Social Science & Medicine 109 (2014) 55–65

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Is breast truly best? Estimating the effects of breastfeeding on long-term child health and wellbeing in the United States using sibling comparisons

Cynthia G. Colen*, David M. Ramey

Department of Sociology, Ohio State University, United States

ARTICLE INFO

Article history:
Available online 29 January 2014

Keywords:
Breastfeeding
Child health
Race
Socioeconomic status
Life course epidemiology
Sibling comparisons

ABSTRACT

Breastfeeding rates in the U.S. are socially patterned. Previous research has documented startling racial and socioeconomic disparities in infant feeding practices. However, much of the empirical evidence regarding the effects of breastfeeding on long-term child health and wellbeing does not adequately address the high degree of selection into breastfeeding. To address this important shortcoming, we employ sibling comparisons in conjunction with 25 years of panel data from the National Longitudinal Survey of Youth (NLSY) to approximate a natural experiment and more accurately estimate what a particular child's outcome would be if he/she had been differently fed during infancy. Results from standard multiple regression models suggest that children aged 4 to 14 who were breast- as opposed to bottle-fed did significantly better on 10 of the 11 outcomes studied. Once we restrict analyses to siblings and incorporate within-family fixed effects, estimates of the association between breastfeeding and all but one indicator of child health and wellbeing dramatically decrease and fail to maintain statistical significance. Our results suggest that much of the beneficial long-term effects typically attributed to breastfeeding, per se, may primarily be due to selection pressures into infant feeding practices along key demographic characteristics such as race and socioeconomic status.

Introduction

In 2012, the American Academy of Pediatrics (AAP) updated its original policy statement concerning breastfeeding and summarized findings from a substantial body of research to provide evidence for "diverse and compelling advantages for infants, mothers, families, and society from breastfeeding and use of human milk for infant feeding" (AAP, 2012). Similarly, *Healthy People 2020*, which provides empirically based population health objectives to improve wellbeing for all Americans, has taken an emphatic stance on infant feeding practices by declaring breastfeeding a national priority (U.S. Department of Health and Human Services, 2010). Between 2000 and 2009, the proportion of U.S. infants who were still being breastfed at six months increased from 34% to 47% (Centers for Disease Control and Prevention [CDC], 2013). It is now commonplace for expectant mothers to be counseled that "breast is best" for their infant.

Targeted policies have been initiated at both the national and local level to promote breastfeeding (AAP, 2012; Farley, 2012). Health officials hope to increase the proportion of U.S. mothers who breastfeed at all from 74% to 82% and who continue breastfeeding at 6 months from 44% to 61% (U.S. Department of Health and Human Services, 2010). Moreover, medical professionals and public health advocates are not simply recommending that new mothers breastfeed their infants. Rather, they are emphasizing the perceived benefits of exclusive breastfeeding and hope to ensure that babies receive only human milk during the first six months of life (AAP, 2012; World Health Organization [WHO], 2013).

Clearly, these recommendations are meant to promote the health and wellbeing of both mothers and their newborns. Besides being the most economical choice, it is thought that human breast milk offers infants the most nutrient rich, easily digestible form of nourishment that will contribute to beneficial outcomes during the perinatal period, throughout childhood, and possibly into adulthood (Ip et al., 2007; U.S. Surgeon General, 2011; WHO, 2013).

Breastfeeding is thought to affect child outcomes due to superior nutrients unique to breast milk that are absent from infant formula as well as the biochemical reactions triggered by the act of breastfeeding, itself. For example, breast milk contains enzymes, hormones, growth factors, and immunologic substances that assist in creating an effective host defense to infectious agents (Guilbert, et al, 2007). These cellular attributes are particularly helpful in

* Corresponding author. 217 Townshend Hall, 1885 Neil Avenue Mall, Columbus, Ohio 43210, United States.
E-mail address: colen.3@osu.edu (C.G. Colen).

0277-9536/$ – see front matter
http://dx.doi.org/10.1016/j.socscimed.2014.01.027

combating respiratory infections in infancy and may prevent the subsequent development of asthma and allergies (Oddy, 2004). Concerning obesity as an endpoint, the causal pathway is likely to follow two distinct mechanisms, the first of which concerns the ability of breastfed infants to more quickly and easily recognize feelings of satiety and the second of which is related to specific nutrient combinations that may influence insulin resistance and/or metabolic responses (Gillman & Mantzoros, 2007). Finally, breast milk contains long-chain polyunsaturated fatty acids that play an essential role in normal retinal and neural development (Innis, Gilley, & Werker, 2001; Rey, 2003) and might be implicated in later cognitive functioning (McCann & Ames, 2005).

That the benefits of breastfeeding are sufficiently large and long-term to support such an intense policy commitment to universalizing the behavior is assumed, but deserves systematic study. Total commitment to 6 months of exclusive breastfeeding is a very high expectation of mothers, especially in an era when a majority of women work outside the home, often in jobs with little flexibility and limited maternity leave, and in a country that offers few family policies to support newborns or their mothers (Guendelman et al, 2009; Rippeyoung & Noonan, 2012). The line between providing information about the benefits of breastfeeding and stigmatizing mothers facing structured, valid, and often difficult trade-offs in the care and financial support of their children or in fulfilling their own human potential must be drawn sensitively.

Currently, breastfeeding rates in the U.S. are socially patterned. Previous research has documented startling racial and socioeconomic disparities in infant feeding practices. Data from the 2008 National Immunization Survey reveal that 75% of White infants but only 59% of Black infants were ever breastfed. Similarly, 47% of White infants but only 30% of Black infants were still being breastfed at six months (CDC, 2013). With regard to differences in infant feeding practices according to socioeconomic status, 74% of children whose family incomes were above 185% of the federal poverty threshold but only 57% of children whose family incomes were equivalent to or fell below this threshold had ever been breastfed (Forste and Hoffman, 2008). Furthermore, mothers who completed some high school, were high school graduates, or attended some college were 64%, 60% and 39%, respectively, less likely to initiate breastfeeding than mothers who graduated from college (CDC, 2013).

The social patterning of breastfeeding has important social and scientific implications. Socially, if breastfeeding were as advantageous in both the short- and long- term as is often assumed, one would not want black or poor children to be disproportionately deprived of its benefits. (Whether current approaches to breastfeeding promotion are the best ones is another question beyond the scope of this paper.) Scientifically, disparities in infant feeding practices raise the critical question of the degree to which unobserved heterogeneity between children who were breastfed and those who were not may be driving the frequently noted positive association between breastfeeding and a wide variety of childhood outcomes. If this is the case, a well-intentioned, narrow emphasis on breastfeeding promotion would, at best, fail to realize positive benefits and, at worst, be a source of oppression for women who do not nor cannot breastfeed.

Much of the empirical evidence regarding the effects of infant feeding practices does not adequately address the high degree of selection into breastfeeding. In particular, it must be viewed as inconclusive with regard to conditions that emerge later in the life course -for example, among school-age children or teenagers as opposed to infants – since, of necessity, it often relies on observational, and in many cases cross-sectional, data and study designs that are unable to account for unobserved heterogeneity between breast- and bottle-fed children that are likely to be driving observed differences in health and developmental trajectories. Given the greater likelihood of breastfeeding among socially and economically advantaged groups in the U.S. (Singh, Kogan, & Dee, 2007) and the extent to which race/ethnicity and socioeconomic position is known to influence childhood health and wellbeing (Currie, 2009; Mehta, Lee, & Ylitalo, 2013), these findings are likely to exaggerate the benefits of breastfeeding, per se. The current study was designed to address this possibility.

We examine eleven different outcomes – body mass index, obesity, asthma, hyperactivity, parental attachment, behavioral compliance, reading comprehension, vocabulary recognition, math ability, memory based intelligence, and scholastic competence. In order to separate the impact of factors that predict selection into breastfeeding from the "true" consequences of breastfeeding, we employ sibling comparisons to approximate a natural experiment and more accurately estimate the counterfactual question, "What would this particular child's outcomes be if she/he had been breastfed instead of bottle-fed?" Once between-family differences are taken into account, we find relatively little empirical evidence to support the notion that breastfeeding results in improved health and wellbeing for children between 4 and 14 years of age.

