 <sup>***c</sup>	103.06	102.2
	Effect size		0.18	0.14	3 0.09
Broad reading	Mean	99.06	100.36	101.22	98.88
	Effect size		0.05	0.08	−0.01
WISC <sup>b</sup> digit span test	Mean	14.40	14.30	14.63	14.42
	Effect size		0.02	0.05	0.00

<sup>a</sup> Effect size represents the difference in means divided by the standard deviation for the “never” group

<sup>b</sup> Wechsler Intelligence Scale for Children-Revised

<sup>c</sup> *P*-values refer to the null hypothesis that the mean for this level of breastfeeding is equivalent to that for the “never” category. \* Significant at 5%; \*\* significant at 1%; \*\*\* significant at 0.1%

Second, we use alternative cutoffs for breastfeeding durations. For example, we divide the duration into less than 3 months, 3–12 months, and more than 12 months, and found a significant increase in the applied problem scores for the less-than-3-month category. But none of the other categories shows statistically significant effects. The relationship across duration is still non-linear, and all the effect sizes are less than 0.20. This result is available from the first author.

Third, both regression and the propensity score matching assume that unobserved confounding does not exist [50]. We examine the potential effects of such confounding if they were to exist in a regression framework. Suppose the true relationship between the child cognition and breastfeeding is as follows:

$$Y_i = \alpha + \beta E_i + \gamma H_i + \varepsilon_i$$

$Y_i$  is the child’s cognitive score.  $E_i$  is the child’s breastfeeding status.  $H$  is an unobserved confounder beyond all the covariates we have included in the analysis.  $\beta$  is the (true) effect of breastfeeding on the cognitive score. The resulting estimate of  $\beta$  ( $\hat{\beta}$ ) is the sum of the true treatment effect and some effect caused by the unobserved confounder:

$$\hat{b} = \beta + \gamma \frac{\delta H_i}{\delta E_i} = \beta + \frac{\gamma}{\lambda}$$

The bias depends on the relationship between  $H$  and the outcome ( $\gamma$ ) and between  $H$  and the treatment ( $1/\lambda$ ). Note that this exercise does not involve actually generating  $H$ . The intent of the analysis is to quantify a qualitative

concern, “how large would any remaining bias have to be to change the study’s key figures?”

Following the literature, we assume  $H$  is dichotomous, and for convenience it indicates whether people are healthy or not (over and above the covariates we have included). For further convenience, we assume this unobserved confounder predicts the treatment perfectly. In our case, all healthy women would breastfeed their children while all unhealthy women would not. In essence,  $P(E = 1 | H = 1) = P(E = 0 | H = 0) = 1$ , and this effectively sets  $\lambda$  to 1. Then the remaining bias  $\gamma$  is strictly a function of the link between  $H$  and the outcome, and the effect of this bias is expressed in the same metric as the effect of the exposure/treatment.

How big would  $\gamma$  have to be to change the interpretation of our findings? In our case, the key issue is whether the bias would make a relationship appear significant when it is not. In effect, the question is whether the bias pulled the observed *t*-statistic more than  $|t| - 1.96$  standard errors away from zero. For a positive true effect, for example,

$$\gamma_{\text{critical}} = \hat{b} - 1.96 \times \hat{se}$$

$\gamma_{\text{critical}}$  is the smallest value of gamma to produce an estimate that is just statistically significant at 0.05 level (while in fact the true effect is not significant).

One can see that the size of this bias depends on both the magnitude of the estimated effect as well as its precision. A larger, more precise effect will be more robust to possible unobserved confounding.

Using the standard normal distribution and the estimated standard error, we can calculate the critical value of gamma

above. For example, in the propensity-score-matched sample, the estimated effect of breastfeeding initiation on child's letter word test is 3.43 with a standard error of 1.05 (see Table 2). The critical value of gamma for a significant effect with the above standard error is  $3.43 - 1.96 \times 1.05 = 1.38$ . So the unobserved confounder would have to have an absolute effect of 1.38 or larger to change the statistical significance of our result. This estimate is about 40% of the observed effect and suggests that the unobserved confounding would have to be rather large. Note that this effect has to exist over and above the balancing of all the observed characteristics. Though there is no way to know for sure (as with any sensitivity analysis), it appears our finding is rather robust to unobserved confounding. On the other hand, the magnitude of bias for the WISC test needs to be as small as 0.01 (or 1% of the observed effect) to flip the significance of the result. In general, these findings suggest that the two strongest relationships in the table are unlikely to be caused by modest confounding. (The full set of estimates is available from the first author.)

