Explaining the Effect of Breastfeeding Promotion on Infant Weight Gain: The Role of Nutrition*

Anne Ardila Brenøe*[1]
University of Zurich

Jenna Stearns[2]
University of California, Davis

Richard M. Martin[3]
University of Bristol

September 2020

*Corresponding author: Anne Ardila Brenøe. 1: Anne Ardila Brenøe. ORCID 0000-0002-8045-8032. anne.brenoe@econ.uzh.ch. University of Zurich, Department of Economics, Schönberggasse 1, 8001 Zurich, Switzerland. 2: Jenna Stearns. jestearns@ucdavis.edu. University of California, Davis, Department of Economics, 1 Shields Avenue, Davis, CA 95616, the United States. 3: Richard M. Martin. richard.martin@bristol.ac.uk. University of Bristol, Bristol Medical School, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, the United Kingdom; National Institute for Health Research (NIHR) Bristol Biomedical Research Centre; and University Hospitals Bristol NHS Foundation Trust.

Abstract

Understanding the mechanistic effects of breastfeeding on child development is key to designing cost-effective policies that support optimal infant feeding practices. Benefits from breastfeeding may arise due to the consumption of breast milk itself, the substitution away from alternate food sources, or the physical act of breastfeeding. However, the prior literature has not determined which of these mechanisms is most important. We study the causal effects of an intervention to promote prolonged and exclusive breastfeeding on infant feeding patterns and child health through adolescence, using data from the largest randomized controlled trial on breastfeeding ever conducted. We present three key results. First, compared to the control group, infants exposed to the intervention were breastfed twice as much, received more calorie-dense feedings, and consumed more calories overall throughout infancy. Notably, their mothers substituted frequent use of water and juice for breast milk. Second, the intervention only significantly and persistently increased weight-for-age, in contrast to other health outcomes. Third, we show the increase in calories can almost entirely explain the weight gain in early infancy. Thus, we provide novel evidence indicating that the mediating mechanism of the effect of breastfeeding on weight gain is improved infant nutrition. We conclude that breastfeeding has beneficial effects on infant health in countries where breast milk is a replacement for low-quality and low-calorie liquids. In contrast, we caution policy makers from drawing conclusions about the effects of breastfeeding in environments where the alternative to breastfeeding is high-quality infant formula without more causal evidence from these settings.

Keywords: Breastfeeding, infant feeding, infant illness, anthropometric health, the Promotion of Breastfeeding Intervention Trial (PROBIT)

Significance To design cost-effective policies supporting optimal infant feeding practices, it is important to understand the mechanistic effects of breastfeeding. Using data from the only large-scale randomized controlled trial promoting prolonged and exclusive breastfeeding, we find it significantly and persistently increased weight-for-age, but did not have robust effects on other child health measures. To explain this result, we provide novel evidence of changes in infant feeding patterns. The estimated increase in calories treated infants received can fully explain the weight gain in early infancy. Our findings highlight the importance of the specific environment when considering the generalizability of the results. The effects of breastfeeding on child health may be very different in settings where higher-nutrition breast milk alternatives are more common.

1 Introduction

The World Health Organization (WHO) recommends that mothers worldwide should exclusively breastfeed infants during the child's first six months of life "to achieve optimal growth, development and health."[1] This recommendation is based on evidence of the health effects of exclusive breastfeeding from studies in both developed and developing countries (Kramer and Kakuma, 2012; Victora et al., 2016).

This breastfeeding recommendation has extensive public health and economic consequences. The Global Breastfeeding Collective (2017) estimates reaching a target 50 percent exclusive breastfeeding rate among infants under six months of age by 2025 will cost $5.7 billion. Breastfeeding advocates argue that this is a worthwhile investment. Estimates of the economic returns to breastfeeding indicate that the savings through lower infant mortality, improved health, and higher levels of human capital associated with exclusively breastfed infants substantially exceed the costs. Across the globe, initiatives to limit or regulate the marketing of infant formula, enact workplace breastfeeding policies, and promote paid leave to encourage breastfeeding are based on the idea that breastfeeding causally improves child outcomes.

But despite a large medical and epidemiological literature on breastfeeding, the causal evidence of the effects of breastfeeding on child outcomes is scarce. Much of the evidence comes from observational studies, where self-selection into breastfeeding and other types of infant feeding affects the interpretation of the results (Ip et al., 2009; Kramer et al., 2002*a*). Only few studies convincingly estimate the causal effects of breastfeeding through observational data (Del Bono and Rabe, 2012; Fitzsimons and Vera-Hernández, 2015). Arguably, the best causal evidence comes from the Promotion of Breastfeeding Intervention Trial (PROBIT) in Belarus—the only large-scale randomized controlled trial (RCT) on breastfeeding to date. This hospital-level intervention was based on the WHO/UNICEF Baby Friendly Hospital Initiative and substantially increased the duration and exclusivity of breastfeeding (Kramer et al., 2001). In this RCT, the overall benefits of the promotion of breastfeeding were of lower magnitude than those found in correlational settings. Nevertheless, studies evaluating this RCT have found positive effects on measures of length and weight in infancy, lower risk of rash or gastrointestinal infection by 12 months, improved cognitive development at age six, and higher risk of being overweight in adolescence (Kramer et al., 2001, 2002*b*, 2008*b*; Martin

[1] WHO (2011). WHO also recommends that children should continue to be breastfed in addition to receiving nutritious foods until at least two years of age.

et al., 2017).

A key limitation of the current causal literature is the lack of evidence on mechanisms. Any benefits of breastfeeding on child development could be driven by improved infant nutrition or calorie intake, the unique elements of breast milk that are not found in high-quality infant formula, or increased social stimulation. Understanding which of these mechanisms are most relevant is crucial for designing optimal breastfeeding policies—whether it is to promote and support breastfeeding itself, enable breast pumping at work (without the need of providing costly and lengthy maternity leaves), or promote programs to more broadly foster early childhood stimulation. Furthermore, the relative importance of these factors may vary in different settings, suggesting a need for country-specific recommendations.

Using data from the PROBIT RCT, we analyze the causal effects of promoting breastfeeding duration and exclusivity on child health and provide novel insight into the possible nutrition mechanism. We first study the effect of the breastfeeding promotion intervention on measures of infant and childhood health. We use an infant illness index to summarize episodes of gastrointestinal, respiratory, and other illness, rash, and hospitalizations during the first 12 months of life. We also study anthropometric health measures at nine time points from early infancy through age 16. These measures include standardized z-scores for weight, height, and body mass index (BMI), as well as indicators for being underweight and overweight and an overall anthropometric health index.

More importantly, we next provide novel evidence of how the breastfeeding promotion intervention affected infant feeding patterns. With this analysis, we shed light on the mediating mechanism of the effects of breastfeeding on child health in this particular context. Because much of the causal evidence used to set global breastfeeding policy comes from this single intervention, evaluating the external validity of these results is important. In particular, the alternative to breastfeeding may matter. In Belarus, less nutritious liquids, such as water and juice, were common alternatives to breast milk, and infant formula was less widely used compared to the United States (Grummer-Strawn, Scanlon and Fein, 2008). Our analysis highlights the importance of changes in nutrition as the primary mechanism by which breastfeeding improves child health in Belarus. This finding is important for designing cost-effective policies that support optimal infant feeding practices. But it also stresses the importance of the specific policy environment in thinking about the generalizability of the results. We caution that the benefits of breastfeeding and resulting policy implications may be different in settings where higher-nutrition alternatives to breast milk are more common.

2 The PROBIT Study

The PROBIT study was a cluster RCT in Belarus based on the WHO/UNICEF Baby-Friendly Hospital Initiative's "10 Steps to successful breastfeeding". As the only large-scale breastfeeding RCT conducted among healthy full-term infants, it was designed to identify the causal effects of breastfeeding promotion among mothers who had expressed a prenatal intention to breastfeed on breastfeeding duration and infant health (Kramer et al., 2001). Randomization occurred at the hospital level (the cluster), and treatment hospitals were given extensive training in methods to promote and prolong breastfeeding, maintain lactation, and resolve common problems.

We use data on 16,774 children born between June 1996 and December 1997 in 30 maternity hospitals.[2] Eligible infants were born weighing at least 2500 grams and at a gestational age of 37 weeks or greater. Participants were followed six times throughout their first year of life, and again at ages 6.5, 11.5, and 16. The hospitals were geographically dispersed across Belarus and were matched in pairs stratified on region, degree of urbanization, the annual number of deliveries, and the pre-intervention breastfeeding initiation rate. Treatment status within each hospital pair was assigned randomly. More details about the PROBIT design are available in Kramer et al. (2001).

Treatment hospitals received an intervention promoting breastfeeding (for details see SI A.1), while the control hospitals continued the standard practices in effect at the time of randomization. Across all hospitals, the postpartum stay after a routine vaginal delivery was 6–7 days (Kramer et al., 2001). The conventional practices at control sites included routine separation of mother and child, delayed onset of breastfeeding, scheduled feedings, routine use of water, formula, and other liquids in newborn diets, and recommendation of early introduction of solid foods (Patel et al., 2013).

As previously documented by Kramer et al. (2001), the breastfeeding promotion intervention substantially changed mothers' breastfeeding behavior. SI Figure A2 illustrates that mothers who gave birth at treated hospitals were considerably more likely

[2]Previous studies included 31 hospitals; however, we drop one unmatched hospital in our analysis. There is no evidence of differential attrition from the study by treatment status (SI Figure A1), and the intervention had no effect on infant mortality (Kramer et al., 2001).

to breastfeed exclusively for up to six months compared to mothers at control hospitals. For example, treated mothers were two and six times more likely to exclusively breastfeed infants at ages one and three months, respectively, compared to mothers in the control group with exclusive breastfeeding rates of only 27 percent at one month and seven percent at three months. The intervention also increased breastfeeding duration, with exposed mothers significantly more likely to breastfeed for at least twelve months.

During the late 1990s, Belarus resembled Western developed countries in terms of basic health services and sanitation. The country had high rates of adult literacy and immunization, and low rates of infant and child mortality. Maternity and postpartum infant care practices were comparable to those in North America and Western Europe 20 years earlier. At the time of the study, infant formula was readily available but expensive. Exclusive use of infant formula cost nearly 20 percent of the average monthly salary, compared to about 2.5 percent of the median monthly income in the United States. However, other social supports for new mothers were relatively generous. For instance, in contrast to the United States, mothers in Belarus had three years of maternity leave (often obligatory), possibly making it relatively easier for mothers to breastfeed.

3 Results

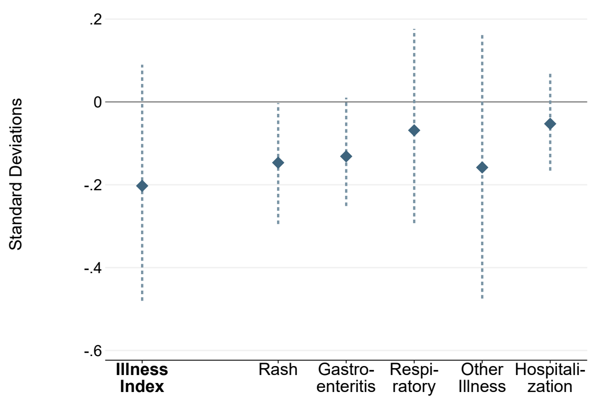
3.1 Child Health

We start by characterizing the health status of children who were not exposed to the breastfeeding promotion intervention. Children in the control group had healthy weights and heights during infancy and childhood, with average weight(height)-for-age z-scores of -0.11 (0.10) and 0.23 (0.31) standard deviations at age 1 month and 16 years, respectively (SI Table A2). Compared to reports from Canada, the United Kingdom, and the United States (Baker and Milligan, 2008; Fitzsimons and Vera-Hernández, 2015; Haider et al., 2014), the rates of illnesses during the first year of life were similar or slightly lower in our setting (SI Table A3).

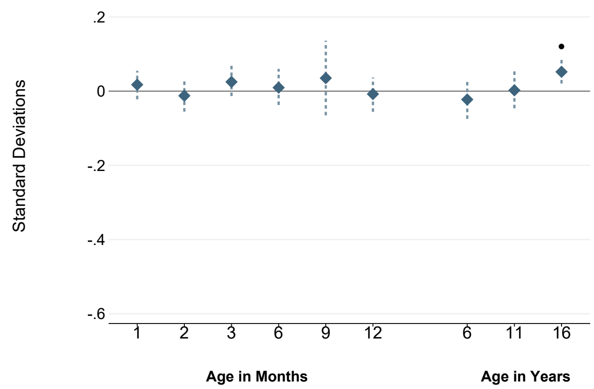
Figure 1 shows the effects of the breastfeeding promotion intervention on infant illness and anthropometric health throughout childhood (see Section 5.2 and SI A.2 for details on the outcome measures). Although we find suggestive evidence that the breastfeeding intervention decreased the likelihood of infant illness, our summary index and its individual components are not statistically significant. Nevertheless, the point es-

Figure 1
The Effect of Breastfeeding Promotion on Infant Illness and Childhood Anthropometric Health

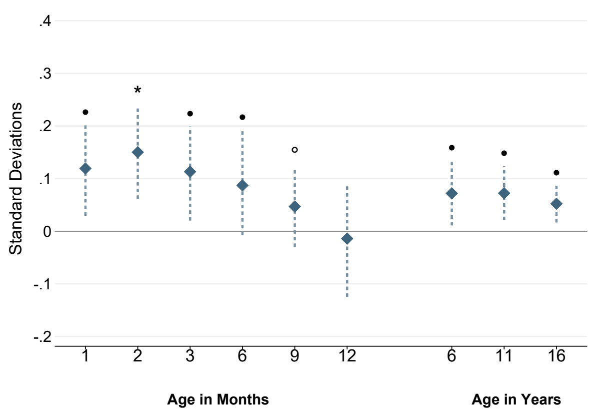
(a) Infant Illness



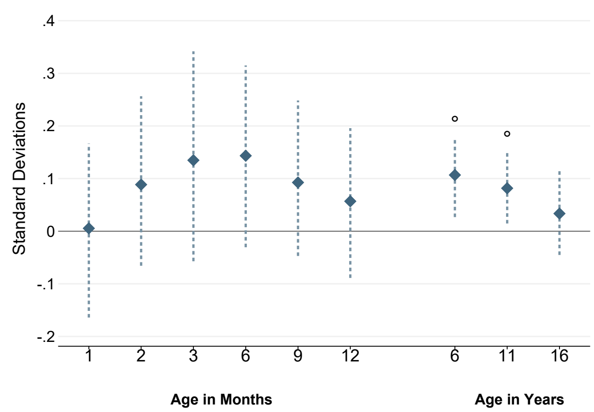
(b) Anthropometric Index



(c) Weight-for-age



(d) Height-for-age



Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. The models in graph 1a also control for age at last visit and the total number of observations during wave 1. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

timates are all negative and we cannot rule out medium-sized reductions in illness of up to 0.2-0.4 standard deviations. We do not find any heterogeneity in the effects on dimensions proxying the infectious environment (SI Figures A3). Our empirical design uses standard errors clustered at the hospital level and an adjustment for multiple hypothesis testing (see Section 5.1 for more details). The results suggest these adjustments are important, as other work has identified significant negative effects on the risk of any gastrointestinal tract infection, atopic eczema, and rash in the first 12 months (Kramer et al., 2001). If we instead consider each component of the illness index by month during infancy, we still find negative and mostly statistically insignificant estimates (SI Figure A4).