Breastfeeding and childhood health and wellbeing: current evidence

At first glance, the extant literature concerning the association between breastfeeding and long-term child health and wellbeing seems to be straightforward. Previous studies suggest that breastfed children are significantly less likely than their bottle-fed counterparts to be classified as obese (Arenz, Rucker, Koletzko, & von Kries, 2004; Armstrong & Reilly, 2002; Harder, Bergman, Kallischnigg, & Plageman,, 2005; Weden, Brownell, and Rendall, 2012); develop asthma (Oddy, 2004); and be diagnosed with autoimmune diseases, such as Type I diabetes, (Young et al, 2002) and childhood cancers (Martin et al., 2005). Moreover, infant feeding practices appear to be associated with cognitive ability during childhood, such that full-term infants who are breast- as opposed to bottle-fed score 3–6 points higher on IQ tests (Quigley, Hockley, & Carson, 2011). However, upon more recent and rigorous evaluation, these findings appear less conclusive.

A more detailed examination of existing epidemiological studies regarding the effects of breastfeeding on subsequent child health and development reveals more questions than it does answers. Results often vary depending on the study sample employed, the age at which outcomes were measured, whether breastfeeding was defined dichotomously or in terms of duration, and which potential confounders were included in statistical analyses (Baker & Milligan, 2008; Der, Batty, & Deary, 2006; Evenhouse & Reilly, 2005; Metzger & McDade, 2010; Nelson, Gordon-Larsen, & Adair, 2005). The most problematic aspect of this literature is the extent to which children are selected into breastfeeding based upon several sociodemographic dimensions that are simultaneously associated with infant feeding practices and long-term child outcomes. Compared to bottle-fed infants, breastfed infants are significantly more likely to be white, be born into families with above average incomes, have parents with advanced educational attainment, maintain easier access to health care services, and live in safer neighborhoods with lower levels of environmental toxins (Singh et al., 2007; Van Rossen et al., 2009). Thus, comparisons of breast- and bottle-fed infants are likely to be biased by both observed and unobserved heterogeneity, of which the latter poses a greater risk when trying to assess the "true" effects of breastfeeding on subsequent childhood outcomes since these characteristics cannot be taken into account by traditional statistical approaches (ie. OLS or logit regression models).

Findings from the few existing studies that have more rigorously controlled for selection into breastfeeding are not conclusive but suggest that the beneficial effects of breastfeeding on child outcomes may be overstated (Baker & Milligan, 2008; Carlsen, Jacobsen, & Vanky, 2010; Der et al., 2006; Evenhouse & Reilly, 2005; Kramer, 2010). Kramer et al. (2001) conducted a hospital- and clinic-based randomized experiment designed to promote exclusive breastfeeding among Belarusian women. While this intervention resulted in a significantly higher prevalence of breastfeeding among the treatment as opposed to the control group, there was no reduction in the risk of asthma or allergy (Kramer, Matush, et al, 2007), adiposity (Kramer, Matush, et al, 2007), dental caries (Kramer, Vanilovich, et al, 2007), or problem behavior (Kramer et al., 2008) among 6 year-olds.

Most investigators, however, are not able to conduct randomized controlled trials. Relying on natural experiments is often times the next best approach to reducing selection bias. The few existing studies that exploit sibling comparisons to more accurately estimate the impact of breastfeeding on long-term childhood health and wellbeing, for the most part, focus solely on obesity and reveal inconsistent findings (Gillman et al., 2006; Metzger & McDade, 2010; Nelson et al., 2005; O'Tierney et al., 2009). There are two notable exceptions. In a subset of their original analysis which examines the effect of breastfeeding on childhood intelligence, Der et al. (2006) restrict their sample to differently fed sibling pairs, calculate corresponding differences in academic achievement scores, and find the impact of breastfeeding status (yes/no) and duration (in months) no longer reaches statistical significance. We identified only one previous study that compared between- and within-family estimates of breastfeeding across multiple dimensions of child health (Evenhouse & Reilly, 2005). However, these findings are based on the first wave of data from the National Longitudinal Study of Adolescent Health (Add Health) which was collected when members of the original cohort were in grades 7 through 12. Thus, breastfeeding status was obtained solely through retrospective maternal report and outcome measures were captured at a single point in time.

The current study builds upon the existing literature in a number of important ways. First, we employ 13 waves of nationally representative, prospective cohort data that span a time period of 24 years during which NLSY children are 4–14 years of age. Therefore, negative outcomes, which could be the result of not breast-feeding, that are discernable only after a latency period or experienced intermittently will still be captured in our analyses. For example, if a respondent's BMI follows a trajectory where it only exceeds normal limits after age 10, he/she will still be classified as overweight in the years that correspond to these later ages. If we were to average this respondent's BMI across all waves of data, we might erroneously determine that he/she was never overweight. Second, both indicators of breastfeeding (i.e. status and duration) are assessed within two years of a child's birth; therefore, our measures of breastfeeding are not subject to recall bias to the same extent as those derived from retrospective questions. This is a particular strength of the current study since recent evidence suggest maternal recall of breastfeeding initiation and duration is accurate when obtained within 3 years of the birth (Li, Scanlon, & Serdula, 2005). Third, we incorporate a number of outcome measures designed to capture multiple dimensions of childhood health and wellbeing – including physical, behavioral, and mental/intellectual capabilities. If breastfeeding is shown to be a more powerful predictor of behavioral as opposed to physical outcomes once selection into infant feeding practices is more fully taken into account, we will be able to rule out differences across study samples as a potential explanation. Finally, we include some outcome measures that have not been previously assessed in prior research efforts that rely on sibling comparisons to more accurately capture the association between breastfeeding and child health and wellbeing – most notably, behavioral outcomes such as hyperactivity, parental attachment, and compliance as well as self-perceptions of scholastic ability.

Data and methods

Description of the data

We utilize data from the National Longitudinal Study of Youth, 1979 Cohort (NLSY79) for the years between 1986 and 2010 to examine the association between infant feeding practices and child health and wellbeing. The NLSY79 is a nationally representative, prospective cohort study containing information on 12,686 young men and women who were between the ages of 14 and 22 in 1979. In 1986, a separate biennial survey of all children born to original NLSY79 female respondents was initiated. By 2010, the NLSY-Children included 11,504 children from 4932 mothers, ranging from 0 to 14 years of age. Children are directly assessed by trained interviewers and additional information is obtained from his/her mother. Given the longitudinal nature of the study design, the rigor with which surveys have been conducted, and the wealth of information spanning the early life course, the NLSY remains one of the best datasets with which to examine the influence of early life factors and subsequent health and wellbeing.

We limit the study sample to NLSY children who were between 4 and 14 years of age for the years between 1986 and 2010 who have valid information on either one of the breastfeeding measures (status or duration) and the outcome of interest. The longitudinal nature of the NLSY-CYA allows us to capture multiple outcome measures over time; thus, each respondent can contribute between 1 and 13 years of data. We began with 9101 NLSY children who were between 4 and 14 years of age during the time period of interest and born after 1978 (so we had prospective data on breastfeeding). We dropped 675 (7.4%) cases due to missing values on either breastfeeding status or duration. We also exclude children who are multiples resulting in the deletion of 97 sets of twins and 2 sets of triplets. New paragraph beginning here. We generate findings for three subgroups. The full sample includes all respondents who were interviewed at least once between 1986 and 2010. The sibling sample is restricted to NLSY children for which a sibling was also assessed. Finally, the discordant sibling sample contains only siblings who were differently fed in infancy. If more than two children are born to the same mother, we include all of them in the sibling sample and adjust sampling weights accordingly by dividing the average custom weight for all siblings in a given family by the total number of siblings in that family. The full sample includes information from 8237 children from 4071 families who contribute 35,572 person-years of data. The sibling sample is limited to 7319 respondents from 3153 families who provide 31,815 person-years of data. The discordant sibling sample consists of 1773 children from 665 families who generate 7663 person-years of data. Thus, 89% of the original sample is retained in sibling analyses that compare outcomes within rather than across families.

Our decision to restrict the sample to children who were 4 years of age or older was primarily driven by this study's focus on the long- as opposed to short-term effects of breastfeeding on child outcomes, since much less is known about how infant feeding practices impact health and wellbeing in this age group. Second, we sought to limit the extent to which our results would be biased by reverse causality that could stem from instances in which sicker or more "difficult" children have a harder time breastfeeding, especially for extended periods of time. Finally, many of the outcome measures we were interested in examining, especially those that captured academic achievement, were not assessed until NLSY respondents were at least 4 or 5 years of age.