## Discussion

Previous research has indicated that in unadjusted comparisons with children who are formula fed, children who are breastfed score higher on measures of cognitive functioning [16, 17, 63, 64]. The reasons for this difference, however, are unclear. Breast milk contains nutrients that promote optimum cognitive development; alternatively, the benefits could reflect not breastfeeding but the characteristics of the mothers who breastfeed. The distinction is the key: in the latter case, promoting breastfeeding will not confer those advantages to the women and their children.

Our analyses suggest that adjusting for observed confounding substantially reduces the apparent benefits of breastfeeding initiation. Although the effects remain statistically significant for at least two of the outcomes, the adjusted difference between breastfed and non-breastfed children translates into an effect size between one-tenth and one-fifth of a standard deviation, an effect which is small according to Cohen's effect size rules.

More tellingly, we did not find that children who were breastfed for longer periods of time had higher cognitive scores than children who were never breastfed. In the duration analyses, positive effects on the outcomes were found for children who were breastfed less than six months when compared to children who were not breastfed. But no beneficial effect of breastfeeding was found when a child was breastfed for more than six months. Moreover, children who were never breastfed compare favorably on some outcomes to those who were breastfed for more than 12 months (e.g., passage comprehension and broad reading

test). These results suggest that there may be some critical window (e.g., the first six months of life) in early infant development during which breastfeeding can affect cognitive outcomes; after that window has closed, breastfeeding is not associated with better cognitive scores. This pattern also hints at unobserved differences—perhaps the women who start but then discontinue breastfeeding are disadvantaged in ways not captured here.

Our findings do suffer from limitations. First, data on breastfeeding practice are based on maternal recall. We address this issue in several ways. First, we compare the prevalence of breastfeeding initiation in this sample to what was reported in the National Health and Nutrition Examination Survey during the same period (NHANES 1988–1994) [65] and found them comparable. For example, the overall breastfeeding prevalence, that for non-Hispanic white, and that for non-Hispanic black was 53.6, 60.2, and 26.3% for NHANES sample. [66] The corresponding prevalence was 46, 61.0, and 22.7% for the PSID sample. The amount of breastfeeding was also close in the two samples. Overall, 22.4 and 8.9% of the NHANES III sample were still breastfeeding at 6 and 12 months [66]. Those for the PSID sample were 23.9 and 10.9%, respectively. The prevalence for premature and full-term babies was similar among the two samples, too. This shows two things. First, our sample was compatible with what might be expected for a nationally representative sample. Second, these comparisons suggest that recall problems may be rather small. This consistency with other figures is reassuring.

In addition, when we restrict our sample to younger children in the supplemental analysis, we find similar modest effect. Furthermore, for maternal recall to influence the estimated effect of breastfeeding, the relationship between actual and reported breastfeeding practice would have to be moderated by the outcomes of interest. We cannot rule out selective recall bias, but it is hard to imagine how that might be related to the child outcomes.

Another limitation is that our data lack a measure of breastfeeding exclusivity. According to NHANES III data, 9.5% of the children were still exclusively breastfeeding at the age of 6 months [66]. Other studies focus on the period during which breast milk is not supplemented with other, solid food [67, 68]. The effects presented here may mask larger benefits for the sub-group of participants who breastfeed exclusively. However, as an empirical matter, the impact of exclusivity is difficult to assess in observational data. It is reasonable to assume that those who breastfeed exclusively are not a representative sub-sample of those who breastfeed at all; in that case, the potential for confounding is likely even greater.