We do not find evidence of meaningful effects on the anthropometric index (Graph 1b). However, the null effects on the anthropometric index mask significant positive effects on weight-for-age (Graph 1c). Infants exposed to the breastfeeding promotion intervention were about 0.10 standard deviations heavier during the first six months of life compared to those in the control group. Interestingly, the effects on weight re-emerged later in childhood and persisted throughout adolescence. At age 16, children in the treatment group had a 0.04 standard deviation higher weight-for-age. Treated children had a 0.10 standard deviation higher height-for-age at ages 6 and 11, but the effect was small and no longer statistically significant at age 16. Children were not statistically more likely to be either overweight or underweight, suggesting that the effect on weight-for-age was driven by weights within the healthy range (SI Figure A5). The results on increased adiposity and growth are overall consistent with previous findings from the PROBIT study (Kramer et al., 2002*a*; Martin et al., 2013, 2017). In contrast to Martin et al. (2017), however, we do not find evidence of an increased probability of being overweight in the overall sample.

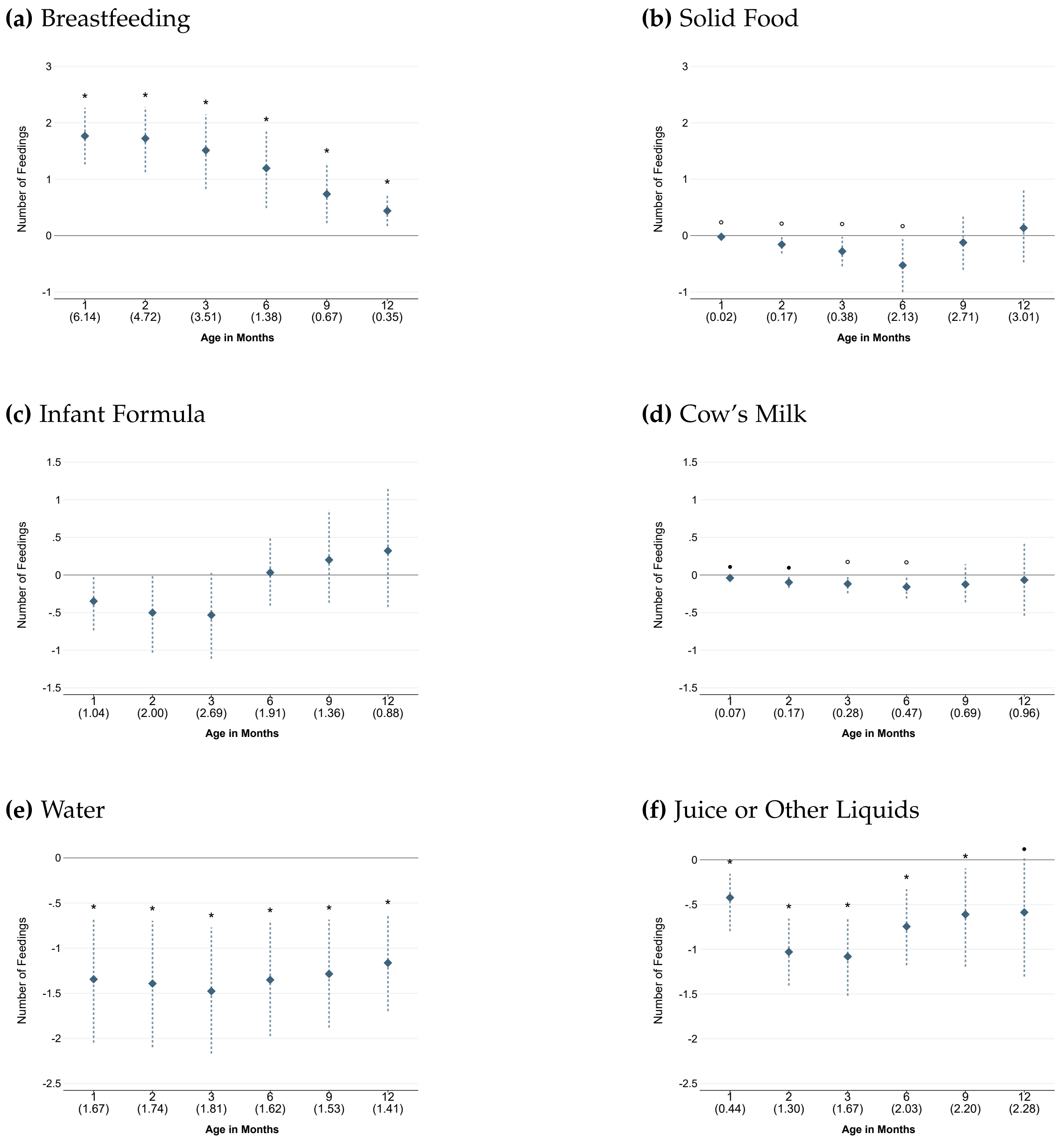
3.2 Infant Feeding Patterns

Studying infant illness and childhood anthropometric health, we only find robust, statistically significant effects of the breastfeeding promotion intervention on weight-for-age. This effect persisted until at least age 16, but appears to be driven by early weight gain during months also associated with higher rates of exclusive breastfeeding. What can explain these weight gain effects?

To examine the nutrition mechanism through which breastfeeding promotion affected weight, we turn to detailed data on infant feeding patterns. Figure 2 shows

Figure 2

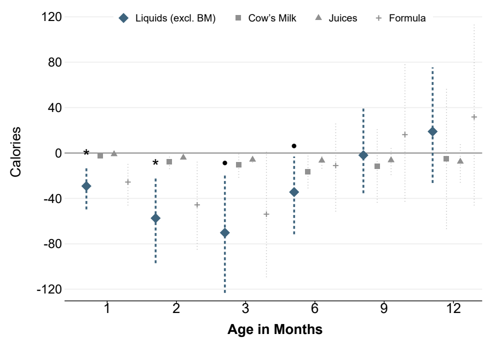
The Effect of Breastfeeding Promotion on Frequency of Infant Feedings per Day



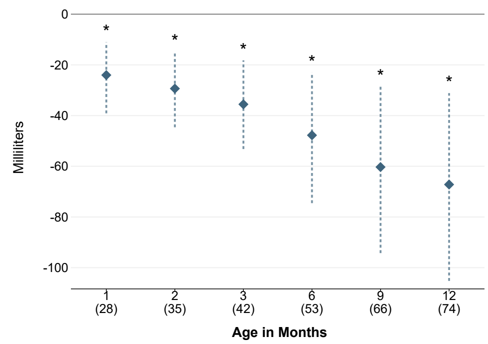
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable. Each outcome measures the number of feedings of that particular liquid or food that the infant recieved during the previous 24 hours.

Figure 3
The Effect of Breastfeeding Promotion on Estimated Liquid Intake by Calories and Volume

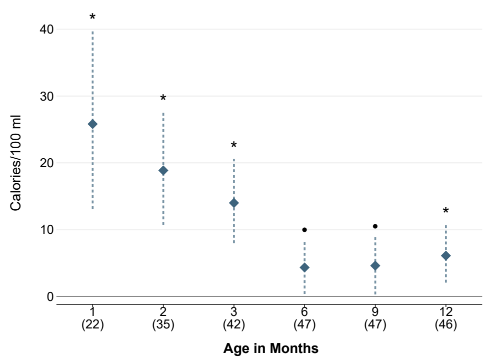
(a) Calories from Liquids (excl. BM)



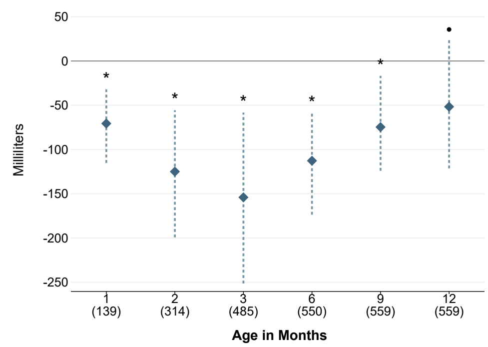
(b) Volume of Water



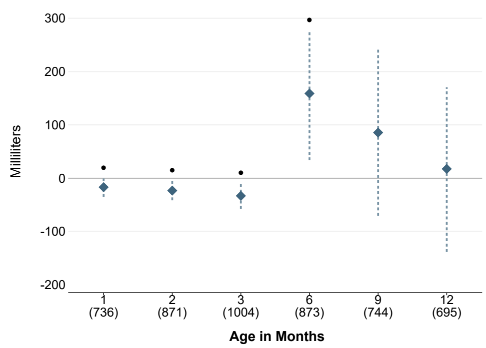
(c) Calories/100 ml Liquids (excl. BM)



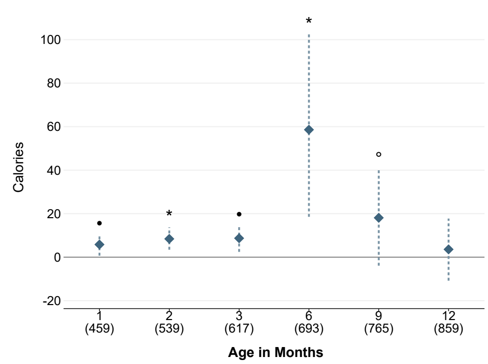
(d) Total Volume Liquids (excl. BM)



(e) Volume of All Liquids (incl. BM)



(f) Calories of All Liquids (incl. BM)



Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph except for Figure 3a, where we only perform multiple hypothesis testing across the estimates of Liquids (excl. BM). Multiple hypothesis results for all estimates in Graph 3a are given in SI Figure A6. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable. See SI A.2 for the exact construction of the outcome variables.

the number of feedings per day over the first 12 months of life by nutrition type (breast milk, solid food, infant formula, cow's milk, water, and juice or other liquids). As expected, given earlier work on breastfeeding rates, infants exposed to the treatment were breastfed about 1.7 more times per day during the first three months of life, and 0.8 more times per day between 6 and 12 months. This corresponds to approximately 428 more breastfeedings during the first year, an increase of 50 percent relative to the control group. While the frequency of infant formula feedings seemed to decrease during the first three months of life, the biggest substitutes for breastfeeding in terms of the number of feedings were water and juice, which are much less nutritionally rich. Infants in the treated group received about 1.4 and 0.7 fewer feedings per day of water and juice or other liquids, respectively, compared to those in the control group, and this persisted throughout the full first year of life. There was also a small reduction in the frequency of receiving cow's milk and solid food during the first six months. Thus, we observe a stronger compliance to the recommendation of first introducing solid food at age six months in the treatment group.

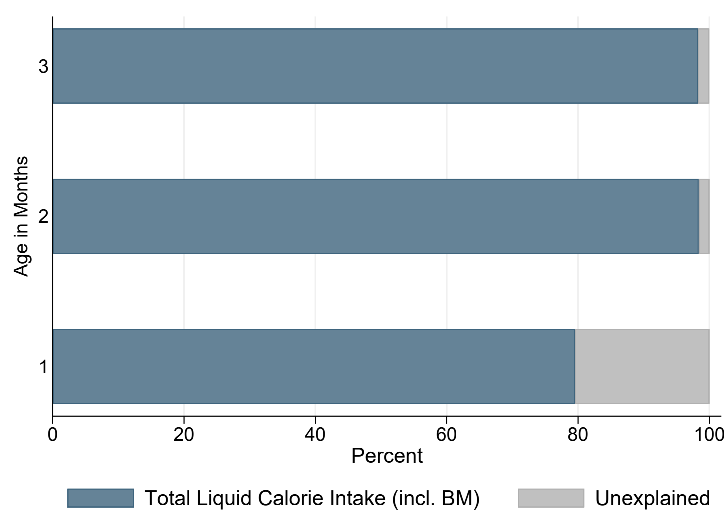
Next, we estimate differences in the volume of and caloric intake from liquids other than breast milk (Graphs 3a–3d of Figure 3). The intervention reduced the daily caloric intake from liquids excluding breast milk by about 50 kcal during the first six months. Although water and juice were the biggest substitutes for breast milk in terms of feeding frequency, the decrease in the caloric intake from infant formula among treated infants was ten times larger than that from juices (SI Figure A6). We also see a substantial reduction in the volume of water consumed throughout the first year of life. Together, these findings imply an increase in the density of calories treated infants received. Treated infants received around 20 kcal more per 100 ml liquids (excluding breast milk) during the first three months in particular, compared to control infants. Consistent with these results, treated infants received a smaller total volume of liquids other than breast milk on the order of 50–150 ml per day compared to infants in the control group.

Finally, we estimate the total volume of and caloric intake from liquids including breast milk. Because we do not observe the quantity of breast milk consumed (only the frequency of breastfeeding), this analysis is naturally associated with more measurement error, as we must estimate the volume per feeding (SI A.2). When including breast milk in the volume of all liquids, the negative effect is very small for ages 1–3 months and becomes positive at age 6 months (Graph 3e).[3] Graph 3f shows the estimated total caloric

[3] We note that the spike at age six months most likely is due to the delayed introduction of solid foods

intake from all liquids including breast milk. Treated infants received around 8 calories more per day from liquids than control infants at ages 1–3 months and even more at later ages. It is thus evident that the breastfeeding promotion intervention increased both the nutritional density and the total amount of calories received from liquids. One relevant data limitation is that we do not know which type of solid food infants received or how much they ate. Because there is little evidence of meaningful differences in solid food feedings between the treatment and control group, we interpret the results for liquids to imply that treated infants received more calories during their first year of life.

Figure 4
Mediation Analysis on Weight in Grams



Note: This figure shows the results of the mediation analysis, decomposing the overall treatment effect of weight-for-age at 1, 2, and 3 months. We use total caloric intake from all liquids including breast milk as the mediator. See SI A.3 for details.

Can this increase in calorie consumption explain the effects on weight gain? To formally test how much of the increase in weight can be explained by the effect of the breastfeeding promotion intervention on total calorie intake, we conduct a mediation analysis (SI A.3). We limit this analysis to ages 1–3 months because calories received from solid foods are minimal at these ages; thus, we have good estimates for the total calorie intake of all infants during this period. At age 1 month, we estimate that a notable 80 percent of the effect of the breastfeeding intervention on weight gain is mediated by the increase in the total calorie intake (Figure 4). Remarkably, the estimated increase in

compared to control children; we therefore do not claim that this represents a net effect on total calorie intake at this age.

calories from the intervention explains nearly the entire difference in body weight at ages 2 and 3 months (98 percent). Meanwhile, the reduction in the incidence of gastroenteritis in early infancy does not explain more than 0.2 percent of the effect on weight gain (SI Figure A8).

4 Discussion

WHO's recommendation of six months exclusive breastfeeding has extensive public health and economic consequences. However, this global recommendation is based on limited causal evidence from specific policy settings. Little is known about the external validity of the effects, or the specific mechanisms that might drive the effects of breastfeeding on child health and development. Answering these questions is important from a policy perspective in order to issue efficient and cost effective recommendations to support optimal infant feeding practices across the world.

In this paper, we study the causal effects of a breastfeeding promotion intervention on infant feeding patterns and childhood health, using data from the PROBIT study—the only large-scale RCT on breastfeeding to date. Consistent with prior work, we confirm that infants exposed to the breastfeeding intervention had a significantly and consistently higher weight-for-age from infancy through at least age 16. More importantly, we provide novel insights on the effects of the intervention on the nutritional composition of the diet that infants received. Infants exposed to the breastfeeding intervention were breastfed more and received less water, juice, and other liquids throughout their first year of life. This resulted in a more calorie-dense and calorie-rich diet. Treated infants consumed at least 8 calories more per day compared to those in the control group over the whole first year of life, which represents almost a two percent increase in calories per day compared to the recommended nutritional guidelines for infants.

These effects on infant feeding patterns are important for interpreting the mediating mechanisms of breastfeeding on child health. Our mediation analysis shows that the increased total caloric intake explains nearly the entire effect of the breastfeeding intervention on weight gain in early infancy. The results indicate that improved nutrition during key periods of growth have contemporaneous and lasting effects on weight gain. They also suggest that, at least in this setting where water and juice are frequent substitutes for breast milk, the primary benefit of breastfeeding is improved nutrition. The importance of nutrition as a mechanism is also consistent with evidence from Jamaica and Guatemala showing that nutritional supplementation among infants and toddlers improves contemporaneous weight gain (Schroeder et al., 1995; Walker et al., 1991). In contrast to these two studies, it is worth noting that our findings are from a setting where infants were not generally underweight or particularly unhealthy even in absence of the breastfeeding intervention. Furthermore, we do not find evidence of any noticeable heterogeneity in the effects of the breastfeeding promotion intervention on child health or feeding patterns with respect to either gender or socioeconomic status (SI A.4), supporting the nutrition channel rather than differences in parenting behavior or environment.