Description of the measures

In order to capture multiple dimensions of child health and wellbeing, we purposefully selected eleven dependent variables that are often invoked in the literature to provide empirical evidence of the beneficial effects of breastfeeding. They include three measures of physical health (body mass index, obesity, and asthma diagnosis), three behavioral indicators (hyperactivity, parental attachment, and behavioral compliance), and five outcomes specifically designed to predict academic achievement (Peabody Picture Vocabulary Test, Peabody Individual Achievement Test - Reading Recognition, and Peabody Individual Achievement Test – Math Ability, Wechsler Intelligence Scale (WISC), and scholastic competence). A full description of all outcome measures and the manner in which they were obtained is provided in Table 1.

We rely on two independent variables to capture infant feeding practices. Breastfeeding status (yes/no) was coded as 0 if the mother did not breastfeed and 1 if she breastfed him/her for any length of time. Breastfeeding duration (in weeks) was based on a question that asked how many weeks old the child was when the NLSY mother quit breastfeeding altogether. The distribution of this explanatory variable is positively skewed; therefore, we conducted sensitivity analyses in which we included the logarithmic transformation. Results remained qualitatively unchanged; thus, we present results for breastfeeding duration in its original metric.

To control, at least in part, for characteristics that are likely to confound the association between infant feeding practices and subsequent child health and wellbeing, we incorporate a number of covariates measured either *at the time of interview* or *at the time of the respondent's birth* in all multivariate regression models. Those that capture time varying circumstances at the date of interview include survey year, respondent's age (in years) and race (non-Hispanic White, non-Hispanic Black, and Hispanic), maternal marital status (currently married, cohabiting, single, or never married), region of the country (Northeast, Midwest, South, or West), maternal educational attainment (in years), total family income (adjust for inflation and reported in 2010 dollars), maternal employment status (full-time, part-time, or unemployed), and insurance coverage (private, public, or none). Variables that reflect time-invariant circumstances at the respondent's birth are as follows: maternal age (in years), birth order (first, second, third, etc.), maternal educational attainment (in years), total family income (adjusted for inflation and reported in 2010 dollars), maternal employment status (full-time, part-time, and unemployed), preterm birth (<37 weeks), maternal smoking during pregnancy (yes/no), maternal alcohol consumption during pregnancy (yes/no), and prenatal care initiation during the first trimester (yes/no).

Statistical modeling strategy

To more accurately estimate the effect of breastfeeding on long-term childhood health and wellbeing, we exploit a unique attributed of the NLSY-Children's data – its inclusion of multiple siblings born to the same mother, some of whom were breastfed and some of whom were not. Sibling comparisons offer a way to better estimate the counterfactual condition than OLS or logit regression models (Geronimus & Korenman, 1992; Geronimus, Korenman, & Hillemeier, 1994; Moffitt, 2005; Morgan & Winship, 2007). Their primary strength derives from the fact that each mother serves as her own "control" by embodying both the observed and unobserved social processes that determine whether or not she breastfeeds her child as well as the biological and behavioral implications that stem from this decision. We restrict our analyses to siblings and incorporate within-family fixed effects for each mother into regression models; thus, we are able to estimate how much of the effect of breastfeeding on the eleven outcomes of interest is due to unobserved differences between mothers who breastfeed and those who bottle-feed their children.

Sibling comparisons are a power methodological strategy to reduce selection bias, but they can only account for unobserved potential confounders that differ across – not within – families. However, all of the scenarios we can call to mind in which siblings are differently fed favor the breastfed sibling – for example, sibling A is born full-term and breastfed while sibling B is born preterm and bottle-fed. In this regard, we caution against interpreting within-family estimates as reflecting the "true" effect of breastfeeding on childhood health and wellbeing; rather, they provide a more stringent upper bound limit on the association of interest (Bound & Solon, 1999; Griliches, 1979).

We estimate the following statistical model:

$$y^*_{ijt} = \beta_0 + \beta_1 X_{ijt} + \beta_2 Z_{ij} + \beta_3 W_j + \gamma \text{Breast}_{ij} + \alpha_j + \delta_{ij} + \varepsilon_{ijt} \tag{1}$$

where i indexes the individual, j the family (i.e. mother), and t the year. X represents observable characteristics that vary within individuals (i.e. age); Z captures observable characteristics that vary across but not within individuals (i.e. birth order); and W includes observable characteristics that vary across mothers (i.e. maternal educational attainment). Unobserved heterogeneity is captured by α_j for mothers and δ_{ij} for individuals. y^* is a latent variable. For continuous outcomes, $y = y^*$. For dichotomous outcomes, $y = 1$ if $y^* > 0$ and 0 otherwise; furthermore, we assume that $\delta_{ij} + \varepsilon_{ijt}$ follows a logistic distribution.

Conventional regression models that are typically invoked to investigate the effect of breastfeeding on childhood health and wellbeing control for W_j by including as many covariates as possible. However, the fear is that unobserved factors that vary across mothers are also correlated with Breast_{ij}. With sibling data, we are able to condition on α_j; thus, comparisons now occur within mothers (or families). To maximize the amount of available information on each NLSY child, we include all observations on individual *i*. Alternatively, we could have averaged the value of each time-varying variable across *t* but this would have reduced intra-individual variation and, more importantly, might have led us to misclassify negative outcomes that occur intermittently or only after a latency period. It is possible that child-specific characteristics (i.e. the δ_{ij}s) are correlated with both Breast_{ij} and y^*_{ijt}. We cannot condition our analyses on δ_{ij}, but most arguments suggest that if δ_{ij} confounds the relationship between Breast_{ij} and y^*_{ijt}, the bias will most likely amplify rather than attenuate the effect of breastfeeding (Bound & Solon, 1999; Griliches, 1979).

First, we generate weighted descriptive statistics for all three subsamples (full, sibling, and discordant sibling). Next, we present unadjusted means for each outcome of interest, comparing children who were breastfed to those who were bottle-fed. This provides an initial sense of the extent to which estimates of the effect of breastfeeding on child health and wellbeing are likely to be biased by unobserved family background characteristics.

Then, we extend our analyses to a multivariate context by conducting between-family comparisons for both the full (Model 1) and sibling samples (Model 2). We generate a succession of least squares (for continuous outcomes) and logit (for dichotomous outcomes) population-average regression models using the xtreg and xtlogit commands, respectively, in Stata 13.0. This approach is

Table 1
Description of long-term child wellbeing outcomes.