One can bound the potential effect of exclusivity by thinking of it as a form of compliance. Suppose that non-exclusive breastfeeding has no effect whatsoever. Roughly

40% of babies that are ever breastfed are breastfed through the first three months of life [69]. In that case, the effect we find here should be multiplied by 2.5 (or 1/.40). Multiplying a small estimate still produces a rather modest effect [70]. (This effect is the so-called compliance-adjusted effect.) Note that as we shrink the size of the group for which breastfeeding may be effective (and thereby inflate the effect of breastfeeding), the public health implications are diminished.

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 breastfeeding research. A body of research considers the circumstances under which observational analyses produces findings similar to a randomized experiment [71, 72]. That literature suggests that a well done observational study can produce findings quite similar to those of a randomized trial (where both exist). In terms of addressing selection bias, regression has limitations in their underlying assumptions and the number of variables they can match. Propensity score matching relaxes the linear assumption of the regression and simplifies the matching variables into a single predicted probability.

We do have a key assumption underlying the propensity score model, that is, there is no unmeasured confounder. However, we include a wide range of covariates and at least by the standard or prior research, our list is comprehensive. In addition, we estimate how large the potential bias due to unobserved confounding has to be to change the statistical significance of our results. This sensitivity analysis reveals that for some outcomes, such as the letter word and the applied problem tests, our results are rather robust; while for other outcomes like the passage comprehension and the WISC tests, our results are sensitive. Note that for the two outcomes that are sensitive, they are only significant at  $P = 0.05$  level (instead of at 0.01 or 0.001 level). This links back to the issue of multiple testing. When using adjustment such as Bonferroni correction, these two outcomes are no longer statistically significant. Therefore, the effects of these two tests could indeed be explained by some unobserved confounding.

Do these findings mean that breastfeeding should not be encouraged? Of course not. The health benefits of breastfeeding for both the child and the mother have been well demonstrated [1]. On the other hand, women who did not breastfeed their babies (either because they cannot or they do not want to) should not be made to feel that they are damaging the cognitive potential of their children. Such guilt may be more damaging to the child, especially if it interferes with attachment and/or contributes to maternal depression [73, 74]. Nor should the benefits of breastfeeding be overstated, given the barriers to breastfeeding

[75], which remain formidable in the workplace and other contexts.

Furthermore, from a broader public health perspective, the practical importance of breastfeeding on cognitive development may be small. Other factors, such as poverty, low birth weight, or maternal risky behaviors, may be much more important in determining a child's intellectual well-being [61, 76, 77]. If society wishes to increase the cognitive outcomes of at-risk children, it may be more sensible to direct attention to factors that have a larger impact on child outcomes.