Interestingly, early breastfeeding appears to have a lasting effect on feeding patterns. Even after mothers exposed to the breastfeeding intervention stop exclusively breastfeeding, they are less likely to feed less-nutritious liquids, such as water or juice, than are control mothers, and are more likely to delay the introduction of solid food until the child reaches six months of age. Treated mothers are also actually somewhat more likely to use infant formula after they stopped exclusively breastfeeding, which is much more nutritionally rich and similar to breast milk than other liquids. These patterns may suggestively indicate that breastfeeding leads to a longer-term change in nutrition that could potentially explain the persistent effects on weight gain through adolescence. In support of this idea, we also find a robust decline in an index measuring problematic eating attitudes among treated children at age 11 (SI Figure A7).

These findings have important policy implications. Because mothers in Belarus did not use breast milk as a perfect substitute for infant formula, and instead breastfeeding replaced less nutritious liquid alternatives, the results on child health may be context specific. It is not clear that breastfeeding interventions would affect weight gain in populations where infant formula is already widely used as the primary alternative to breast milk. Our results suggest that studying the causal effects of breastfeeding in settings where high-quality infant formula is the alternative to breastfeeding is necessary to be able to provide clear policy recommendations for those environments. Although more work is needed, there is currently no evidence of causal effects of breastfeeding on early childhood health in other higher-income countries (Baker and Milligan, 2008; Del Bono and Rabe, 2012; Fitzsimons and Vera-Hernández, 2015).

Further, we do not find any effect of the breastfeeding promotion intervention on socioemotional or cognitive skills through age 16 (SI A.5). While we cannot rule out that the lack of effects on these human capital measures are specific to Belarus, the results suggest that any causal effects of breastfeeding on intelligence are small. In support of this, a recent review of correlational studies shows that controlling for maternal IQ substantially decreases the association between breastfeeding and child IQ (Horta, Loret de Mola and Victora, 2015). However, recent evidence from the United Kingdom suggests that breastfeeding improves children's cognitive development at age 7 (Fitzsimons and Vera-Hernández, 2015). Although we find weak evidence for improvements in some measures of cognitive skills at age 6 as well, these effects disappeared by age 16. Much like with the effects on health, more work is needed to understand the importance of cultural context in determining the degree of external validity, as well as to understand the mechanisms driving any causal effects of breastfeeding.

There are several channels through which breastfeeding could plausibly affect child health and development outcomes. For example, breast milk could provide a more calorie-dense diet, breast milk may contain health-improving antibodies or other non-nutrient benefits, or the physical act of breastfeeding may cause socioemotional stimulation for both the infant and mother. These three mechanisms may lead to very different policy recommendations. While the latter implies the need for policies that encourage breastfeeding itself, less costly policies that call only for the feeding of expressed breast milk may be sufficient for the first two. Furthermore, if the nutritional composition of breast milk compared to its common alternatives drives most of the effects of breastfeeding–as we find in Belarus–then recommending or subsidizing high-quality infant formula might be sufficient to capture the main benefits of breastfeeding.

While it is likely that all of these potential mechanisms may matter to some degree, nutrition appears to be by far the most important factor through which breastfeeding impacts child health and development in this setting. The increased calorie consumption of breastfed infants can explain over 90 percent of the effect on weight gain in the first three months of life. In contrast, reductions in illness (potentially driven by increased antibodies or decreased exposure to non-sterile foods or feeding equipment) explains less than one percent. We also find no evidence of significant socioemotional effects that could be driven by the physical act of breastfeeding.[4] Taken together, these results suggest that nutrition is the most important mechanism driving the effects of breastfeeding on child health. While more work is needed to explore the importance of these other mechanisms directly, it may be possible to interpret the effects of the breastfeeding intervention on child health and development in Belarus as an upper bound for effects in other settings

[4]In this paper, we do not consider any potential benefits of breastfeeding for mothers. Evidence suggests breastfeeding may reduce the risk of breast and ovarian cancer in mothers (Victora et al., 2016). It may also affect maternal mental health (Borra, Iacovou and Sevilla, 2015).

where high-quality infant formula is widely used and nutrition is likely to be relatively less important.

5 Materials and Methods

5.1 Empirical Strategy

We analyze the intent to treat (ITT) effects of the PROBIT breastfeeding promotion intervention by estimating the following specification:

$$Y_{iph} = \gamma_0 + \gamma_1 Treatment_h + Z_i'\delta + \theta_p + \varepsilon_{iph}, \tag{1}$$

where Y_{iph} is the outcome of interest for individual i at a specific age, born at hospital pair p, and hospital h. The variable *Treatment* is an indicator for whether the hospital received the breastfeeding intervention. We control for a vector of individual baseline characteristics, Z (birth weight in grams (squared); maternal and paternal age (squared); indicators for gender, cesarean section, gestational age at birth in weeks, maternal smoking and alcohol use during pregnancy, parents' marital and cohabitation status, number of siblings, maternal and paternal educational attainment, and quarter-by-year of birth). SI Table A1 shows descriptive statistics at baseline. Finally, θ_p are hospital pair fixed effects.

We allow for the errors, ε_{iph}, to be correlated at the hospital level. Due to the small number of clusters, we use the wild cluster bootstrap (WCB) to estimate p-values (Cameron, Gelbach and Miller, 2008) and conduct 999 replications. Finally, because we estimate effects for a number of related outcomes, all results are corrected for multiple hypothesis testing using the method developed in Benjamini, Krieger and Yekutieli (2006). Known as the *krieger* method, this is a "step-up" method of multiple hypothesis testing that is less data-intensive than methods designed for use in very large samples, and controls for the false discovery rate (FDR) rather than the family-wise error rate.

5.2 Data and Outcome Variables

We consider two broad groups of outcomes for children: health and infant feeding. SI A.2 provides more details regarding the data and construction of the variables.

We examine two dimensions of child health: infant illness and anthropometric health throughout childhood. We construct summary indices for these health outcomes, following Anderson (2008). The infant illness index includes indicators for rash, gastrointestinal illness, respiratory illness, other illness, and hospitalization during infancy. The anthropometric health index includes z-scores for weight-for-age, height-for-age, BMI-for-age, and indicators for being underweight and overweight at each survey wave. The anthropometric health index is constructed so that larger values reflect more beneficial outcomes.

Infant feeding data is based on maternal reports of feedings during the past 24 hours before each of the six infant checkups. We use information on the frequency and volume of feedings of breast milk, infant formula, cow's milk, water, juices or other liquids, and solid food (including cereals).

Acknowledgments: We are grateful to Maximilian Mähr, Claude Raisaro, and Yanzun Yang for excellent research assistance. We thank Pietro Biroli, Ernst Fehr, Claudio Schilter, Hannes Schwandt, Xiaoyue Shan, Miriam Wüst and seminar participants at the International Health Economics Association Conference 2019, the Importance of Early-Life Circumstances Workshop at the University of Copenhagen, and the University of Zurich for helpful discussions and comments. We are grateful to the cohort members, their parents and the study pediatricians and auditors who participated so willingly in the study.

Richard M. Martin was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol, Weston NHS Foundation Trust, and the University of Bristol. Anne Ardila Brenøe is grateful for the support of the Family Larsson-Rosenquist Foundation (LRF). The views expressed in this article are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care at the University of Bristol, or LRF.

References

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association*, 103(484): 1481–1495.

Baker, Michael, and Kevin Milligan. 2008. "Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates." *Journal of health economics*, 27(4): 871–887.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika*, 93(3): 491–507.

Borra, Cristina, Maria Iacovou, and Almudena Sevilla. 2015. "New evidence on breastfeeding and postpartum depression: the importance of understanding women's intentions." *Maternal and child health journal*, 19(4): 897–907.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics*, 90(3): 414–427.

Cole, Tim J, Mary C Bellizzi, Katherine M Flegal, and William H Dietz. 2000. "Establishing a standard definition for child overweight and obesity worldwide: international survey." *Bmj*, 320(7244): 1240.

Del Bono, Emilia, and Birgitta Rabe. 2012. "Breastfeeding and child cognitive outcomes: Evidence from a hospital-based breastfeeding support policy."

Fitzsimons, Emla, and Marcos Vera-Hernández. 2015. "Breastfeeding and Child Development." *University College London and Institute for Fiscal Studies, London (May 2015), http://www. home pages. ucl. ac. uk/-uctpamv/papers/breastfeeding. pdf*.

Gelbach, Jonah B. 2016. "When do covariates matter? And which ones, and how much?" *Journal of Labor Economics*, 34(2): 509–543.

Global Breastfeeding Collective. 2017. "Nurturing the health and wealth of nations: the investment case for breastfeeding." *World Health Organization*.

Grummer-Strawn, Laurence M, Kelley S Scanlon, and Sara B Fein. 2008. "Infant feeding and feeding transitions during the first year of life." *Pediatrics*, 122(Supplement 2): S36–S42.

Haider, Steven J, Lenisa V Chang, Tracie A Bolton, Jonathan G Gold, and Beth H Olson. 2014. "An evaluation of the effects of a breastfeeding support program on health outcomes." *Health services research*, 49(6): 2017–2034.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review*, 103(6): 2052–86.

Horta, Bernardo L, Christian Loret de Mola, and Cesar G Victora. 2015. "Breastfeeding and intelligence: a systematic review and meta-analysis." *Acta paediatrica*, 104: 14–19.

Ip, Stanley, Mei Chung, Gowri Raman, Thomas A Trikalinos, and Joseph Lau. 2009. "A summary of the Agency for Healthcare Research and Quality's evidence report on breastfeeding in developed countries." *Breastfeeding medicine*, 4(S1): S–17.

Kramer, Michael S, and Ritsuko Kakuma. 2012. "Optimal duration of exclusive breastfeeding." *Cochrane database of systematic reviews*, , (8).

Kramer, Michael S, Beverley Chalmers, Ellen D Hodnett, Zinaida Sevkovskaya, Irina Dzikovich, Stanley Shapiro, Jean-Paul Collet, Irina Vanilovich, Irina Mezen, Thierry Ducruet, et al. 2001. "Promotion of Breastfeeding Intervention Trial (PROBIT): a randomized trial in the Republic of Belarus." *Jama*, 285(4): 413–420.

Kramer, Michael S, Eric Fombonne, Sergei Igumnov, Irina Vanilovich, Lidia Matush, Elena Mironova, Natalia Bogdanovich, Richard E Tremblay, Beverley Chalmers, Xun Zhang, et al. 2008*a*. "Effects of prolonged and exclusive breastfeeding on child behavior and maternal adjustment: evidence from a large, randomized trial." *Pediatrics*, 121(3): e435–e440.

Kramer, Michael S, Frances Aboud, Elena Mironova, Irina Vanilovich, Robert W Platt, Lidia Matush, Sergei Igumnov, Eric Fombonne, Natalia Bogdanovich, Thierry Ducruet, et al. 2008*b*. "Breastfeeding and child cognitive development: new evidence from a large randomized trial." *Archives of general psychiatry*, 65(5): 578–584.

Kramer, Michael S, Tong Guo, Robert W Platt, Stanley Shapiro, Jean-Paul Collet, Beverley Chalmers, Ellen Hodnett, Zinaida Sevkovskaya, Irina Dzikovich, Irina Vanilovich, et al. 2002*a*. "Breastfeeding and infant growth: biology or bias?" *Pediatrics*, 110(2): 343–347.

Kramer, Michael S, Tong Guo, Robert W Platt, Stanley Shapiro, Jean-Paul Collet,

Beverley Chalmers, Ellen Hodnett, Zinaida Sevkovskaya, Irina Dzikovich, Irina Vanilovich, et al. 2002*b*. "Breastfeeding and infant growth: biology or bias?" *Pediatrics*, 110(2): 343–347.

Kramer, MS, B Chalmers, ED Hodnett, Z Sevkovs Kaya, I Dzikovich, S Shpiro, et al. 2000. "Promotion of breastfeeding intervention Trial (PROBIT): A cluster randomized trial in the Republic of Belarus. Koletzok, KF Michaelsen, and O. Hernell (Eds.), Short and Long Term Effects of Breastfeeding on Child Health (PP327-345)."

Lupton, Joanne R, JA Brooks, NF Butte, B Caballero, JP Flatt, SK Fried, et al. 2002. "Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids." *National Academy Press: Washington, DC, USA*, 5: 589–768.

Martin, Richard M, Michael S Kramer, Rita Patel, Sheryl L Rifas-Shiman, Jennifer Thompson, Seungmi Yang, Konstantin Vilchuck, Natalia Bogdanovich, Mikhail Hameza, Kate Tilling, et al. 2017. "Effects of promoting long-term, exclusive breastfeeding on adolescent adiposity, blood pressure, and growth trajectories: a secondary analysis of a randomized clinical trial." *JAMA pediatrics*, 171(7): e170698–e170698.

Martin, Richard M, Rita Patel, Michael S Kramer, Lauren Guthrie, Konstantin Vilchuck, Natalia Bogdanovich, Natalia Sergeichick, Nina Gusina, Ying Foo, Tom Palmer, et al. 2013. "Effects of promoting longer-term and exclusive breastfeeding on adiposity and insulin-like growth factor-I at age 11.5 years: a randomized trial." *Jama*, 309(10): 1005–1013.

Organization, World Health, et al. 2006. "WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development."

Patel, Rita, Emily Oken, Natalia Bogdanovich, Lidia Matush, Zinaida Sevkovskaya, Beverley Chalmers, Ellen D Hodnett, Konstantin Vilchuck, Michael S Kramer, and Richard M Martin. 2013. "Cohort profile: the promotion of breastfeeding intervention trial (PROBIT)." *International journal of epidemiology*, 43(3): 679–690.

Schroeder, Dirk G, Reynaldo Martorell, Juan A Rivera, Marie T Ruel, and Jean-Pierre Habicht. 1995. "Age differences in the impact of nutritional supplementation on growth." *The Journal of nutrition*, 125(4): 1051S–1059S.

Skugarevsky, Oleg, Kaitlin H Wade, Rebecca C Richmond, Richard M Martin, Kate Tilling, Rita Patel, Konstantin Vilchuck, Natalia Bogdanovich, Natalia Sergeichick,

George Davey Smith, et al. 2014. "Effects of promoting longer-term and exclusive breastfeeding on childhood eating attitudes: a cluster-randomized trial." *International journal of epidemiology*, 43(4): 1263–1271.

Victora, Cesar G, Rajiv Bahl, Aluísio JD Barros, Giovanny VA França, Susan Horton, Julia Krasevec, Simon Murch, Mari Jeeva Sankar, Neff Walker, Nigel C Rollins, et al. 2016. "Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect." *The Lancet*, 387(10017): 475–490.

Vidmar, Suzanna, John Carlin, Kylie Hesketh, and Tim Cole. 2004. "Standardizing anthropometric measures in children and adolescents with new functions for egen." *The Stata Journal*, 4(1): 50–55.

Walker, Susan P, Christine A Powell, Sally M Grantham-McGregor, John H Himes, and Susan M Chang. 1991. "Nutritional supplementation, psychosocial stimulation, and growth of stunted children: the Jamaican study." *The American journal of clinical nutrition*, 54(4): 642–648.