Measure	Age range	Objective	Format	Method of assessment
Body Mass Index	4–14 years	To measure weight to height ratio. BMI is considered to be reliable indicator of body fat for most people.	Measurements of height and weight obtained during interview. BMI calculated by dividing current weight by height squared. Reported in kilograms per squared meters (kg/m2).	63.96% obtained by interviewer; 33.29% obtained via maternal report; and 2.76% obtained by child report.
Obesity	4–14 years	To determine if respondent's BMI is exceeds the 95th percentile.	Dichotomous variable coded as 1 if child's BMI is at or exceeds 95th percentile for age- and sex-specific distributions and 0 if child's BMI falls below the 95th percentile.	All calculations based on sex-specific BMI-for-age growth charts for the U.S. generated by the Centers for Disease Control (CDC) and conducted by NLSY staff.
Asthma	4–14 years	To measure whether the respondent currently has asthma	Dichotomous variable coded as 1 if parent reported that child has asthma and 0 if parent reported child does not have asthma.	Maternal Report
Hyperactivity[a]	4–14 years	To measure the frequency and range of childhood behavioral problems attributable to hyperactivity	Subset of six questions from Behavior Problem Index (BPI): (1) has difficulty concentrating or paying attention; (2) is easily confused or seems to be in a fog; (3) is impulsive or acts without thinking; (4) has a lot of difficulty getting his/her mind off certain thoughts; and (5) is restless or overly active and cannot sit skill. Answer of "not true" is given value of 0 and answers of "sometimes true" or "often true" are given value of 1.	Maternal report
Parental attachment	4–7 years	To measure aspects of the child's usual behavior related to secure/insecure parental attachment.	Subset of seven questions based on Campos and Kagan's Compliance Scale: (1) trouble soothing child; (2) child stays close when playing; (3) child copies your actions; (4) child upset when you leave; (5) child is demanding; (6) child is empathetic; (7) child wants to help with things.	Maternal report
Behavioral compliance	4–7 years	To measure aspects of the child's usual behavior regarding following/not following household rules.	Subset of seven questions based on Campos and Kagan's Compliance Scale: (1) child resists eating meals; (2) child obeys when told to eat; (3) child resists going to bed; (4) child obeys going to bed; (4) child protests TV rules; (6) child obeys TV rules.	Maternal report
PIAT math[a]	5–14 years	To measure academic achievement in mathematics as taught in mainstream education for children ages 5 through 14.	Test consisting of 84 multiple-choice items of increasing difficulty, beginning with such early skills as recognizing numerals and progressing to measuring advanced concepts in geometry and trigonometry.	Interviewer assessment
PIAT reading[a]	5–14 years	To measure word and letter recognition as well as pronunciation ability for children ages 5 through 14.	Test of 84 questions of increasing difficulty; child matches letters, names letters, and reads single words aloud.	Interviewer Assessment
Peabody picture vocabulary[a,b]	4–14 years	To measure hearing and receptive vocabulary for Standard American English.	Interviewer says a word and the child points to 1 of 4 pictures that best portrays the word's meaning.	Interviewer Assessment
Wechsler Intelligence Scale (WISC)[a,b]	7–14 years	To measure child's short-term auditory memory and ability to manipulate verbal information from temporary storage	Digits Forward: The child listens to and repeats a sequence of numbers said by the interviewer. Digits Backwards: The child listens to a sequence of numbers and repeats them in reverse order.	Interviewer Assessment
Scholastic competence[b]	8–14 years	To measure child's sense of self-competence in the domain of academic skills.	Six item Likert scale measure that asks child, "How true of you is this statement?" (1) Some kids feel they are very good at school work; (2) Some kids feel they are just as smart as other kids their age; (3) Some kids are pretty slow in finishing their school work; (4) Some kids often forget what they learn; (5) Some kids do very well at their school work; (6) Some kids have trouble figuring out the answers in school.	Child Report

Source: National Longitudinal Survey of Youth, 1979 – Children's sample (NLSY-Childrens).
[a] Dependent variables are standardized by age.
[b] Age range did vary slightly over time.

mathematically equivalent to using generalized equation estimation (GEE) with an exchangeable correlation structure. It allows us to estimate the average effect of breastfeeding on a specific child outcome across our sample while accounting for the nonindependence of observations (Rabe-Hesketh & Skrondal, 2012; Zeger, Liang, & Albert, 1988). In contrast to random effects models, which require that we correctly specifying the joint distribution of observed values and random effects, population-average models only demand that we appropriately identify the mean of an outcome given a set of covariates (Hubbard et al. 2010). In linear GEE and random effects models, the estimator of β has the same structural form as the Generalized Least Square (GLS) estimator – it is the method they use to calculate the variance of β that is different and some suggest less biased (Gardiner, Luo, & Roman, 2009; Hubbard et al., 2010).

Finally, we carry out within-family comparisons by restricting the sample to NLSY Children who had at least one sibling and incorporating fixed effects for each mother (Model 3). In this study, we are using the term, fixed effect model, as it is elucidated in the econometrics literature. Statistically, it is equivalent to entering a dummy variable for each individual mother so that the effect of any family-specific (ie. maternal) characteristic that does not vary over time is removed from the coefficient for breastfeeding. A fixed-effects approach treats specific family (i.e. maternal) characteristics as nuisance parameters rather than parameters of interest to be quantified. This is appropriate since the overarching objective of this paper is to quantify the extent to which typical estimates of the effects of breastfeeding on long term childhood outcomes are biased due to maternal characteristics that predict both infant feeding practices and subsequent childhood wellbeing. Standard errors are calculated using the Huber/White correction method and adjusted for intracluster correlation to account for the presence of heteroskedasticity and likelihood that the independence of error terms assumption has been violated (Moulton, 1990; Wooldridge, 2002, 2003).

When using the full sample, we incorporate custom sampling weights provided by the NLSY, which are calculated to account for the probability of the original NLSY79 mother being selected into the survey as well as the rate of sample attrition among both mothers and their children (Zagorsky, 2012). When using the sibling sample, we construct our own sampling weights by dividing the average custom weight of all siblings within a given family by the total number of siblings in that family. This is done because the unit of analysis has shifted from the individual child to the family (i.e. the mother), and we include all siblings rather than randomly selecting one sibling pair. Thus, we want to assure that each sibling within a given family is weighted equally and the total family weight does not exceed a value of 1. This prevents siblings from large families, which may be outliers on other key characteristics, from having an undo influence on the results.

To handle issues of missing data, we rely on multiple imputations with chained equations to generate values for all covariates (Royston, 2005). Conditional distributions for missing data on all variables are created using Gibbs sampling techniques. (van Buuren, 2012; Royston, 2005). Typically, the number of imputed datasets is dependent on the amount of total missing information, with scholars recommending that 3 to 5 datasets is acceptable for models containing up to 20% missing information (van Buuren, 2012; Royston, 2005; Rubin, 1987). In our analysis, the largest amount of total missing information for any model was 7%. For all analysis, we created five distinct data sets, which were all used in conjunction with the mi command in Stata to complete both descriptive and multivariate analyses. Following Von Hippel (2007), we impute values for all variables in a given model and then delete observations with missing data on either breastfeeding or dependent variables before running regression analyses.

Results

Descriptive findings for the full, sibling, and discordant sibling samples are presented in Table 2. It is apparent from these results that all three subgroups are remarkably similar to one another along a wide range of key sociodemographic indicators. The distributions of respondents according to mother's marital status as well as mother's employment status are consistent across the three subsamples employed in this study. For example, only 10–11% of

Table 2
Descriptive statistics for covariates across different NLSY-Children's subsamples, 1986–2010.

	Full sample[a]	Sibling sample[b]	Discordant sibling sample[b]
Measured at Interview			
Age	8.87	8.88	8.92
Female	48.85	48.76	50.84
Race/ethnicity			
Non-Hispanic White	74.49	73.72	66.60
Non-Hispanic Black	17.28	17.68	19.90
Hispanic	8.23	8.60	13.50
Mother's marital status			
Currently married	71.10	72.35	70.29
Cohabitating	5.37	5.25	4.91
Never married	6.65	6.23	5.90
Previously married	16.88	16.17	18.90
Region of country			
Northeast	17.84	17.84	19.95
Midwest	30.13	30.87	28.36
South	35.28	34.51	32.50
West	16.75	16.79	19.19
Insurance Coverage			
Private Insurance	82.57	82.05	78.33
Public insurance	11.46	11.92	14.68
No Insurance	5.97	6.04	6.99
Mother's Education (Yrs)	13.09	13.03	12.52
Mean Family Income (2010 $)	80,079	79,667	70,150
Mother's Employment Status			
Full-time	30.22	28.85	26.26
Part-time	43.81	44.12	44.83
Unemployed	25.98	27.03	28.92
Measured at birth			
% Preterm (<37 weeks)	12.61	12.31	15.12
Birth Order	1.97	2.08	2.26
Age of Mother	27.13	27.03	26.33
Mother's Education (Yrs)	12.86	12.81	12.25
Mean Family Income (2010 $)	67,861	67,222	61,864
Mother's employment status			
Full-time	10.69	10.01	10.02
Part-time	46.05	45.45	46.11
Unemployed	43.26	44.53	43.87
Mother Smoked During Pregnancy	30.31	29.87	31.35
Mother Drank During Pregnancy	48.74	48.63	42.68
Prenatal Care During 1st Trimester	83.34	83.34	82.06
N (Person-years)	35,572	31,815	7663
N (Individuals)	8237	7319	1773

Source: National Longitudinal Survey of Youth, 1979 – Children's Sample (NLSY-Childrens).
Notes: All data are weighted to reflect the complex sampling design of the NLSY79 study.
[a] The full sample is weighted using longitudinal custom probability weights provided by the NLSY.
[b] We calculate weights for the sibling sample by dividing the average custom weight of all siblings within a given family by the total number of siblings from that family.

mothers worked full-time the year in which they gave birth. This stands in comparison to 45%–46% who worked part-time and 43%–44% who were unemployed the year in which their child was born. Maternal health behaviors, such as cigarette smoking and timely prenatal care, also adhere to the same patterning across subgroups with 30%–31% of women reporting that they smoked cigarettes while they were pregnant and 82%–83% obtaining prenatal care during the first trimester. Interestingly, a slightly *smaller* percentage of mothers in the discordant sibling sample (43%) said they drank alcohol while pregnant compared to mothers in the full and sibling sample (49%).