**Acknowledgments** This study was not funded by any governmental or non-governmental agency.

## References

1. American Academy of Pediatrics. (2005). Breastfeeding and the use of human milk. *Pediatrics*, *115*(2), 496–507.
2. Uauy, R., & Peirano, P. (1999). Breast is best: Human milk is the optimal food for brain development. *American Journal of Clinical Nutrition*, *70*, 433–434.
3. Oddy, W. H., Kendall, G. E., Blair, E., De Klerk, N. H., Stanley, F. J., Landau, L. I., et al. (2003). Breast feeding and cognitive development in childhood: A prospective birth cohort study. *Paediatric and Perinatal Epidemiology*, *17*(1), 81–90.
4. Angelsen, N. K., Vik, T., Jacobsen, G., & Bakketeig, L. S. (2001). Breast feeding and cognitive development at age 1 and 5 years. *Archives of Disease in Childhood*, *85*(3), 183–188.
5. Wigg, N. R., Tong, S., McMichael, A. J., Baghurst, P. A., Vimpani, G., & Roberts, R. (1998). Does breastfeeding at six months predict cognitive development? *Australian and New Zealand Journal of Public Health*, *22*(2), 232–236.
6. Lucas, A., Morley, R., Cole, T. J., Lister, G., & Leeson-Payne, C. (1992). Breast milk and subsequent intelligence quotient in children born preterm. *Lancet (British edition)*, *339*(8788), 261–264.
7. Gomez-Sanchiz, M., Canete, R., Rodero, I., Baeza, J. E., & Gonzalez, J. A. (2004). Influence of breast-feeding and parental intelligence on cognitive development in the 24-month-old child. *Clinical Pediatrics*, *43*(8), 753–761.
8. Carlson, S. E. (1999). Long-chain polyunsaturated fatty acids and development of human infants. *Acta Paediatrica. Supplement*, *88*(430), 72–77.
9. Rey, J. (2003). Breastfeeding and cognitive development. *Acta Paediatrica*, *92*(8 supp 442), 11–18.
10. Koletzko, B., Agostoni, C., Carlson, S. E., Clandinin, T., Hornstra, G., Neuringer, M., et al. (2001). Long chain polyunsaturated fatty acids (LC-PUFA) and perinatal development. *Acta Paediatrica*, *90*(4), 460–464.
11. Lauritzen, L., Hansen, H. S., Jorgensen, M. H., & Michaelsen, K. F. (2001). The essentiality of long chain n-3 fatty acids in relation to development and function of the brain and retina. *Progress in Lipid Research*, *40*(1–2), 1–94.
12. McCann, J. C., & Ames, B. N. (2005). Is docosahexaenoic acid, an n-3 long-chain polyunsaturated fatty acid, required for development of normal brain function? An overview of evidence from cognitive and behavioral tests in humans and animals. *American Journal of Clinical Nutrition*, *82*(2), 281–295.
13. Lucas, A., Stafford, M., Morley, R., Abbott, R., Stephenson, T., MacFadyen, U., et al. (1999). Efficacy and safety of long-chain