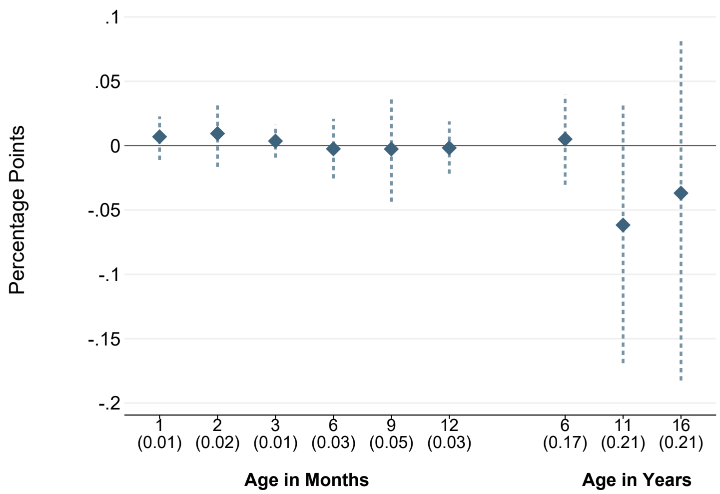
WHO, World Health Organization. 2011. "Exclusive breastfeeding for six months best for babies everywhere." URL: https://www.who.int/mediacentre/news/statements/2011/breastfeeding_20110115/en/, Accessed: 2020-01-07.

WHO, World Health Organization. 2020. "The WHO Child Growth Standards." URL: https://www.who.int/childgrowth/standards/en/, Accessed: 2020-07-16.

Yang, Seungmi, Richard M Martin, Emily Oken, Mikhail Hameza, Glen Doniger, Shimon Amit, Rita Patel, Jennifer Thompson, Sheryl L Rifas-Shiman, Konstantin Vilchuck, et al. 2018. "Breastfeeding during infancy and neurocognitive function in adolescence: 16-year follow-up of the PROBIT cluster-randomized trial." *PLoS medicine*, 15(4): e1002554.

A Supplementary Information (SI)

Figure A1
Attrition by Treatment Status

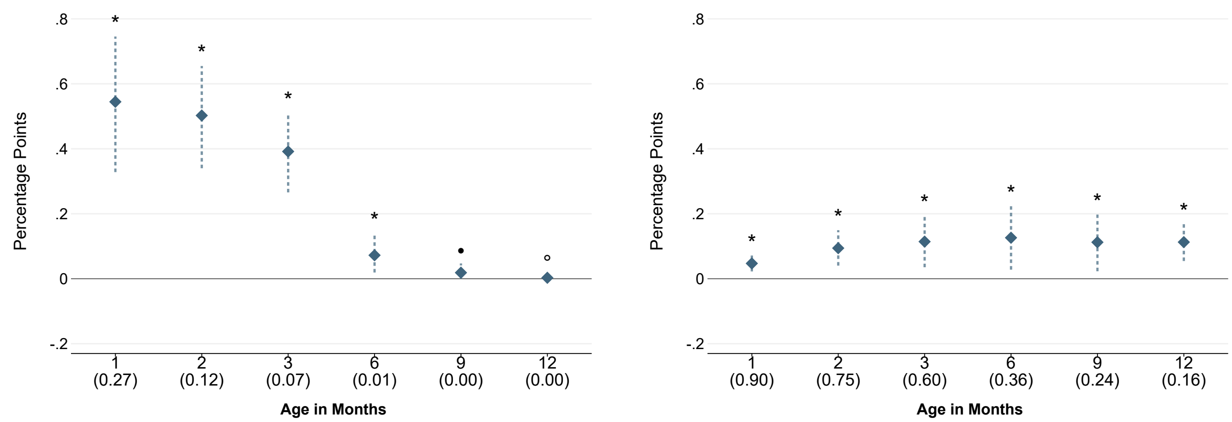


Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable. The number of non-missing observations for the treatment (control) arm for each wave (baseline, 1 month, 2 months, 3 months, 6 months, 9, months, 12 months, 6 years, 11 years, and 16 year) are 8,596 (8,178), 8,416 (8,078), 8,282 (7,985), 8,496 (8,123), 8,304 (7,916), 8,162 (7,771), 8,308 (7,921), 6,943 (6,788), 7,247 (6,472), 7,063 (6,491).

Figure A2

The Effect of Breastfeeding Promotion on Breastfeeding Exclusivity and Duration

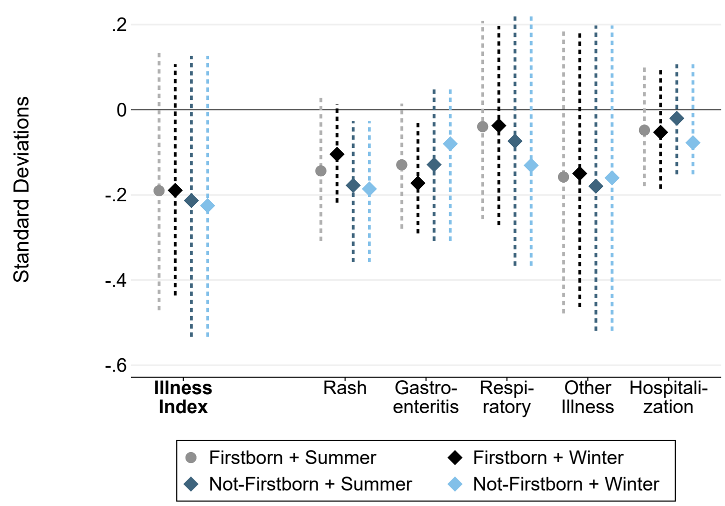
(a) Exclusive Breastfeeding **(b)** Any Breastfeeding



Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable.

Figure A3

The Effect of Breastfeeding Promotion on Infant Illness: Heterogeneity by Infectious Environment

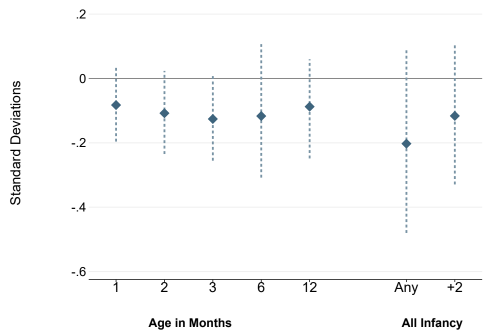


Note: Each estimate comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable.

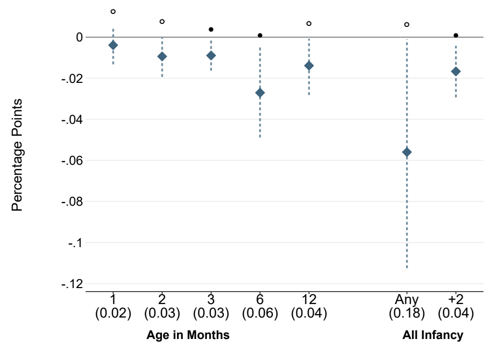
Figure A4

The Effect of Breastfeeding Promotion on Infant Illness by Age

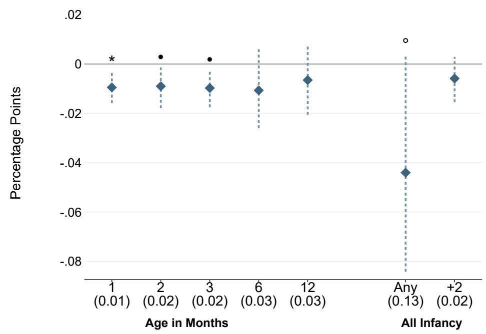
(a) Illness Index



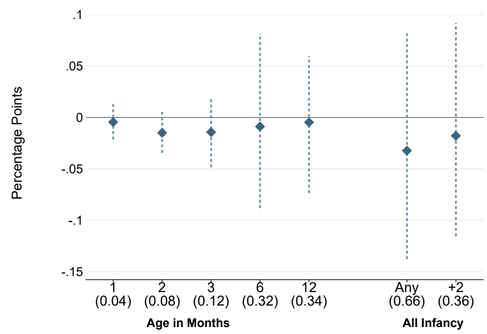
(b) Rash



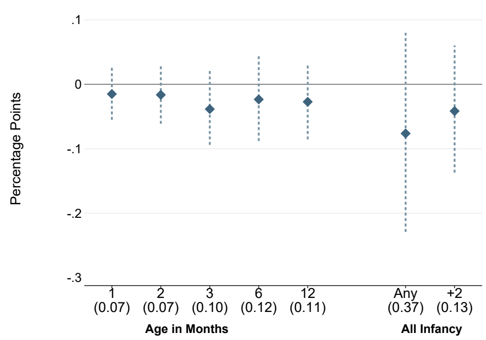
(c) Gastroenteritis



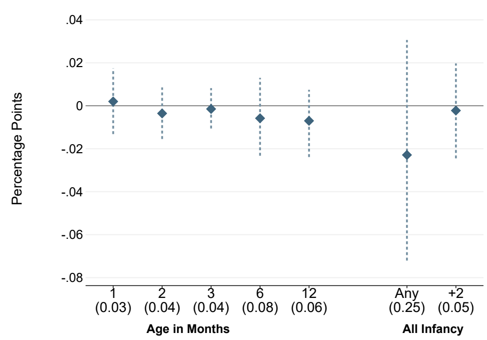
(d) Respiratory



(e) Other Illness



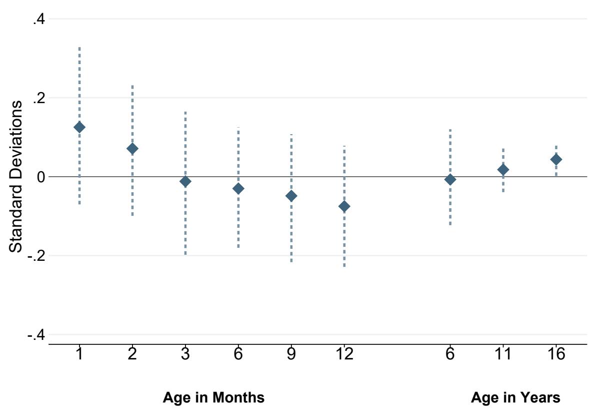
(f) Hospitalization



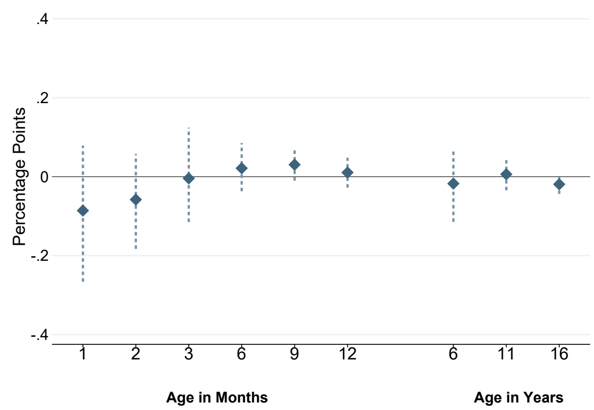
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$. The Illness Index for each age group is constructed as a weighted covariance summary index with indicators for any episode of each of the five illness measures since last health visit.

Figure A5
The Effect of Breastfeeding Promotion on Anthropometric Health Components

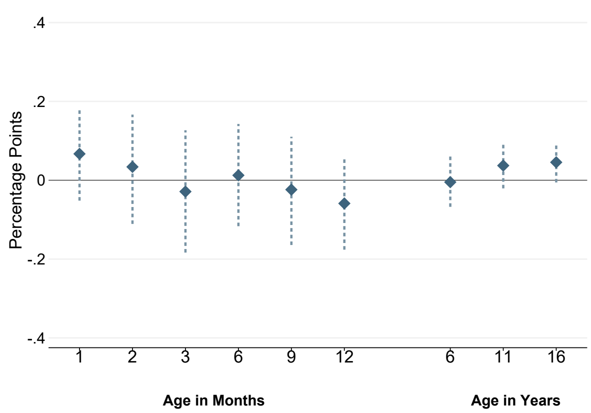
(a) Weight-for-length / BMI



(b) Probability of Underweight



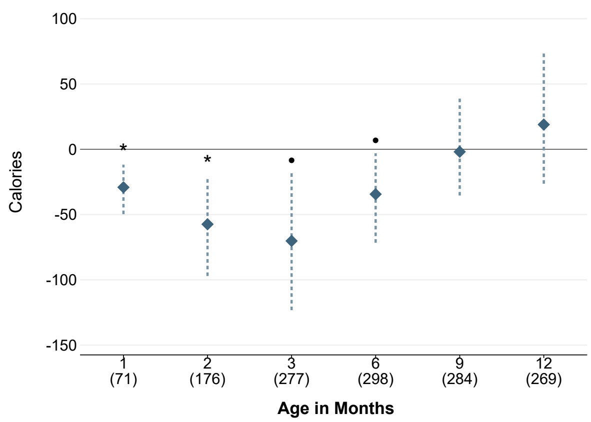
(c) Probability of Overweight



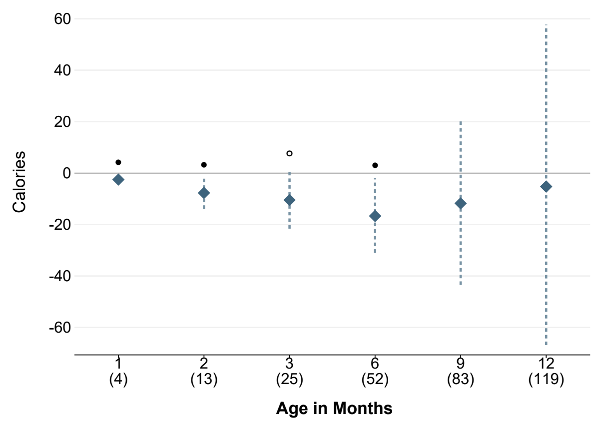
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

Figure A6
The Effect of Breastfeeding on Estimated Infant Liquid Calorie Intake

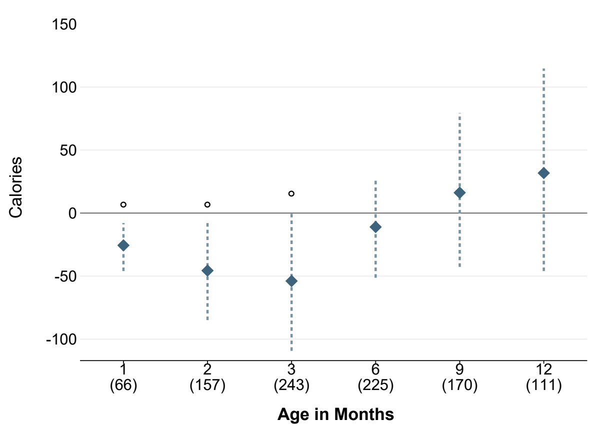
(a) Calories from Liquids (excl. BM)



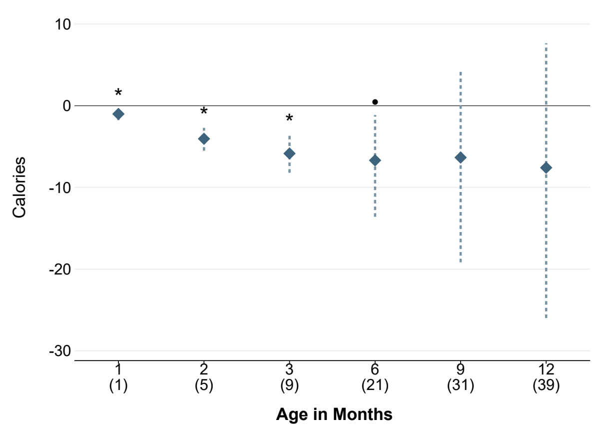
(b) Calories from Cow's Milk



(c) Calories from Formula

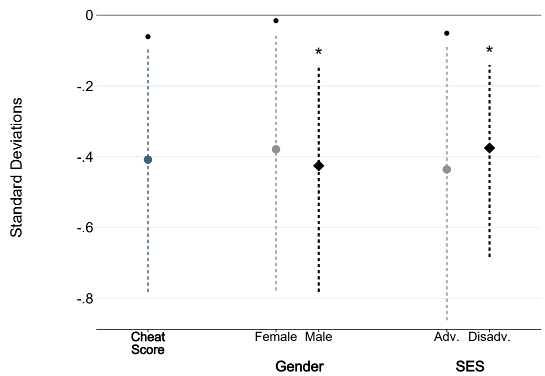


(d) Calories from Juices



Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ○$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