In some ways, however, the discordant sibling sample is more disadvantaged than either the full or sibling sample, but these differences are slight and should not be overstated. For example, the proportions of racial minorities are greater in the discordant sibling sample (20% Black and 14% Hispanic) compared to both the full sample (17% Black and 8% Hispanic) and the sibling sample (18% Black and 9% Hispanic). Mothers in the full and sibling sample complete approximately one-half a year more of schooling than their counterparts in the sibling sample, both at birth (12.8 years vs. 12.3 years) and at time of interview (13.0 years vs. 12.5 years). Similarly, family incomes are lower, on average, for respondents who comprise the discordant subgroup ($62,000 at birth and $70,000 at interview) than either the full ($68,000 at birth and $80,000 at interview) or sibling ($67,000 at birth and 80,000 at interview) subgroups. Finally, 15% of children in the discordant sample were born preterm compared to 13% of children in the full sample and 12% in the sibling sample.

In Table 3, we present descriptive statistics by breastfeeding status (yes/no) for the eleven outcomes of interest across three different subgroups – the full NLSY-Children's sample, the sibling sample, and the discordant sibling sample. Results for the first two subgroups are remarkably similar. Mean levels of BMI, hyperactivity, math skills, reading recognition, vocabulary word identification, digit recollection, and scholastic competence as well as the percentage of respondents who are obese all appear to significantly ($p < 0.05$) differ between children who were breastfed and those who were not and are in the predicted direction, with breastfed children exhibiting better outcomes. When the sample is restricted to discordant siblings, mean scores across all eleven indicators of child health and wellbeing are comparable and differences between breast- and bottle-fed respondents are small enough to be attributable to random chance alone.

Regression coefficients and corresponding standard errors for select indicators of child health and wellbeing are reported in Table 4. All Models contain the full set of covariates listed in Table 2. Model 1 relies on the full NLSY-Children's sample and reveals between-family estimates; thus, it reflects the standard multiple regression approach in which observed potential confounders are entered into the analysis as covariates. Model 2 also presents findings from between-family comparisons but is restricted to NLSY siblings. Therefore, it demonstrates the extent to which our findings are consistent across the full and sibling samples. Model 3 is the most unique in that it incorporates maternal fixed effects and limits estimates to those occurring within, as opposed to between, families. Thus, Model 3 is the most rigorous test of the hypothesis that breastfeeding positively influences childhood health and wellbeing.

Findings from the full sample (Model 1) suggest that children who were breastfed during the first year of life were significantly better off than their bottle-fed counterparts. Asthma was the only endpoint of interest that did not adhere to the expected patterning of results, in which breastfed children do better than their bottle-fed counterparts. Results from Model 2 demonstrate the consistency of our findings between the full and sibling subsamples. With the exception of one outcome (hyperactivity), regression coefficients remain remarkably similar between Models 1 and 2, standard errors increase only slightly, and the conclusions that can be drawn are virtually identical.

The most stringent test of the hypothesis that breastfeeding during infancy positively influences long-term childhood outcomes occurs when we include fixed effects for each NLSY79 mother, thus limiting comparisons to within rather than across families (Table 4, Model 3). What is most striking about these findings is the extent to which regression coefficients are attenuated, with a few even changing signs. Furthermore, *none* of the estimates maintain statistical significance ($p < 0.05$). For example, the coefficient for BMI decreased by 66% from −0.41 in Model 2 to −0.14 in Model 3. Therefore, once unobserved heterogeneity across families is more stringently taken into account, breastfed children tend to have BMIs that are, on average, only 0.14 kg/m^2 smaller than bottle-fed children. Similarly, the magnitude of coefficients for scholastic achievement and memory based intelligence declines by 69% (PIAT Math), 57% (PIAT Reading), 78% (PPVT), and 29% (WISC) when switching from a standard regression approach to one that incorporates within-family fixed effects. Coefficients for parental

Table 3
Unadjusted means and (sample sizes) for select child wellbeing outcomes by breastfeeding status (yes/no), 1986–2010: All NLSY Children and sibling subsamples.

	Full sample[a]			Sibling sample[b]			Discordant sibling sample[b]		
	Breastfed		Not breastfed	Breastfed		Not breastfed	Breastfed		Not breastfed
Body Mass Index	17.83 (15,518)	***	18.55 (17,984)	17.78 (13,911)	***	18.47 (16,120)	18.40 (3471)		18.59 (3733)
Obesity (%)	11.91 (15,518)	***	17.38 (17,984)	11.63 (13,911)	***	17.03 (16,120)	16.36 (3471)		18.14 (3733)
Asthma (%)	7.91 (17,150)	+	6.79 (18,382)	7.43 (14,981)		6.40 (15,673)	7.95 (3768)		8.89 (3718)
Hyperactivity score[c]	101.79 (16,312)	***	104.68 (17,515)	101.91 (14,277)	***	104.47 (14,949)	102.97 (3582)		103.81 (3543)
Parental attachment	19.94 (5386)		19.29 (5715)	20.04 (4801)		19.39 (5095)	19.68 (1160)		19.54 (1193)
Behavioral compliance	25.19 (5358)		24.65 (5716)	25.23 (4778)		24.67 (5095)	24.93 (1166)		24.88 (1182)
PIAT math skills[c]	106.87 (13,783)	***	100.11 (15,113)	107.11 (12,114)	***	100.38 (12,968)	102.39 (3093)	+	101.06 (3042)
PIAT reading recognition[c]	109.36 (13,734)	***	103.35 (15,043)	109.58 (12,069)	***	103.43 (12,906)	106.30 (3078)	+	104.81 (3027)
Peabody picture vocabulary test[c]	100.40 (7639)	***	90.43 (8762)	100.91 (6666)	***	90.97 (7476)	94.54 (1743)		93.26 (1766)
Wechsler Intelligence Scale (WISC)[c]	10.38 (7039)	***	9.58 (8122)	10.38 (6317)	***	9.55 (7287)	9.91 (1579)		9.61 (1666)
Scholastic competence	178.63 (5015)	***	169.39 (7084)	178.49 (4568)	***	169.05 (6393)	173.27 (1266)		169.84 (1414)

Source: National Longitudinal Survey of Youth, 1979 – Children's Sample (NLSY-Children).
Notes: All data are weighted to reflect the complex sampling design of the NLSY79 study.
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $+p < 0.10$.
[a] The full sample is weighted using longitudinal custom probability weights provided by the NLSY.
[b] We calculate weights for the sibling sample by dividing the average custom weight of all siblings within a given family by the total number of siblings from that family.
[c] Dependent variables are standardized by age.

Table 4
Unstandardized coefficients and corresponding standard errors for breastfeeding initiation (yes/no) from regression models predicting select outcomes among NLSY Children aged 4–14, 1986–2010.