- polyunsaturated fatty acid supplementation of infant-formula milk: A randomised trial. *Lancet*, 354(91914), 1948–1954.
14. Gustafsson, P. A., Duchon, K., Birberg, U., & Karlsson, T. (2004). Breastfeeding, very long polyunsaturated fatty acids (PUFA) and IQ at 6 1/2 years of age. *Acta Paediatrica*, 93(10), 1280–1287.
  15. Gomez-Sanchiz, M., Canete, R., Rodero, I., Baeza, J. E., & Avila, O. (2003). Influence of breast-feeding on mental and psychomotor development. *Clinical Pediatrics*, 42(1), 35–42.
  16. Rao, M. R., Hediger, M. L., Levine, R. J., Naficy, A. B., & Vik, T. (2002). Effect of breastfeeding on cognitive development of infants born small for gestational age. *Acta Paediatrica*, 91(3), 267–274.
  17. Quinn, P., O'Callaghan, M., GM, W., Najman, J., Andersen, M., & Bor, W. (2001). The effect of breastfeeding on child development at 5 years: A cohort study. *Journal of Pediatrics & Child Health*, 37(5), 465–469.
  18. Oddy, W. H., Kendall, G. E., Blair, E., De Klerk, N. H., Stanley, F. J., Landau, L. I., et al. (2003). Breast feeding and cognitive development in childhood: A prospective birth cohort study. *Pediatric and Perinatal Epidemiology*, 17(1), 81–90.
  19. Horwood, L. J., Darlow, B. A., & Mogridge, N. (2001). Breast milk feeding and cognitive ability at 7–8 years. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 84, f23–f27.
  20. Johnson, D. L., Swank, P. R., Howie, V. M., Baldwin, C. D., & Owen, M. (1996). Breast feeding and children's intelligence. *Psychological Reports*, 79(3 Pt 2), 1179–1185.
  21. Anderson, J. W., Johnstone, B. M., & Remley, D. T. (1999). Breast-feeding and cognitive development: A meta-analysis. *American Journal of Clinical Nutrition*, 70(4), 525–535.
  22. Reynolds, A. (2001). Breastfeeding and brain development. *Pediatric Clinics of North America*, 48(1), 159–171.
  23. Drane, D. L., & Logemann, J. A. (2000). A critical evaluation of the evidence on the association between type of infant feeding and cognitive development. *Pediatric and Perinatal Epidemiology*, 14(4), 349–356.
  24. Jacobson, S. W., Chiodo, L. M., & Jacobson, J. L. (1999). Breastfeeding effects on intelligence quotient in 4- and 11-year-old children. *Pediatrics*, 103, e71.
  25. Der, G., Batty, G. D., & Deary, I. J. (2006). Effect of breast feeding on intelligence in children: Prospective study, sibling pairs analysis, and meta-analysis. *British Medical Journal*, 333(7575), 945–950.
  26. Jain, A., Concato, J., & Leventhal, J. M. (2002). How good is the evidence linking breastfeeding and intelligence? *Pediatrics*, 109(6), 1044–1053.
  27. Greene, L. C., Lucas, A., Livingstone, M. B., Harland, P. S., & Baker, B. A. (1995). Relationship between early diet and subsequent cognitive performance during adolescence. *Biochemical Society Transactions*, 23(2), 376S.
  28. Horwood, L. J., & Fergusson, D. M. (1998). Breastfeeding and later cognitive and academic outcomes. *Pediatrics*, 101(1), e9.
  29. Mortensen, E., Michaelsen, K., Sanders, S., & Reinisch, J. (2002). The association between duration of breastfeeding and adult intelligence. *The Journal of the American Medical Association*, 287(18), 2365–2371.
  30. Rogan, W. J., & Gladen, B. C. (1993). Breast-feeding and cognitive development. *Early Human Development*, 31(3), 181–193.
  31. Smith, M. M., Durkin, M., Hinton, V. J., Bellinger, D., & Kuhn, L. (2003). Influence of breastfeeding on cognitive outcomes at age 6–8 years: Follow-up of very low birth weight infants. *American Journal of Epidemiology*, 158(11), 1075–1082.
  32. Hernan, M. A., Hernandez-Diaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology*, 15(5), 615–625.
  33. Greenland, S. (2003). Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology*, 14(3), 300–306.
  34. Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, U.K.: Cambridge University Press.
  35. McGonagle, K. A., & Schoeni, R. F. (2006). The panel study of income dynamics: overview & summary of scientific contributions after nearly 40 years. Retrieved May 21, 2007, from <http://www.psidonline.isr.umich.edu/Publications/Papers/montreal.pdf>.
  36. Duncan, G. J., & Brooks-Gunn, J. (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.
  37. PSID Data Center. CDS frequently asked questions. Retrieved May 21, 2007, from <http://psidonline.isr.umich.edu/CDS/faq.aspx>.
  38. Woodcock, R. W. (1989). *Woodcock-Johnson tests of achievement-Revised*. Allen, TX: DLM Teaching Resources.
  39. Mainieri, T. (2006). The panel study of income dynamics child development supplement: User guide for CDS-II. Retrieved December 5, 2009, from [http://psidonline.isr.umich.edu/CDS/cdsii\\_userGd.pdf](http://psidonline.isr.umich.edu/CDS/cdsii_userGd.pdf).
  40. Wechsler, D. (1974). *Wechsler intelligence scale for children-revised*. New York, NY: The Psychological Corporation.
  41. Alexander, G. R., Himes, J. H., Kaufman, R. B., Mor, J., & Kogan, M. (1996). A United States national reference for fetal growth. *Obstetrics and Gynecology*, 87(2), 163–168.
  42. Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the environment*. Little Rock: University of Arkansas.
  43. Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
  44. Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, 65(2), 261–294.
  45. Blundell, R., Dearden, L., & Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3), 473–512.
  46. Lee, M. (2005). *Micro-econometrics for policy, program, and treatment effects*. New York: Oxford University Press.
  47. Rosenbaum, P. R. (2002). *Observational studies*. New York: Springer.
  48. Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25(1), 659–707.
  49. Gibson-Davis, C. M., & Brooks-Gunn, J. (2006). Breastfeeding and verbal ability of 3-year-olds in a multicity sample. *Pediatrics*, 118(5), e1444–e1451.
  50. Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge University Press.
  51. Heckman, J. J., & Navarro-Lozano, S. (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics*, 86(1), 30–57.
  52. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
  53. Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706–710.
  54. Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82(398), 387–394.
  55. Horvitz, D. G., & Thompson, D. J. (1952). A Generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663–685.