Figure A7
Treatment Effect on Children's Problematic Eating Attitudes



Note: The outcome is the Children's Eating Attitude Test (ChEAT), standardized with a mean of zero and standard deviation of one for the control group. A higher value indicates more problematic eating attitudes. For more details on the ChEAT measure, see Skugarevsky et al. (2014). The estimate most to the left comes from a separate regression as specified in equation (1), while the pairs of estimates come from regressions as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

Table A1
Descriptive Statistics and Balancing Test

	Control Mean (1)	Standard Deviation (2)	Difference (3)
Pregnancy and infant characteristics			
Male	0.52	0.50	-0.00
Birth weight (g)	3437.91	420.98	3.21
Birth length (cm)	52.02	2.18	-0.21
Head circumference at birth (cm)	34.79	1.64	0.39
Delivery complications	0.08	0.28	-0.00
Caesarean section	0.10	0.31	0.02
Gestational age at birth (weeks)	39.33	0.98	0.16
Firstborn	0.56	0.50	0.04
Smoking during pregnancy	0.02	0.13	0.01
Alcohol during pregnancy	0.02	0.13	0.03
Household characteristics			
Married at birth	0.91	0.29	-0.04***
Absent father	0.04	0.19	-0.01
Mother's Age (years)	24.44	4.91	-0.08
Father's Age (years)	27.34	5.15	-0.04
Number of other children in HH	0.58	0.86	-0.05
Previous exclusive breastfeeding $\geq$ 3 months	0.25	0.43	-0.00
Socio-economic disadvantage	0.36	0.48	-0.10*
Mother's education			
University degree	0.13	0.34	0.01
Adv. secondary or partial university	0.53	0.50	-0.06
Secondary degree	0.30	0.46	0.04
Incomplete secondary	0.03	0.17	0.01*
Father's education			
University degree	0.13	0.34	0.01
Adv. secondary or partial university	0.50	0.50	-0.10*
Secondary degree	0.32	0.47	0.09
Incomplete secondary	0.06	0.23	0.00
Mother's occupation			
Agriculture or industry	0.32	0.47	0.03
Services	0.45	0.50	-0.05
Housewife	0.13	0.34	-0.00
Unemployed	0.07	0.26	0.01
Student	0.03	0.17	0.01*
Father's occupation			
Agriculture or industry	0.50	0.50	0.05
Services	0.31	0.46	-0.06
Unemployed	0.14	0.35	-0.01
Student	0.01	0.10	0.01***
Observations	Total: 16774; Treatment: 8596; Control: 8178		

Note: The table shows descriptive statistics at baseline. Columns (1) and (2) report the mean and standard deviation of each background variable for the control group. Column (3) shows the difference (and indicates its significance level) between the treatment and control groups, when accounting for hospital pair fixed effects. Of the 34 variables, six are statistically different at the 10 percent level, suggesting that families in the treatment and control groups are relatively comparable within hospital pairs. Due to imbalance in some of the characteristics, we include the rich vector of individual controls as explained in Section 5.1. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

A.1 Experimental Content

The intervention was modeled on the Baby-Friendly Hospital Initiative (BFHI) developed by WHO and UNICEF and was a promotion and support of increased breastfeeding duration and exclusivity. The core aspects of the BFHI prescribe that the hospital should have a written breastfeeding policy that all staff should have the skills necessary to implement, mothers should be helped to initiate breastfeeding within half an hour after a normal birth, and unless medically indicated, newborn babies should have breast milk only. The head obstetrician and head pediatrician from each of the experimental maternity hospitals and polyclinics received the 18 hour BFHI lactation management training course organized by WHO. The aim of this course was to help hospitals transform their maternity facilities into baby-friendly institutions that implement the "Ten Steps to Successful Breastfeeding" and to assist them in implementing lasting policy changes.

The "Ten Steps to Successful Breastfeeding" are the following:

1. the hospital should have a written breastfeeding policy,
2. all staff should be trained in the skills necessary to implement the policy,
3. all pregnant women should be informed about the benefits and management of breastfeeding,
4. mothers should be helped to initiate breastfeeding within half an hour after a normal birth,
5. health workers should know how to assist in starting breastfeeding and how to maintain lactation during temporary separations,
6. unless medically indicated, newborn babies should have breast milk only,
7. babies should remain with their mothers 24 hours a day,
8. breastfeeding on demand should be encouraged,
9. pacifiers should not be given, and
10. the establishment of breastfeeding support groups should be fostered, and mothers should be referred to them on discharge (Kramer et al., 2000). Step 10 was implemented in the polyclinics.

After the 18 hour course, the trial participants organized and implemented training programs for midwives, nurses, physicians, and pediatricians working in their postpartum ward and polyclinic, respectively. The full implementation of the intervention required at least 12 months.

A.2 Outcome Measures

This section provides additional details on the data and construction of the outcome measures in the main analysis. The PROBIT study consists of four waves conducted in infancy and at ages 6.5, 11.5, and 16 years. The first wave includes baseline data and data from six routine health checkups when the children were approximately 1, 2, 3, 6, 9, and 12 months.

Age at Measurement in Wave 1 Previous papers have considered the outcomes for each subwave during infancy regardless of infant age at the visit. In contrast, we define the outcomes by actual infant age and consider the ages in months as follows: 1 (< 1.5 months), 2 (1.5–2.5), 3 (2.5-4.5), 6 (4.5-7.5), 9 (7.5-10.5), and 12 (10.5–14). The subwaves generally correspond closely to the actual age; however, we reclassify 0.8 percent of the observations.

Summary Indices For the construction of the summary indices, we reverse the signs of the components when necessary, so that all components in the indices indicate more favorable outcomes (except for the infant illness index; see below). We weight the standardized components by the covariance matrix and standardize the index so that the control group has a mean of zero and a standard deviation of one for each domain and age at survey. In addition to the indices, we also show the results for the individual components. From an economic perspective, identifying significant effects on an index among related outcomes more strongly signals robust differences that may indicate important changes in health or development. It also is another way to limit identification of false positives when evaluating many related outcomes.

Infant Illness Index At each visit during infancy, pediatricians asked mothers to detail any episodes of skin rash, gastrointestinal illness, respiratory illness, other illness, and hospitalization since the previous visit. The infant illness index includes five indicators for whether the infant experienced any of the five outcomes at least once during infancy, and another five indicators representing multiple reports of the same outcome. In contrast to all other indices, a lower value on the infant illness index is better and indicates being less ill.

Anthropometric Index During all waves (including each of the subwaves during infancy), a pediatrician conducted anthropometric measurements. We consider the measures of height and weight and standardize these to age-specific z-scores based on the 2000 growth charts for the United States from the Centers for Disease Control and Prevention (Vidmar et al., 2004). We then construct an anthropometric index for each wave which includes weight-for-age, height-for-age, body mass index (BMI)-for-age, and indicators for being underweight and overweight. We reverse the sign for the last three components when constructing the index. For measurements during infancy, we use length instead of height and weight-for-length instead of BMI. We define underweight (overweight) as being below the 15^{th} (above the 85^{th}) percentile according the WHO weight-for-length growth standards (Organization et al., 2006) for infants less than one year old. For older children, we rely on the overweight and underweight measures developed in Cole et al. (2000) and implemented in Vidmar et al. (2004).

We exclusively consider the height and weight instead of other anthropometric measurements, such as head circumference, for two reasons. First, audit test-retest correlations at age 6.5 were very high for height (0.84) and BMI (0.89) but substantially lower for head circumference (0.65) (Patel et al., 2013). Second, height and weight are consistently measured across all waves in contrast to other measurements. Moreover, we decided not to consider blood pressure, as audit test-retest correlations are particularly low (around 0.50); in results not reported, we do not find an effect on blood pressure, but the confidence intervals are large.

Infant Feeding During the routine health visits in wave 1, study pediatricians assessed infant feeding using standard questionnaires. Previous published studies have only reported the effect of the intervention on breastfeeding exclusivity and duration. We focus on maternal reports of the number of times and the total quantity of breast milk (including expressed and donor milk), infant formula, cow's milk (including other types of animal milk), water, juices or other liquids, and solid food (including cereals) the child received during the previous 24 hours at each visit. Less than 1.00 (0.01) percent of breast milk feedings are expressed (donor) milk, with the rest being breastfed at the breast.

Table A2

Descriptive Statistics for the Control Group: Anthropometric Measures

	Weight-for-Age (1)	**Height-for-Age** (2)	**BMI-for-Age** (3)	**Underweight** (4)	**Overweight** (5)
1 Month	−0.11	0.10	−0.63	0.22	0.13
	(0.80)	(0.84)	(1.11)	(0.41)	(0.33)
2 Months	0.07	−0.01	−0.26	0.17	0.18
	(0.81)	(0.89)	(1.10)	(0.37)	(0.38)
3 Months	0.25	0.14	−0.06	0.13	0.18
	(0.83)	(0.94)	(1.08)	(0.34)	(0.39)
6 Months	0.44	0.23	0.54	0.03	0.36
	(0.87)	(1.00)	(0.98)	(0.18)	(0.48)
9 Months	0.47	0.28	0.81	0.01	0.51
	(0.84)	(0.95)	(0.90)	(0.11)	(0.50)
12 Months	0.51	0.34	0.93	0.01	0.61
	(0.84)	(0.91)	(0.89)	(0.07)	(0.49)
6 Years	0.04	0.21	−0.09	0.12	0.09
	(0.93)	(0.93)	(1.02)	(0.33)	(0.29)
11 Years	−0.01	0.26	−0.10	0.13	0.13
	(1.00)	(0.98)	(1.03)	(0.34)	(0.34)
16 Years	0.23	0.31	0.03	0.09	0.15
	(0.89)	(0.93)	(0.92)	(0.29)	(0.35)

Note: The table shows means and standard deviations (in parenthesis) for anthropometric measured from age 1 month to 16 years. Columns (1)-(3) show age-standardized z-scores for weight, height, and body mass index (BMI), while columns (4)-(5) report the share of children who are under- and overweight respectively. Underweight (overweight) measures are based on being below (above) the 15^{th} (85^{th}) percentile according to the WHO Child Growth Standards (WHO, 2020).

Table A3

Descriptive Statistics for the Control Group: Infant Illnesses

	Rash	**Gastroenteritis**	**Respiratory Illness**	**Other Illness**	**Hospitalization**
	(1)	(2)	(3)	(4)	(5)
Any Illness at ...					
1 Month	0.02	0.01	0.04	0.07	0.03
	(0.13)	(0.12)	(0.19)	(0.25)	(0.18)
2 Months	0.03	0.02	0.08	0.07	0.04
	(0.16)	(0.13)	(0.28)	(0.25)	(0.19)
3 Months	0.03	0.02	0.12	0.10	0.04
	(0.16)	(0.13)	(0.32)	(0.30)	(0.20)
6 Months	0.06	0.03	0.32	0.12	0.08
	(0.24)	(0.18)	(0.47)	(0.32)	(0.26)
9 Months	0.05	0.04	0.36	0.11	0.08
	(0.23)	(0.18)	(0.48)	(0.31)	(0.26)
12 Months	0.04	0.03	0.34	0.11	0.06
	(0.21)	(0.18)	(0.47)	(0.31)	(0.24)
... Illness during first year					
Any	0.20	0.13	0.67	0.37	0.25
	(0.40)	(0.34)	(0.47)	(0.48)	(0.43)
At Least 2	0.07	0.04	0.45	0.18	0.09
	(0.26)	(0.20)	(0.50)	(0.39)	(0.29)

Note: The table shows means and standard deviations (in parenthesis) for indicators of any episode of the specific infant illness measure (as indicated in the column head) since last regular baby health visit and of having any or multiple illnesses during the first year.

Thus, we refer to the combined group of breast milk as breastfeeding interchangeably. For breast milk and solid foods, mothers did not report the quantity but only the number of times the infant received that type of food. We define a child as being exclusively breastfed when he or she only receives breast milk and nothing else.

We construct the infant feeding outcomes in Figure 3 as follows:

- *Calories from Formula* and *Calories from Cow's Milk* respectively indicate the calories the infant received from infant formula and cow's or other types of animal milk, calculated based on the quantity received and assuming that both types of liquid contain 65 kcal per 100 ml.

- *Calories from Juices* indicates the calories the infant received from juices and other liquids, assuming that two-thirds being apple juice (45 kcal per 100 ml) and one-third tea (0 kcal per 100 ml). From anecdotal evidence, *juices* would typically be apple juice and *other liquids* would be black tea.

- *Calories from Liquids (excl. BM)* is the total calorie intake from liquids (formula, cow's milk, and juices) excluding breast milk.

- *Calories/100 ml Liquids (excl. BM)* is constructed as *Calories from Liquids (excl. BM)* divided by the total volume of liquids excluding breast milk.

- *Total Volume Liquids (excl. BM), ml* indicates the total quantity of infant formula, cow's milk, water, juices, and other liquids measured in ml.

- *Volume of All Liquids (incl. BM), ml* indicates the total volume of liquids the child received including breast milk. For breastfed children at 1–3 months, we estimate breast milk intake based on the recommended daily calorie intake formula $\text{Calorie Intake} = 89 \times \text{Weight in Kg} + 75$ (Lupton et al., 2002), subtract their calorie intake from all other liquids, and calculate the volume assuming that breast milk contains 65 kcal per 100 ml. For breastfed children at 6–12 months, we assume that each breast milk feeding contains 175 ml.

- *Calories of All Liquids (incl. BM)* indicates the total calorie intake from liquids including breast milk, where calories from breast milk is estimated as described above.

A.3 Mediation Analysis

We conduct the mediation analysis following Gelbach (2016) and Heckman, Pinto and Savelyev (2013). This analysis provides insights on the importance of the nutritional mechanism (mediator) in explaining the effect of the breastfeeding promotion intervention of infant weight gain. We perform the mediation analysis by decomposing the treatment effect on weight-for-age obtained from equation (1) ($Y_{iph} = \gamma_0 + \gamma_1 Treatment_h + Z_i'\delta + \theta_p + \varepsilon_{iph}$) in the following way:

$$\frac{dY_{iph}}{dTreatment_h} = \sum \frac{\partial Y_{iph}}{\partial M_i} \frac{\partial M_i}{\partial Treatment_h} + R, \tag{2}$$

where Y_{iph} is the weight of child i, born at hospital pair p, and hospital h. $Treatment_h$ is the treatment indicator, M_i is a vector of k mediators, and R is the unexplained part of the treatment effect. In line with this decomposition, we estimate two additional specifications. First, we estimate the conditional outcome equation (equation (1)) augmented with the vector M_i:

$$Y_{iph} = \alpha_0 + \alpha_1 Treatmemt_h + M_i'\phi + Z_i'\delta^{'} + \theta_p^{'} + \varepsilon_{iph}^{'}. \tag{3}$$

Second, we separately estimate the treatment effect of the intervention on each mediator $j \in k$:

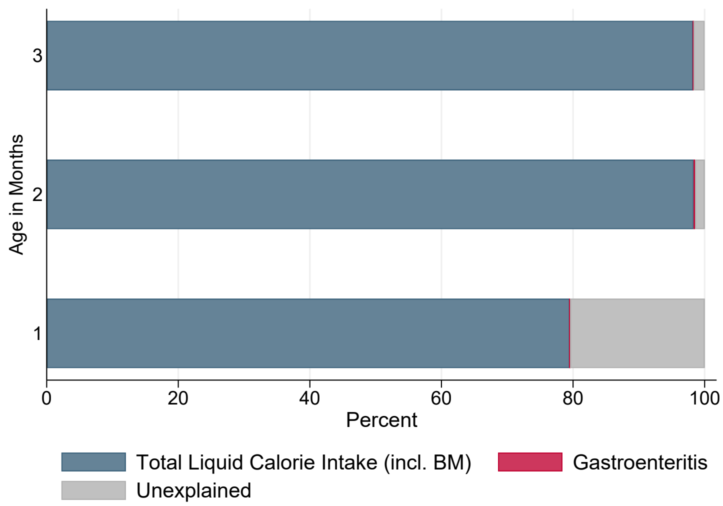
$$M_i^j = \beta_0 + \beta_1^j Treatment_h + Z_i'\delta^{''} + \theta_p^{''} + \varepsilon_{iph}^{''}. \tag{4}$$

The contribution of each mediator $j \in k$ is then computed as the ratio $\frac{\phi^j \times \beta_1^j}{\gamma_1}$. The unexplained part, R, results from $R = 1 - \sum_{j=1}^{k} \frac{\phi^j \times \beta_1^j}{\gamma_1}$.