	Between-family estimates						Within-family estimates		
	Model 1			Model 2			Model 3		
	Full sample[a]			Sibling sample[a]			Sibling sample[b]		
	b		S.E.	*b*		S.E.	*b*		S.E.
Body Mass Index	−0.449	*** (33,502)	0.094	−0.413	*** (30,031)	0.101	−0.141	(30,031)	0.188
Obesity	−0.342	*** (33,502)	0.066	−0.369	*** (30,031)	0.074	−0.173	(30,031)	0.164
Asthma	0.261	* (34,663)	0.106	0.237	* (30,998)	0.117	0.023	(30,998)	0.222
Hyperactivity[c]	−0.631	* (32,973)	0.314	−0.355	(29,513)	0.348	−0.572	(29,513)	0.549
Attachment	0.277	* (11,101)	0.113	0.223	+ (9896)	0.122	−0.047	(9896)	0.205
Compliance	0.227	+ (11,074)	0.119	0.307	* (9873)	0.129	−0.204	(9873)	0.221
PIAT math[c]	2.175	*** (28,179)	0.312	2.066	*** (25,293)	0.331	0.646	(25,293)	0.601
PIAT reading[c]	2.019	*** (28,068)	0.346	2.001	*** (25,190)	0.370	0.868	(25,190)	0.690
Peabody picture vocabulary[c]	3.250	*** (15,969)	0.444	3.181	*** (14,342)	0.474	0.686	(14,342)	0.865
Wechsler Intelligence Scale[c]	0.329	*** (15,161)	0.084	0.311	** (13,604)	0.092	0.221	(13,604)	0.178
Scholastic competence	2.789	* (12,099)	1.204	2.363	+ (10,961)	1.304	−0.353	(10,961)	2.757

Source: National Longitudinal Survey of Youth, 1979 – Children's sample (NLSY-Childrens).
Notes: All data are weighted to reflect the complex sampling design of the NLSY79 study.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; + $p < 0.10$.
[a] Controls measured at the date of interview include: year, age, sex, race, marital status, region, insurance coverage, family income, mother's education, and mother's employment. Controls measured at the time of birth include: preterm birth, birth order, mother's age, family income, mother's education, mother's employment, smoked during pregnancy, drank during pregnancy, and timely prenatal care.
[b] Models include all control variables listed above as well as within family fixed effects.
[c] Dependent variables are standardized by age.

attachment, behavioral compliance, and scholastic competence switch direction, so that breastfed children appear to be worse off than children who were not breastfed on these outcomes of interest. However, we should interpret these findings with caution since point estimates fail to reach statistical significance ($p < 0.05$).

Table 5 presents findings from a set of analyses that are identical to those presented in Table 4 except the independent variable is breastfeeding duration (in weeks) as opposed to breastfeeding status (yes/no). Taken as a whole, these results reveal the same patterning as was evident in Table 4, whereby estimates of the effect of breastfeeding on a diverse set of childhood outcomes are substantially attenuated toward zero when we rely on sibling comparisons. Findings from Model 1, which reflects the standard multiple regression approach, illustrate that each additional week of breastfeeding is associated with significant decreases in BMI, the odds of obesity, and hyperactivity as well as significant increases in parental attachment, math achievement, reading recognition, vocabulary identification, memory based intelligence (WISC), and scholastic competence. Results from Model 2, which is restricted to the sibling sample, are almost identical to those from Model 1, suggesting that selection into the sibling subgroup is not likely to bias our findings.

Finally, estimates from fixed effects regression models that limit comparisons to those occurring within, as opposed to between, families reveal a different story – one in which breastfeeding for longer periods of time does not necessarily result in better childhood health and wellbeing. When moving from Model 2 to Model 3, the coefficients for PIAT Math, PIAT Reading, PPVT, and scholastic competence decrease by 79%, 83%, 92%, and 88%, respectively. Moreover, the sign of the coefficient is not only reduced but actually changes direction for three additional outcomes (BMI, obesity, and WISC). For behavioral outcomes, such as hyperactivity and parental attachment, including fixed effects for each NLSY mother reduces the size of the regression coefficient, but does so in a more stepwise fashion, and renders it statistically insignificant ($p < 0.05$). Point estimates for behavioral compliance remain consistent and marginally significant ($p < 0.10$) across the three models. Similar to the patterning evident in Table 4, asthma is the one outcome for which breastfeeding duration is consistently associated with *poorer* childhood health and wellbeing across all three models.

Discussion

Results from between-family comparisons suggest that both breastfeeding status and duration are associated with beneficial long-term child outcomes. This trend was evident for 10 out of the 11 outcomes examined here. When we more fully account for unobserved heterogeneity between children who are breastfed and those who are not, we are forced to reconsider the notion that breastfeeding unequivocally results in improved childhood health and wellbeing. In fact, our findings provide preliminary evidence to the contrary. When comparing results from between- to within-family estimates, coefficients for 10 of the 11 outcomes are substantially attenuated toward zero and none reach statistical significance ($p < 0.05$). Moreover, the signs of some of the regression coefficients actually change direction suggesting that, for some outcomes, breastfed children may actually be worse off than children who were not breastfed.

The findings presented here are consistent with those from a small but growing literature that seeks to more accurately assess

Table 5
Unstandardized coefficients and corresponding standard errors for breastfeeding duration (in weeks) from regression models predicting select outcomes among NLSY Children aged 4–14, 1986–2010.

	Between-family estimates						Within-family estimates		
	Model 1			Model 2			Model 3		
	Full sample[a]			Sibling sample[a]			Sibling sample[b]		
	b		S.E.	*b*		S.E.	*b*		S.E.
Body Mass Index	−0.007	** (33,502)	0.002	−0.007	** (30,031)	0.003	0.005	(30,031)	0.003
Obese	−0.007	** (33,502)	0.002	−0.006	* (30,031)	0.002	0.001	(30,031)	0.004
Asthma	0.004	* (34,663)	0.002	0.004	+ (30,998)	0.002	0.006	(30,998)	0.008
Hyperactivity[c]	−0.020	** (32,973)	0.007	−0.017	* (29,513)	0.008	−0.015	(29,513)	0.012
Attachment	0.009	*** (11,101)	0.003	0.008	** (9896)	0.003	0.005	(9896)	0.004
Compliance	0.005	+ (11,074)	0.003	0.006	+ (9873)	0.003	0.009	+ (9873)	0.005
PIAT math[c]	0.059	*** (28,179)	0.008	0.056	*** (25,293)	0.008	0.012	(25,293)	0.012
PIAT reading[c]	0.047	*** (28,068)	0.009	0.048	** (25,190)	0.009	0.008	(25,190)	0.014
Peabody picture vocabulary[c]	0.084	*** (15,969)	0.012	0.087	*** (14,342)	0.013	0.007	(14,342)	0.021
Wechsler Intelligence Scale[c]	0.007	*** (15,161)	0.002	0.006	* (13,604)	0.002	−0.005	(13,604)	0.003
Scholastic competence	0.119	*** (12,099)	0.029	0.126	*** (10,961)	0.032	0.015	(10,961)	0.058

Source: National Longitudinal Survey of Youth, 1979 – Children's Sample (NLSY-Childrens).
Notes: All data are weighted to reflect the complex sampling design of the NLSY79 study.
***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; + $p < 0.10$.
[a] Controls measured at the date of interview include: year, age, sex, race, marital status, region, insurance coverage, family income, mother's education, and mother's employment. Controls measured at the time of birth include: preterm birth, birth order, mother's age, family income, mother's education, mother's employment, smoked during pregnancy, drank during pregnancy, and timely prenatal care.
[b] Models include all control variables listed above as well as within family fixed effects.
[c] Dependent variables are standardized by age.

the association between breastfeeding and child health and well-being (Der et al., 2006; Evenhouse & Reilly, 2005; Kramer, 2010), but they extend our understanding of this topic in important ways. To our knowledge, we are the first to examine how infant feeding practices impact behavioral outcomes in childhood such as hyperactivity, secure parental attachment, and compliance. Evenhouse and Reilly (2005) incorporate individual questions that ask whether or not a child "feels close to his/her mother" or "says she is warm and loving"; however, these measures are too restrictive to capture a complex concept such as parental attachment. Moreover, attachment and compliance were two of the five outcomes for which the regression coefficient switched directions when moving from between- to within-family analyses, suggesting that selection into breastfeeding may be a particularly powerful source of bias for behavioral based outcomes.

Although two previous studies using sibling comparisons (Der et al., 2006; Evenhouse & Reilly, 2005) focused on academic achievement as an endpoint of interest, they did not investigate the extent to which breastfeeding impacts self-perceptions of scholastic competence. This is critical since educational achievement appears to be strongly influenced by an individual's perception of his/her academic abilities, especially those that derive from an individual's race or gender (Steele, 2011). We find that once analyses were restricted to within- as opposed to between-family comparisons, breastfed children scored slightly *lower* than children who were not breastfed on the scholastic competence scale. Moreover, breastfeeding for longer periods of time did not result in significant improvements in self-perceptions of scholastic ability. Results for breastfeeding duration reveal a dramatic reduction, but not a reversal, in its influence on scholastic competence. These findings, along with others presented here, suggest that the relationship between breastfeeding and long-term childhood outcomes may not be as consistent and straightforward as once thought.