56. Veroff, J., McClelland, L., & Marquis, K. (1971). *Measuring intelligence and achievement motivation in surveys: Final report to HEW*. OEO.
57. Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
58. Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Version 3.0.0. Retrieved April 7, 2009, from <http://ideas.repec.org/c/boc/bocode/s432001.html>.
59. Royston, P. (2005). Multiple imputation of missing values: Update of ICE. *Stata Journal*, 5(4), 527–536.
60. Stata statistical software: Release 10 [program]. (2007). College Station, TX.
61. Brooks-Gunn, J., Duncan, G. J., & Maritato, N. (1997). Poor families, poor outcomes: The well-being of children and youth. In J. Brooks-Gunn & G. J. Duncan (Eds.), *The consequences of growing up poor* (pp. 1–17). New York: Russell Sage.
62. Greene, W. H. (2008). *Econometric analysis* (6th ed.). Upper Saddle River, N.J.: Prentice Hall.
63. Morrow-Tlucak, M., Haude, R. H., & Ernhart, C. B. (1988). Breastfeeding and cognitive development in the first 2 years of life. *Social Science and Medicine*, 26(6), 635–639.
64. Rey, J. (2003). Breastfeeding and cognitive development. *Acta Pediatric Supplement*, 442(Suppl), 11–18.
65. National Center for Health Statistics. (1996). *NHANES III reference manual and reports [CD-ROM]*. Hyattsville, MD: Centers for Disease Control and Prevention.
66. Li, R., Ogden, C., Ballew, C., Gillespie, C., & Grummer-Strawn, L. (2002). Prevalence of exclusive breastfeeding among US infants: The Third National Health and Nutrition Examination Survey (phase II, 1991–1994). *American Journal of Public Health*, 92(7), 1107.
67. Gdalevich, M., Mimouni, D., David, M., & Mimouni, M. (2001). Breast-feeding and the onset of atopic dermatitis in childhood: A systematic review and meta-analysis of prospective studies. *Journal of the American Academy of Dermatology*, 45(4), 520–527.
68. Oddy, W. H., Holt, P. G., Sly, P. D., Read, A. W., Landau, L. I., Stanley, F. J., et al. (1999). Association between breast feeding and asthma in 6 year old children: Findings of a prospective birth cohort study. *British Medical Journal*, 319(7213), 815–819.
69. Centers for Disease Control and Prevention. (2004). Any and exclusive breastfeeding rates by age among children born in 2004. Retrieved January 3, 2008, from [http://www.cdc.gov/breastfeeding/data/nis\\_data/data\\_2004.htm](http://www.cdc.gov/breastfeeding/data/nis_data/data_2004.htm).
70. Manski, C. F. (2007). *Identification for prediction and decision*. Cambridge, Mass: Harvard University Press.
71. Cook, T. D., Shadish, W. R., & Wong, V. C. (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), 724–750.
72. Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2009). The importance of covariate selection in controlling for selection bias in observational studies.
73. Shakespeare, J., Blake, F., & Garcia, J. (2004). Breast-feeding difficulties experienced by women taking part in a qualitative interview study of postnatal depression. *Midwifery*, 20(3), 251–260.
74. Sibolboro Mezzacappa, E., & Endicott, J. (2007). Parity mediates the association between infant feeding method and maternal depressive symptoms in the postpartum. *Archives of Women's Mental Health*, 10(6), 259–266.
75. Carson, C. (2004). Breastfeeding: Barriers and breakthroughs. *RCM Midwives Journal*, 7(5), 192.
76. Julvez, J., Ribas-Fito, N., Torrent, M., Forns, M., Garcia-Esteban, R., & Sunyer, J. (2007). Maternal smoking habits and cognitive development of children at age 4 years in a population-based birth cohort. *International Journal of Epidemiology*, 36(4), 825–832.
77. Hack, M., Flannery, D. J., Schluchter, M., Cartar, L., Borawski, E., & Klein, N. (2002). Outcomes in young adulthood for very-low-birth-weight infants. *The New England Journal of Medicine*, 346(3), 149.