In the mediation analysis presented in Figure 4, we include as mediator the total calorie intake from all liquids including breast milk through the age at weight measurement. For age one month we multiply the estimate for calorie intake from liquids including breastmilk by the age in days at visit. For age two to three months we take the age difference in days between the previous and current visit and again multiply it by the estimate for calorie intake from liquids including breastmilk. The mediator at two and three months is the cumulative sum of these measures at the respective point in time.

Because we observe a statistically significant decrease in the probability of gastroenteritis in treated infants in the first three months of life (SI Figure A4), we also include, as a robustness check, the occurrence of this type of illness as a mediator. More precisely, we include indicators for any episode of gastrointestinal illness up until the date of weight measurement. Despite gastroenteritis either involves diarrhea or vomiting and thereby a reduced calorie uptake, we do not find evidence that the decrease in episodes of gastrointestinal illness explains any meaningful effect of the intervention on weight gain (SI Figure A8).

Figure A8
Mediation Analysis on Weight-for-Age



Note: This figure shows the results of the mediation analysis, decomposing the overall treatment effect of weight-for-age at 1, 2, and 3 months. We use total caloric intake from all liquids including breast milk as the mediator and an indicator for any instance of gastrointestinal illness up until the date of weight measurement.

A.4 Heterogeneity

In this appendix, we report all the main results when allowing for heterogeneous response on two separate dimensions. First, we consider whether the treatment effects differ by gender. Second, we consider whether the treatment effects differ by socioeconomic status (SES). We define socioeconomic disadvantage to be the case when neither of the parents has a university degree nor works in a non-manual occupation (services), meaning the group of all other parents represents socioeconomic advantage.

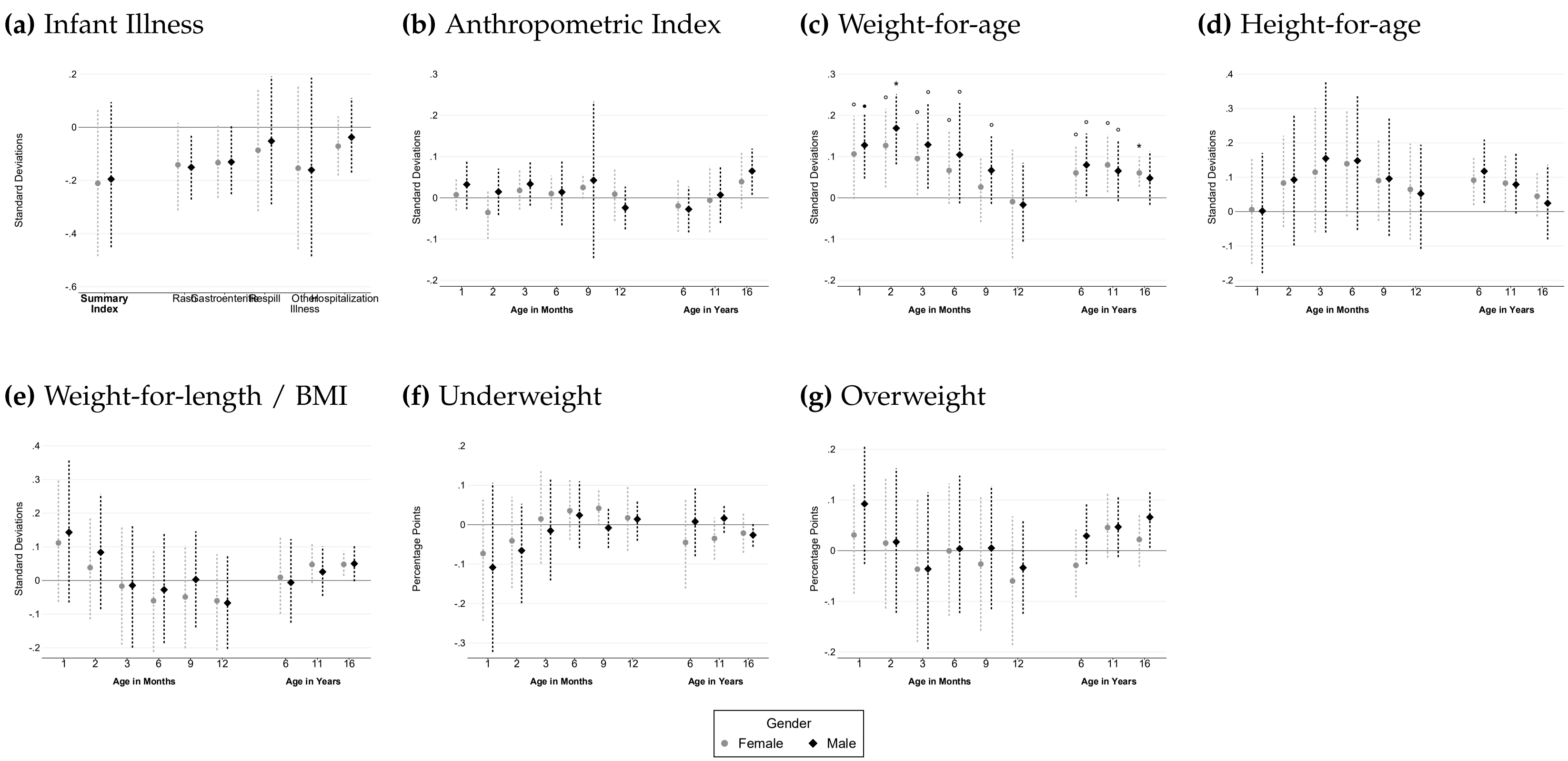
To estimate heterogeneous effects, we modify the empirical specification given in equation (1). We estimate fully interacted models, in which we interact all control variables with the particular dimensions of interest. For the estimation of heterogeneity with respect to gender, the specification is the following (with the model for SES being completely analogous):

$$\begin{aligned} Y_{iph} = &\gamma_0^{female} Female + \gamma_1^{female} Treatment_h \times Female + \gamma_1^{male} Treatment_h \times Male + \\ &Female \times Z_i' \delta^{female} + Male \times Z_i' \delta^{male} + \theta_p^{female} + \theta_p^{male} + \varepsilon_{iph}, \end{aligned} \tag{5}$$

where *Female* (*Male*) takes the value 1 (0) for girls and 0 (1) for boys. The remaining variables are similar to those specified for specification (1).

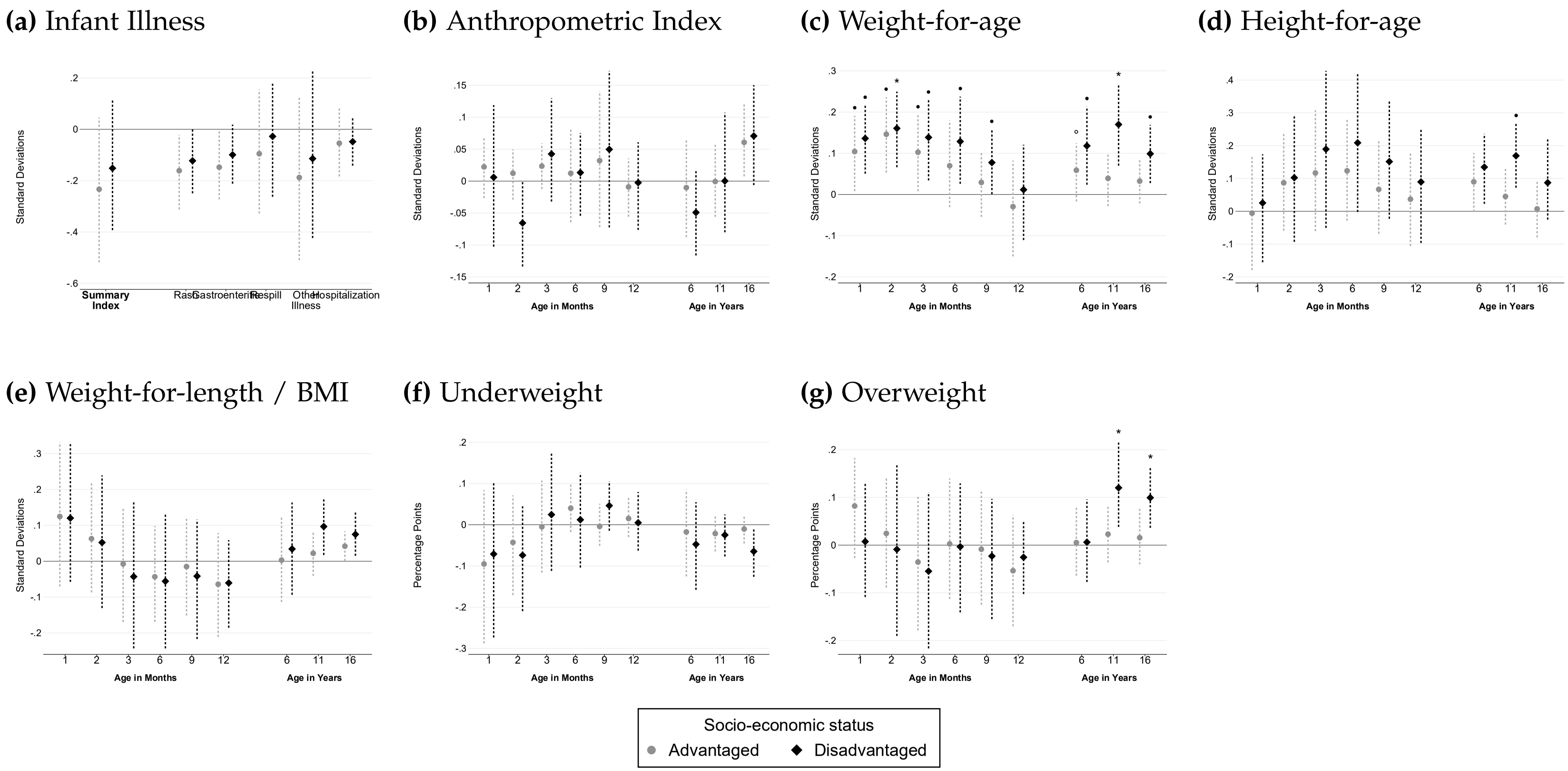
As shown in the following figures, there is little evidence of meaningful heterogeneity in any of the outcomes we study by either gender or SES.

Figure A9
The Effect of Breastfeeding Promotion on Infant Illness and Childhood Anthropometric Health: Heterogeneity by Gender



Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

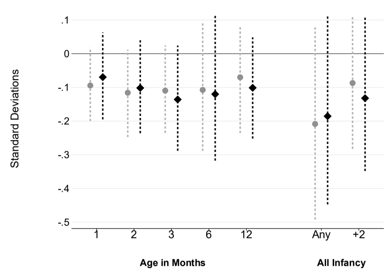
Figure A10
The Effect of Breastfeeding Promotion on Infant Illness and Childhood Anthropometric Health: Heterogeneity by Socioeconomic Status



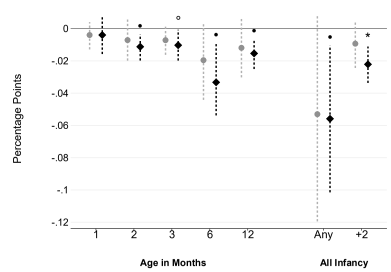
Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

Figure A11
The Effect of Breastfeeding Promotion on Infant Illness: Heterogeneity by Gender

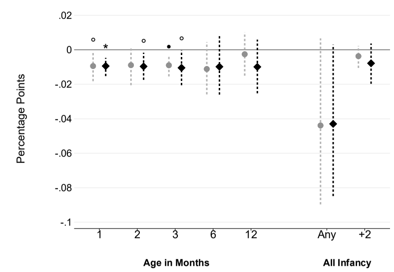
(a) Illness Index



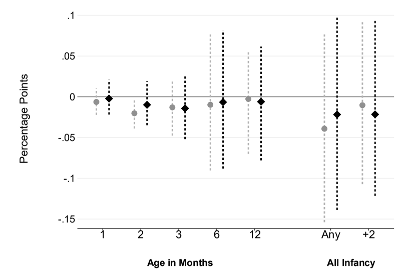
(b) Rash



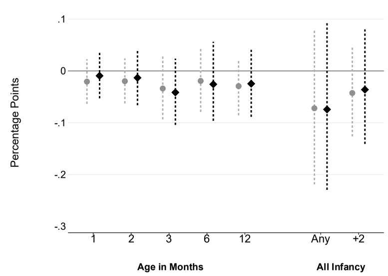
(c) Gastroenteritis



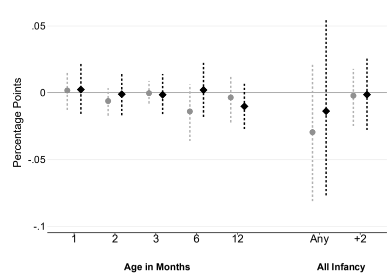
(d) Respiratory



(e) Other Illness



(f) Hospitalization



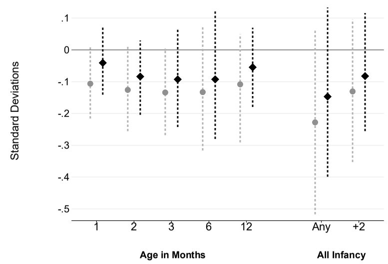
Gender
● Female ◆ Male

Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

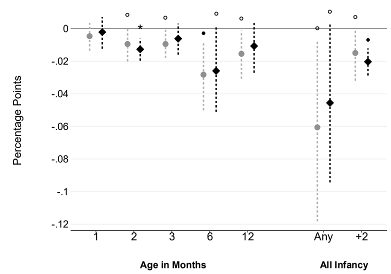
Figure A12

The Effect of Breastfeeding Promotion on Infant Illness: Heterogeneity by Socioeconomic Status