Some of the most notable findings from the current study concern physical health outcomes such as BMI and obesity. Findings from the few extant research efforts that examine the association between breastfeeding and the subsequent risk of obesity using sibling comparisons are conflicting (Gillman et al., 2006; Metzger & McDade, 2010; O'Tierney et al. 2009), even when the study sample used is from the same nationally representative dataset (Evenhouse & Reilly, 2005; Nelson et al., 2005). The results presented here suggest that unobserved family characteristics are likely to upwardly bias "typical" estimates of breastfeeding on BMI and obesity since regression coefficients from sibling comparisons for breastfeeding status are substantially attenuated toward zero while those for breastfeeding duration actually switch directions. These findings stand in sharp contrast to those from recent studies that attribute much of the Black/White disparity in obesity to racial differences in breastfeeding (Harder et al., 2005; Weden, Brownell, & Rendall, 2012).

Study limitations

This study provides evidence that the link between breastfeeding and childhood health and wellbeing may not be as robust as originally thought; however, the conclusions drawn should be considered in the context of its limitations. First, due to social desirability, women might exaggerate the extent to which they breastfed. For between family estimates, this misclassification is likely to result in an underestimation of the positive effect of breastfeeding on child outcomes. The direction toward which this classification error biases within family estimates is less clear. They could be downwardly biased if the variance of the explanatory variable is reduced to such an extent as to outweigh endogenous between sibling differences in breastfeeding (Bound & Solon, 1999).

We do not, however, believe this to be case with the results presented here. Empirical evidence suggests that maternal recall of breastfeeding initiation and duration, especially when obtained within 3 years of birth, provides an accurate measure of actual events (Li et al., 2005). This is a particular strength of the current research effort, especially in comparison to the Evenhouse and Reilly (2005) study which was solely based on retrospective breastfeeding measures 12–18 years prior. Second, we would expect bias attributable to measurement error in the explanatory variable to be more pronounced in analyses for which the response changes from no to yes as opposed to increasing by a few weeks. However, findings presented in Tables 4 and 5 are remarkably consistent and reveal similar patterns concerning reductions in the magnitude of regression coefficients when comparing between and within family estimations. Finally, Bound and Solon (1999) illustrate that means-reverting measurement error, which is the type most likely to lead to bias in our indicators of breastfeeding duration will produce between- *and* within-family estimates that are upwardly biased. Thus, the "true" effect of breastfeeding on child health and wellbeing is likely to smaller than is suggested by our sibling comparisons. How much smaller remains an empirical question.

Second, NLSY-Children are the offspring of women who were between the ages of 14 and 21 in 1979. This cohort experienced childbearing at the same time that infant feeding practices were dramatically changing in the United States. For example, rates of breastfeeding increased from 54% in the early 1980s to 77% in 2010 (CDC, 2013). Women who breastfed their infants during the early years of this study are likely to have been a more select group of individuals than those who breastfed by the end of the study period. To investigate this further, we conducted sensitivity analyses that included interaction terms (breastfed*year) for all outcomes of interest and failed to find evidence that the impact of breastfeeding on child health and wellbeing significantly changed over time.

Finally, the NLSY-Children's sample constitutes a group of individuals born to the *female* participants of a nationally representative prospective cohort study that has experienced at least some lost-to-follow-up over time. Thus, our sample is not representative of the current U.S. population between the ages of 4 and 14, rather it is a representative sample of children who have been born to this population of women. This shortcoming will only introduce bias if participants significantly differ from nonparticipants along dimensions that are related to both breastfeeding and child health and wellbeing. Given that are main findings are based on within–family comparisons, this type of selection bias will be less problematic that if we had relied solely on standard regression models based on between-family comparisons. Further, prevalence of breastfeeding and our outcomes are similar to those found in nationally representative surveys conducted during the same time period (Chase-Lansdale, Mott, Brooks-Gunn, & Phillips, 1991; Der et al., 2006).

Conclusion

This study was undertaken to gain a better understanding of the effects of breastfeeding on long-term childhood health and wellbeing. A mother's decision to breastfeed her child as well as how long she is able to do so is based on a complex web of personal, familial, and social factors. It often requires that women dramatically reduce their hours working outside the home, have jobs with maximum flexibility, and/or rely heavily on wages from partners to make up for lost income. This is a sacrifice for all women, regardless of how much they want to do it or how important they think it is. This trade off, however, may be especially untenable for poor or minority women who already face reduced access to steady, full-time employment, have few or no benefits, and lower than average salaries often in conjunction with the added pressures of single parenthood (Rippeyoung & Noonan, 2012).

Efforts to increase breastfeeding that solely focus on individually based behavior change without addressing the economic and social realities women face and the difficult tradeoffs they are forced to make in the months following the birth of their child risk alienating and stigmatizing the very women they hope to help. Instead, they need to be considered in conjunction with social policies that also influence a mother's ability to breastfeed, especially when current recommendations are that women exclusively do so for at least 6 months of age. For example, parental leave in the United States is limited to 12 unpaid weeks following the birth or adoption of a newborn and stands in sharp contrast to similar policies provided in almost all other developed nations, both in terms of its limited duration and lack of financial remuneration (Guendelman et al., 2009). Furthermore, we face a serious lack of qualified daycare providers in this country, the cost of which is often prohibitive for many working families (Blau & Kahn, 2013), which can make continuous breastfeeding even more difficult for a working mother to maintain. A truly comprehensive approach to increasing rates of breastfeeding in the U.S., with a particular focus on reducing racial and SES disparities, will need to work toward increasing and improving parental leave policies, flexible work schedules and health benefits even for low-wage workers, and access to high quality child care that can ease the transition back to work for both mother and child. Hopefully, this multifaceted approach will allow women who want to breastfeed to do so for as long as possible without promoting a cult of "total motherhood" (Wolf, 2011) in which women's identities are solely constructed in terms of providing the best possible opportunities for their children and the risks associated with a failure to breastfeed are drastically overstated.

Acknowledgment

This research was supported in part by R24-HD058484 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development awarded to the Ohio State University Institute for Population Research.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.socscimed.2014.01.027.

References

American Academy of Pediatrics. (2012). Breastfeeding and the use of human milk. *Pediatrics, 129*(3), e827–e841.

Armstrong, J., & Reilly, J. J. (2002). Breastfeeding and lowering the risk of childhood obesity. *The Lancet, 359*, 2003–2004.

Arenz, S., Rucker, R., Koletzko, B., & von Kries, R. (2004). Breastfeeding and childhood obesity – a systematic review. *International Journal of Obesity, 28*, 1247–1256.

Baker, M., & Milligan, K. (2008). Maternal employment, breastfeeding, and health: evidence from maternity leave mandates. *Journal of Health Economics, 27*(4), 871–887.

Blau, F. D., & Kahn, L. M. (2013). *Female labor supply: Why is the us falling behind?* NBER. Working Paper 18702.