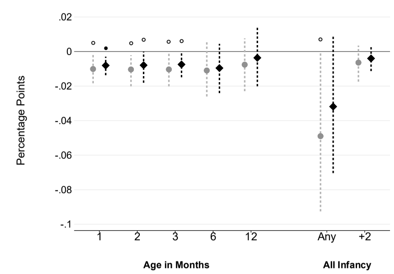
(a) Illness Index



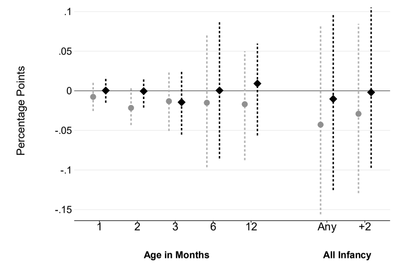
(b) Rash



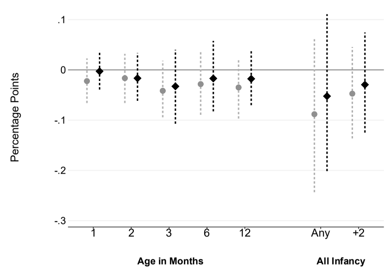
(c) Gastroenteritis



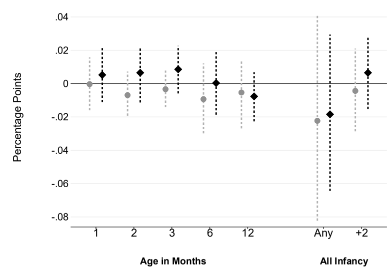
(d) Respiratory



(e) Other Illness



(f) Hospitalization

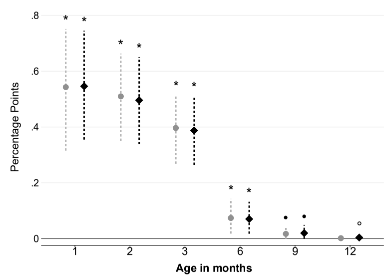


Socio-economic status
● Advantaged ◆ Disadvantaged

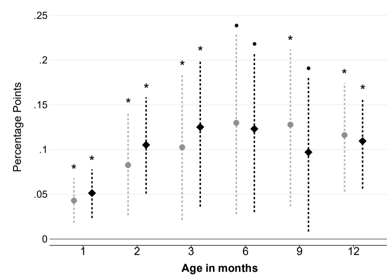
Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

Figure A13
The Effect of Breastfeeding Promotion on Infant Feedings: Heterogeneity by Gender

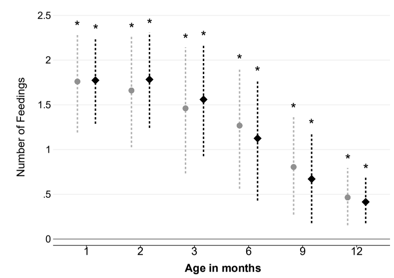
(a) Exclusive Breastfeeding



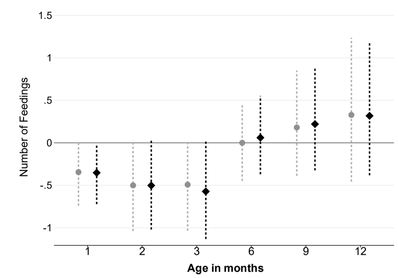
(b) Any Breastfeeding



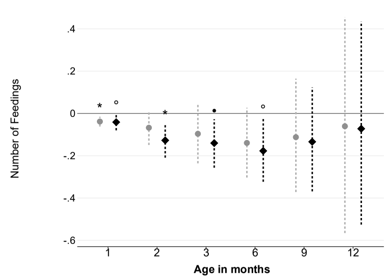
(c) Freq. of Breastfeeding



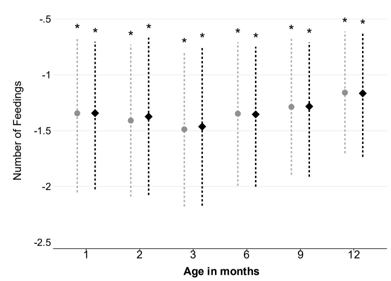
(d) Freq. of Infant Formula



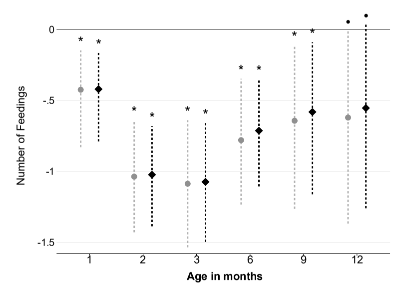
(e) Freq. of Cow's Milk



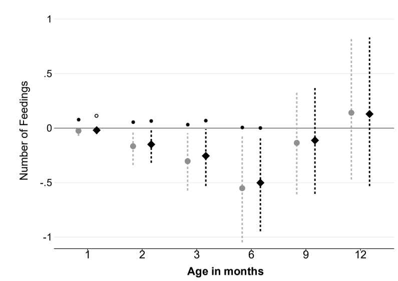
(f) Freq. of Water



(g) Freq. of Juice / Liquids



(h) Freq. of Solid food

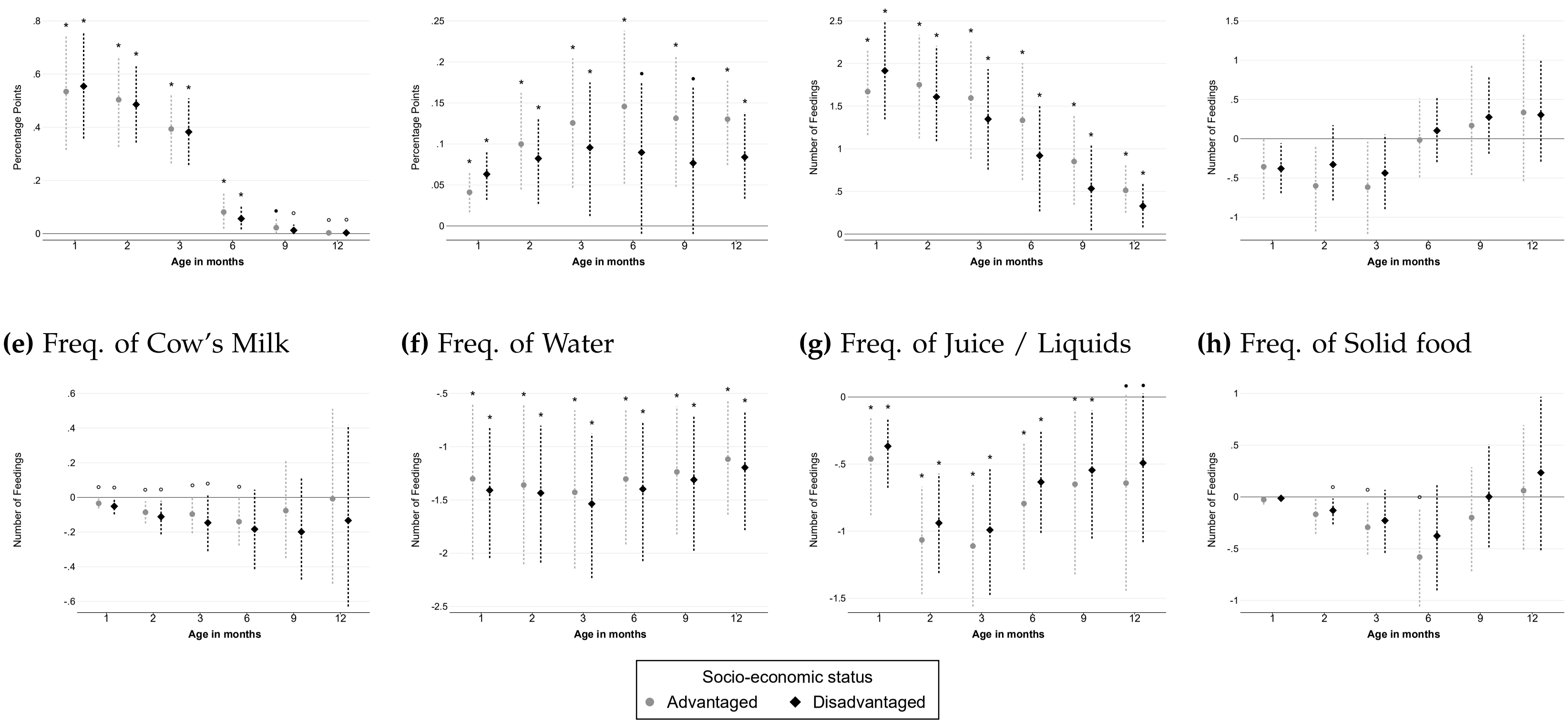


Gender
● Female ◆ Male

Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ○$p < 0.10$, ●$p < 0.05$, *$p < 0.01$.

Figure A14

The Effect of Breastfeeding Promotion on Infant Feedings: Heterogeneity by Socioeconomic Status



(a) Exclusive Breastfeeding **(b)** Any Breastfeeding **(c)** Freq. of Breastfeeding **(d)** Freq. of Infant Formula

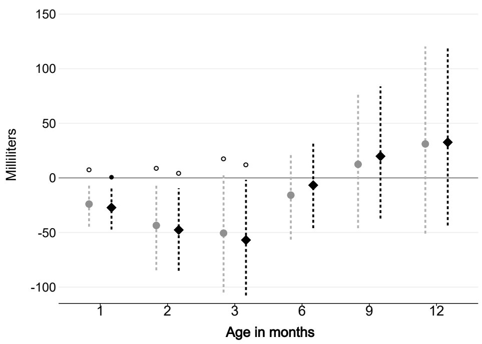
(e) Freq. of Cow's Milk **(f)** Freq. of Water **(g)** Freq. of Juice / Liquids **(h)** Freq. of Solid food

Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

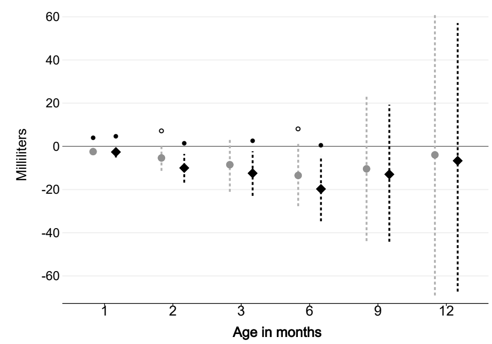
Figure A15

The Effect of Breastfeeding Promotion on Estimated Infant Liquid Calorie Intake, excluding Breast Milk (BM): Heterogeneity by Gender

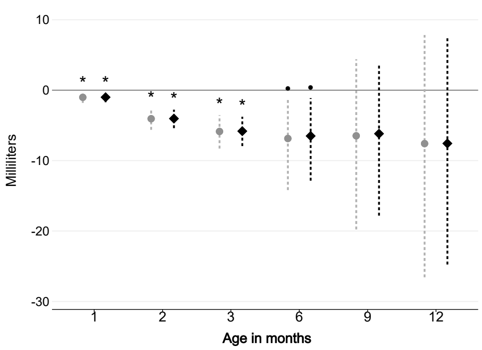
(a) Calories from Formula



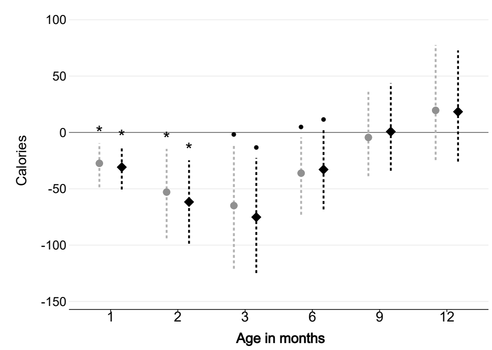
(b) Calories from Cow's Milk



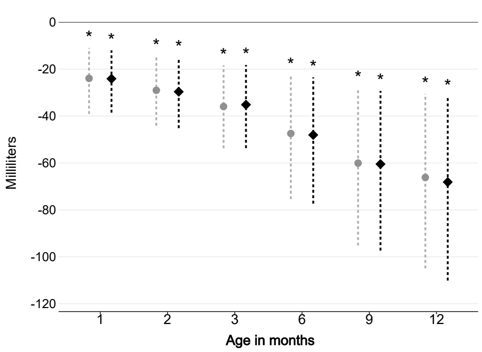
(c) Calories from Juices



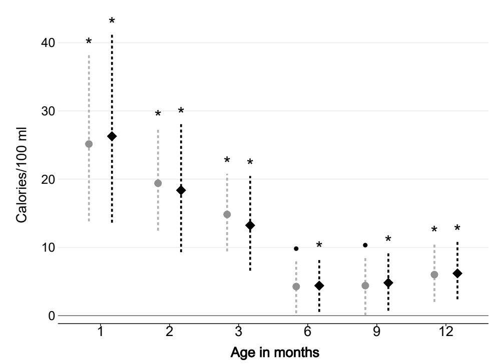
(d) Calories from Liquids (excl. BM)

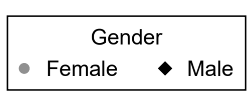


(e) Volume of Water, ml



(f) Calories/100 ml Liquids (excl. BM)



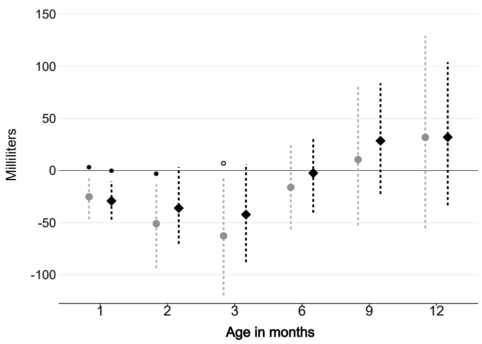


Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

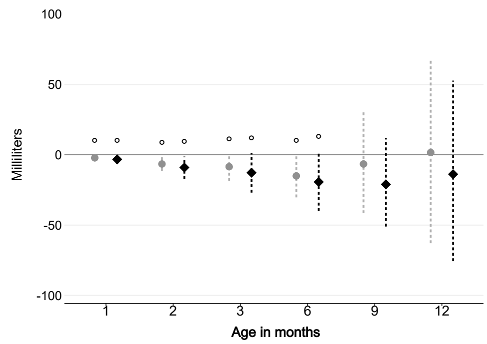
Figure A16

The Effect of Breastfeeding Promotion on Estimated Infant Liquid Calorie Intake, excluding Breast Milk (BM): Heterogeneity by Socioeconomic Status

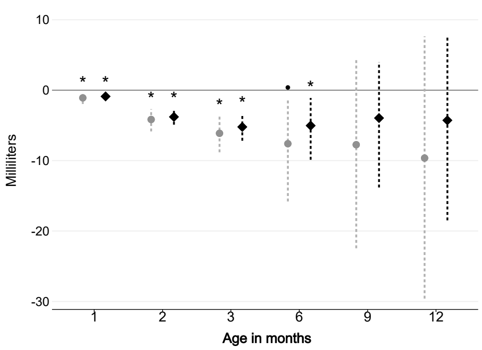
(a) Calories from Formula



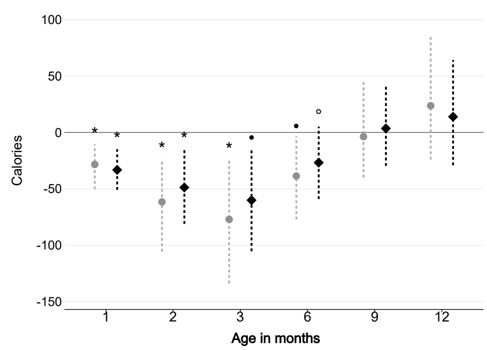
(b) Calories from Cow's Milk



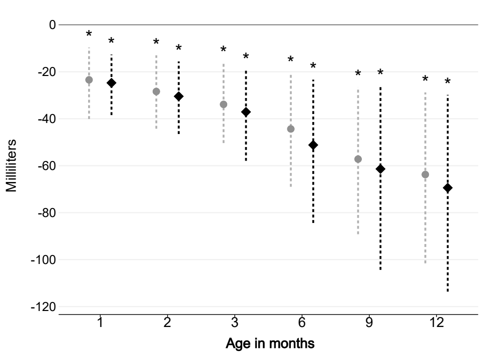
(c) Calories from Juices



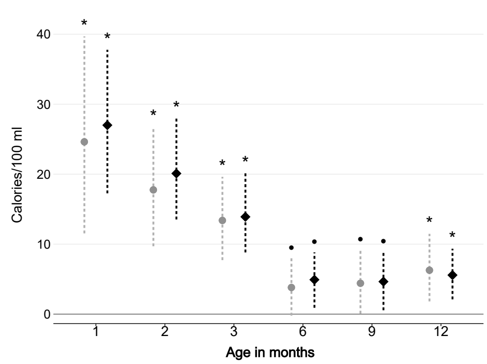
(d) Calories from Liquids (excl. BM)