Bound, J., & Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review, 18*(2), 169–182.
van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.
Carlsen, S. M., Jacobsen, G., & Vanky, E. (2010). Mid-pregnancy androgen levels are negatively associated with breastfeeding. *Acta Obstetricia et Gynecologica Scandinavica, 89*(1), 87–94.
Centers for Disease Control and Prevention [CDC]. (2013). Progress in increasing breastfeeding and reducing racial/ethnic disparities – United States, 2000–2008 births. *Morbidity and Mortality Weekly Report, 62*(5), 77–80.
Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., & Phillips, D. A. (1991). Children of the national longitudinal survey of youth: a unique research opportunity. *Developmental Psychology, 27*(6), 918–931.
Currie, J. (2009). Health, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature, 47*(1), 87–122.
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.
Evenhouse, E., & Reilly, S. (2005). Improved estimates of the benefits of breastfeeding using sibling comparisons to reduce selection bias. *Health Services Research, 40*(6 Pt 1), 1781–1802.
Farley, T. (2012). *Latch on NYC: a hospital-based initiative to support a mother's decision to breastfeed*. New York City Department of Health.
Forste, R., & Hoffman, J. P. (2008). Are U.S. mothers meeting the Healthy People 2010 breastfeeding targets for initiation, duration, and exclusivity? The 2003 and 2004 national immunization surveys. *Journal of Human Lactation, 24*(3), 278–288.
Gardiner, J. C., Luo, Z., & Roman, L. A. (2009). Fixed effects, random effects, and GEE: what are the differences? *Statistics in Medicine, 28*(2), 221–239.
Geronimus, A. T., & Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *The Quarterly Journal of Economics, 107*(4), 1187–1214.
Geronimus, A. T., Korenman, S., & Hillemeier, M. M. (1994). Does young maternal age adversely affect child development? Evidence from cousin comparisons in the united states. *Population and Development Review, 20*(3), 585–609.
Gillman, M. W., & Mantzoros, C. S. (2007). Commentary: breast-feeding, adipokines, and childhood obesity. *Epidemiology, 18*(6), 730–732.
Gillman, M. W., Rifas-Shiman, S. L., Berkey, C. S., et al. (2006). Breastfeeding and overweight in adolescence: within-family analysis. *Epidemiology, 17*(1), 112–114.
Griliches, Z. (1979). Sibling models and data in Economics: beginnings of a survey. *Journal of Political Economy, 87*(5), S37–S64.
Guendelman, S., Kosa, J. L., Pearl, M., Graham, S., et al. (2009). Juggling work and breastfeeding: effects of maternity leave and occupational characteristics. *Pediatrics, 123*(1), e38–e46.
Guilbert, T. W., Stern, D. A., Morgan, W. J., Martinez, F. D., et al. (2007). Effect of breastfeeding on lung function in childhood and modulation by maternal asthma and atopy. *American Journal of Respiratory and Critical Care Medicine, 176*(9), 843–848.
Harder, T., Bergman, R., Kallischnigg, G., & Plageman, A. (2005). Duration of breastfeeding and risk of overweight: a meta-analysis. *American Journal of Epidemiology, 162*(5), 397–403.
Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., et al. (2010). To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology, 21*(4), 467–474.
Innis, S. M., Gilley, J., & Werker, J. (2001). Are human milk long-chain polyunsaturated fatty acids related to visual and neural development in breast-fed term infants? *The Journal of Pediatrics, 139*(4), 532–538.
Ip, S., Chung, M., Raman, G., Chew, P., Magula, N., DeVine, D., et al. (2007). *Breastfeeding and maternal and infant health outcomes in developed countries*. Available at: http://www.ahrq.gov/downloads/pub/evidence/pdf/brfout/brfout.pdf Accessed 30 November 2012.
Kramer, M. S. (2010). "Breast is best": the evidence. *Early Human Development, 86*(11), 729–732.
Kramer, M. S., Chalmers, B., Hodnett, E. D., Sevkovskaya, Z., Dzikovich, I., Shapiro, S., et al. (2001). Promotion of breastfeeding intervention trial(PROBIT). *Journal of the American Medical Association, 285*(4), 413–420.
Kramer, M. S., Fombonne, E., Igumnov, S., Vanilovich, I., et al. (2008). Effects of prolonged and exclusive breastfeeding on child behavior and maternal adjustment: evidence from a large, randomized trial. *Pediatrics, 121*(3), e435–e440.
Kramer, M. S., Matush, L., Vanilovich, I., Platt, R., et al. (2007). Effect of prolonged and exclusive breast feeding on risk of allergy and asthma: cluster randomised trial. *British Medical Journal, 335*(7624), 815.
Kramer, M. S., Vanilovich, I., Matush, L., Bogdanovich, N., et al. (2007). The effect of prolonged and exclusive breast-feeding on dental caries in early school-age children. *Caries Research, 41*(6), 484–488.
Li, R., Scanlon, K. S., & Serdula, M. K. (2005). The validity and reliability of maternal recall of breastfeeding practice. *Nutrition Reviews, 63*, 103–110.
Martin, R. M., Gunnell, D., Owen, C. G., et al. (2005). Breast-feeding and childhood cancer: a systematic review with meta analysis. *International Journal of Cancer, 117*(6), 1020–1031.
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. *The American Journal of Clinical Nutrition, 82*(2), 281–295.
Mehta, N. K., Lee, H., & Ylitalo, K. R. (2013). Child health in the United States: recent trends in racial/ethnic disparities. *Social Science & Medicine, 95*, 6–15.
Metzger, M. W., & McDade, T. W. (2010). Breastfeeding as obesity prevention in the United States: a sibling difference model. *American Journal of Human Biology: The Official Journal of the Human Biology Council, 22*(3), 291–296.
Moffitt, R. (2005). Remarks on the analysis of causal relationships in population research. *Demography, 42*(1), 91–108.
Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.
Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics, 72*(2), 334–338.
Nelson, M. C., Gordon-Larsen, P., & Adair, L. S. (2005). Are adolescents who were breast-fed less likely to be overweight? Analyses of sibling pairs to reduce confounding. *Epidemiology, 16*(2), 247–253.
O'Tierney, P. F., Barker, D. J., Osmond, C., et al. (2009). Duration of breast-feeding and adiposity in adult life. *Journal of Nutrition, 139*(2), 422S–425S.
Oddy, W. H. (2004). A review of the effects of breastfeeding on respiratory infections, atopy, and childhood asthma. *The Journal of Asthma: Official Journal of the Association for the Care of Asthma, 41*(6), 605–621.
Quigley, M. A., Hockley, C., & Carson, C. (2011). Breastfeeding is associated with improved child cognitive development: a population-based cohort study. *The Journal of Pediatrics, 160*(1), 25–32.
Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). College Station, TX: STATA Press.
Rey, J. (2003). Breastfeeding and cognitive development. *Acta Pædiatrica, 92*, 11–18.
Rippeyoung, P. L. F., & Noonan, M. C. (2012). Is breastfeeding truly cost free? Income consequences of breastfeeding for women. *American Sociological Review, 77*(2), 244–267.
Royston, P. (2005). Multiple imputation of missing values: update of ice. *Stata Journal, 5*(4), 527–536.
Singh, G. K., Kogan, M. D., & Dee, D. L. (2007). Nativity/immigrant status, race/ethnicity, and socioeconomic determinants of breastfeeding initiation and duration in the United States, 2003. *Pediatrics, 119*(Suppl. 1), S38–S46.
Steele, C. (2011). *Whistling Vivaldi: How stereotypes affect us and what we can do*. New York, NY: W.W. Norton and Co.
U.S. Department of Health and Human Services. (2010). *Healthy people 2020*. http://healthypeople.gov/2020/.
U.S. Surgeon General. (2011). *The Surgeon General's call to action to support breastfeeding*. Available at: http://www.surgeongeneral.gov/library/calls/breastfeeding/calltoactiontosupportbreastfeeding.pdf Accessed 30.11.2012.
Van Rossen, L., Oenema, A., Steegers, E. A. P., et al. (2009). Are starting and continuing breastfeeding related to educational background? The generation R study. *Pediatrics, 123*(6), e1017–1027.
Von Hippel, P. T. (2007). Regression with missing *Y*s: an improved strategy for analyzing multiply imputed data. *Sociological Methodology, 37*(1), 83–117.
Weden, M. M., Brownell, P., & Rendall, M. S. (2012). Prenatal, perinatal, early life, and sociodemographic factors underlying racial differences in the likelihood of high body mass index in early childhood. *American Journal of Public Health, 102*(11), 2057–2067.
Wolf, J. (2011). *Is breast best? Taking on the breastfeeding experts and the new high stakes of motherhood*. New York: NYU Press.
Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT press.
Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *The American Economic Review, 93*(2), 133–138.
World Health Organization [WHO]. (2013). *Long-term effects of breastfeeding: a systematic review*. Available at: http://apps.who.int/iris/bitstream/10665/79198/1/9789241505307_eng.pdf?ua=1 Accessed 30.11.2012.
Young, T. K., Martens, P. J., Taback, S. P., Sellers, E. A. C., et al. (2002). Type 2 diabetes mellitus in children: prenatal and early infancy risk factors among native Canadians. *Archives of Pediatrics & Adolescent Medicine, 156*(7), 651–655.
Zagorsky, J. (2012). *Custom weighting program documentation*. http://www.nlsinfo.org/pub/usersvc/CustomWeight/CustomWeightingProgramDocumentation.htm Accessed 30.11.2012.
Zeger, S. L., Liang, K., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics, 44*, 1049–1060.