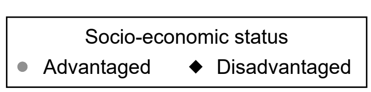


(e) Volume of Water, ml



(f) Calories/100 ml Liquids (excl. BM)

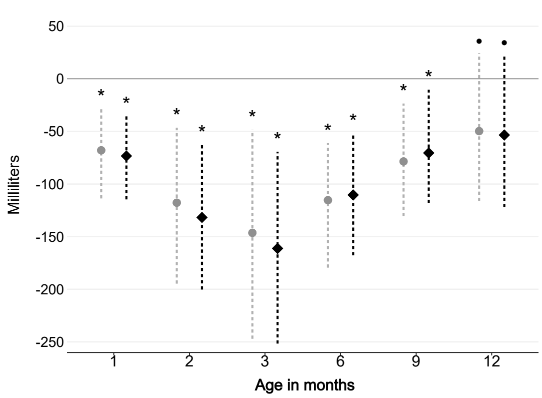




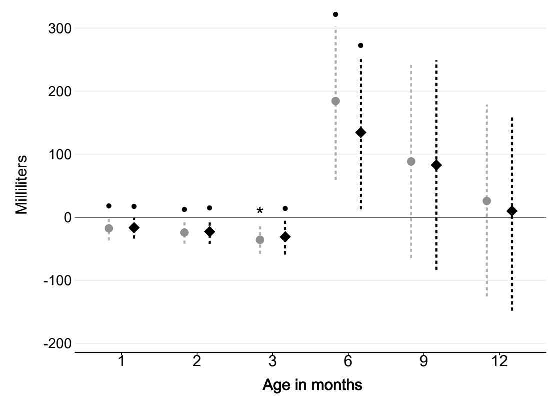
Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

Figure A17
The Effect of Breastfeeding Promotion on Estimated Infant Liquid Calorie Intake, including Breast Milk: Heterogeneity by Gender

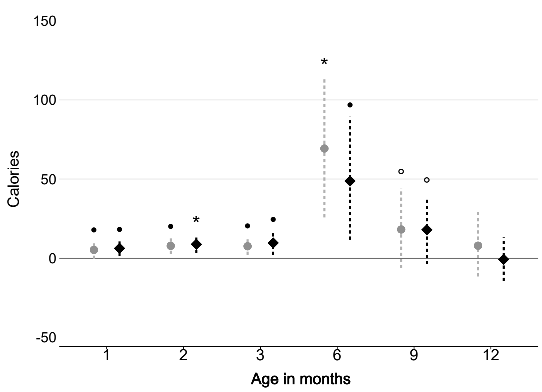
(a) Total Volume Liquids (excl. BM), ml



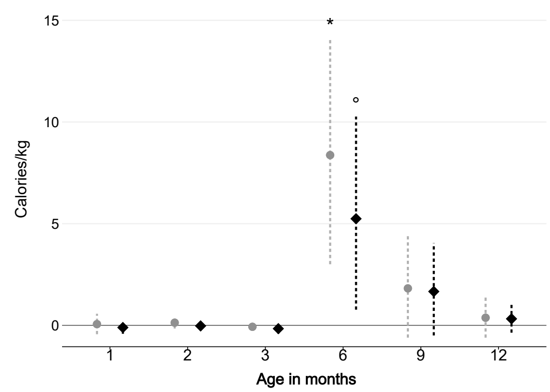
(b) Volume of All Liquids (incl. BM), ml

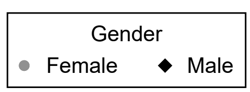


(c) Calories of All Liquids (incl. BM)



(d) Calories of All Liquids (incl. BM) per kg Body Weight

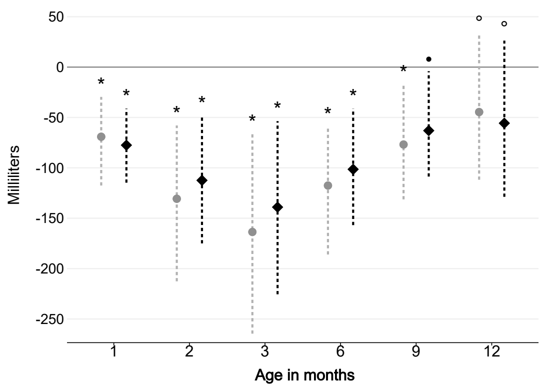




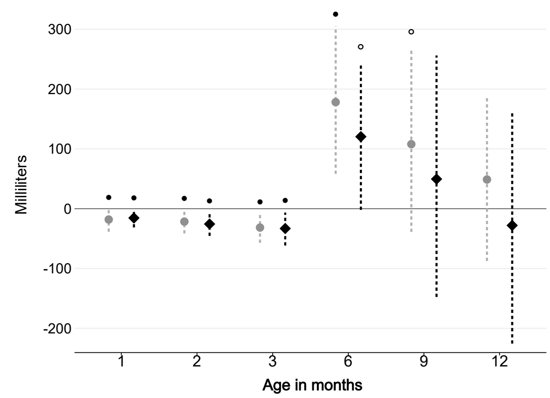
Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

Figure A18
The Effect of Breastfeeding Promotion on Estimated Infant Liquid Calorie Intake, including Breast Milk: Heterogeneity by Socioeconomic Status

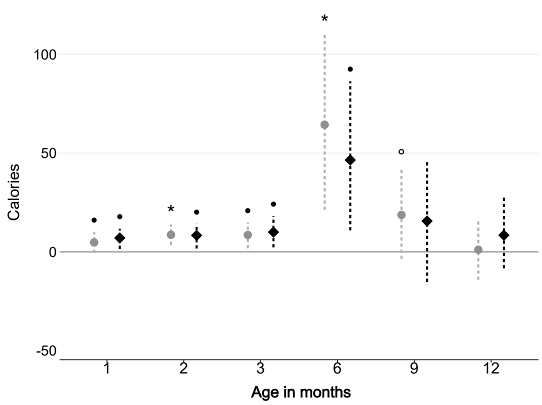
(a) Total Volume Liquids (excl. BM), ml



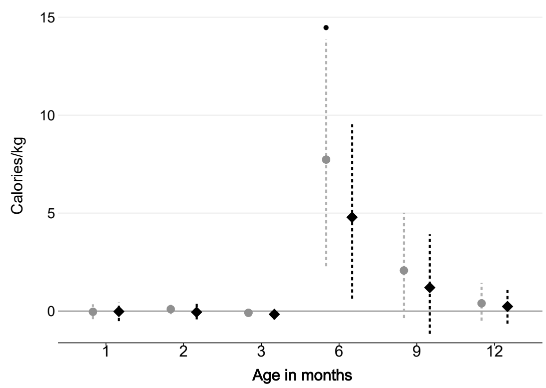
(b) Volume of All Liquids (incl. BM), ml

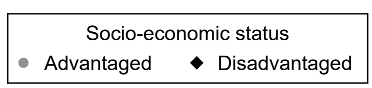


(c) Calories of All Liquids (incl. BM)



(d) Calories of All Liquids (incl. BM) per kg Body Weight





Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

A.5 Child Socioemotional and Cognitive Skills

This subsection analyzes the effect of the breastfeeding promotion intervention on child socioemotional and cognitive skills. Child socioemotional skills are measured at age 6 based on parent and teacher reports on a series of questions summarized into seven standard subscales. The academic performance index at age 6 includes teachers' ratings of each child's performance in reading, writing, and mathematics. The performance IQ index at age 6 includes block design and matrix reasoning from the Wechsler Abbreviated Scales of Intelligence (WASI). The cognitive skills index at age 16 includes scores from the NeuroTrax cognitive tests for seven domains. All indices are scaled so a larger value is more beneficial. We first provide more details about the exact measurements and thereafter, we present the results following a similar estimation strategy as for the main results.

A.5.1 Data and Outcome Variables

Socioemotional Indices We assess child socioemotional skills at age 6 through parent and teacher reports to the Strengths and Difficulties Questionnaire (SDQ) and to supplemental behavioral questions taken from the Canadian National Longitudinal Survey of Children and Youth (NLSCY). Using the SDQ answers, we construct the five standard subscales for emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior for parents and teachers, separately; we reverse the sign of the first four subscales so that a higher score is more favorable. Based on the questions from the NLSCY, we construct two summary indices capturing respectively externalizing and internalizing behavioral problems; we reverse the sign of these measures so that a higher score is more favorable. SI Table A4 illustrates the exact questions used from the NLSCY. We then construct the parent reported socioemotional skills index by including the seven parent-reported components: the five SDQ subscales and the two NLSCY behavioral problem indices. Similarly, we construct the teacher reported socioemotional skills index by including the corresponding seven teacher-reported components.

Academic Performance Index At age 6, teachers also rated each child's academic performance on a five-point Likert scale as far below, somewhat below, at, somewhat above, or far above his or her grade level (Kramer et al., 2008*b*). We use an index for aca-

Table A4
NLSCY Behavioral Questions

	Questionnaire Respondent	
	Parents	**Teachers**
Externalizing Behavioral Problems		
Hits, bites, or pinches other children	✓	✓
Reacts in an aggressive manner when contradicted or teased	✓	✓
Scares other children to get what he wants	✓	✓
Does not seem to feel badly after misbehaving	✓	✓
When mad at someone, says bad things behind the other's back	✓	✓
Reacts in an aggressive manner when contradicted or teased		✓
Is impulsive		✓
Has difficulty waiting for his turn in games		✓
Internalizing Behavioral Problems		
Is nervous, high-strung, or tense	✓	✓
Is not as happy as other children	✓	✓
Has no energy, feels tired	✓	✓
Is shy with children he/she does not know	✓	✓
Clings to adults or is too dependent	✓	✓
Gets very upset when separated from parents	✓	✓
Does not want to sleep alone	✓	
Has trouble enjoying himself	✓	
Takes a long time getting used to being with children he does not know	✓	
Readily approaches children he does not know (reverse sign)	✓	

Note: This table shows the exact questions used for the construction of the two NLSCY behavioral problem indices. The respondent answered whether the statements are not true, somewhat true, or certainly true.

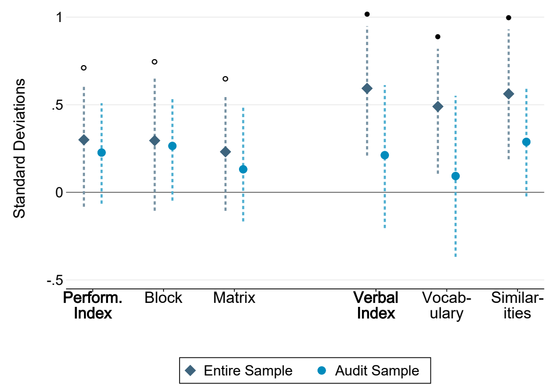
demic performance, including ratings in reading, writing, and mathematics. We observe teacher ratings for about 63 percent of the children; most of the children without a rating had not started school at the time of follow-up. Teachers were unaware of the children's participation in the study (and therefore also of their treatment status) and had only 3–4 PROBIT children on average (Kramer et al., 2008*a*). Consequently, any potential bias in the teacher reports should be unrelated to the treatment status.

WASI Measures at Age 6 At age 6, study pediatricians assessed children's cognitive development using the Wechsler Abbreviated Scales of Intelligence (WASI). The WASI consists of four subtests (vocabulary, similarities, block design, and matrix reasoning) and takes about 30 minutes to administer. One pediatrician assessed the children in 24 polyclinics and two did so in the remaining 7 high-volume polyclinics, resulting in a high intraclass correlation coefficient (ICC) for cognitive scores of 0.31. Unfortunately, the pediatricians were not blind to treatment (Kramer et al., 2008*b*), leading to a potential measurement bias. Kramer et al. (2008*b*) conducted an audit of 190 children and found the lowest test-retest correlations for verbal IQ (vocabulary and similarities) of 0.62, for which the scope for subjective assessment of the child might be largest, compared to correlations of 0.71 in performance IQ (block design and matrix reasoning).

SI Figure A19 shows the effect of the PROBIT intervention on the indices for performance IQ and verbal IQ as well as the four underlying components, using the audit sample and the entire sample. From this, it is clear that the point estimates for verbal IQ and its components in the main sample are much greater than in the audit sample, in the magnitude of 2–5 times. Meanwhile, the differences between the point estimates for performance IQ and its components in the audit and main samples are small. These differences are in line with what Kramer et al. (2008*b*) report. Thus, while the audit data replicate the estimate for performance IQ well, the study pediatricians' assessments of verbal IQ in the experimental sites seem to be upward biased, leading to an overestimate of the effect of the intervention. For this reason, we chose to only consider an index of performance IQ in the following analysis. However, we still caution the interpretation of the WASI results.

Cognitive Skills Index at Age 16 At age 16, children's neurocognitive function was assessed using a computerized battery of the NeuroTrax cognitive tests. The tests were self-administered to minimize potential measurement bias caused by non-blinding of

Figure A19
Estimates on WASI in the Audit and Main Samples



Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

the pediatricians and within polyclinic correlations, resulting in ICCs as small as 0.02 (Yang et al., 2018). The test battery consists of 10 short subtests that assess both verbal and nonverbal domains of cognitive function. From this, we have standardized scores for seven domains, measuring memory, executive function, visual–spatial perception, verbal function, attention, information processing, and fine motor skills. We construct a cognitive skills index, based on these seven scores, to reflect general neurocognitive function.

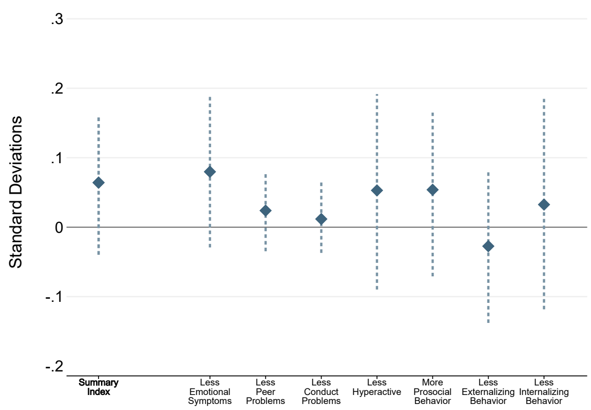
A.5.2 Results on Socioemotional and Cognitive Skills

SI Figure A20 shows effects on socioemotional and cognitive skills. We do not find any statistically significant evidence of effects on socioemotional skills using either parent or teacher-reported measures, but the point estimates are generally positive. We can almost always rule out effects larger than 0.2 standard deviations. Finally, there is no evidence that the intervention significantly affected cognitive skills. Although the point estimates on performance IQ at age 6 are large, the estimates are quite noisy and there are concerns that these pediatrician-measured assessments could be biased. It is also clear that no effects on objective measures of IQ persisted into adolescence. While earlier work has found significant impacts on specific measures of cognition during childhood (Kramer et al., 2008*b*), we do not find robust support for this result when evaluating summary measures rather than individual components and applying multiple hypothesis testing

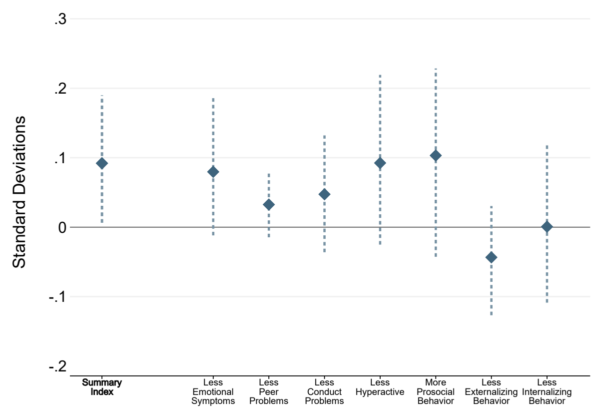
Figure A20

The Effect of Breastfeeding Promotion on Socioemotional and Cognitive Skills

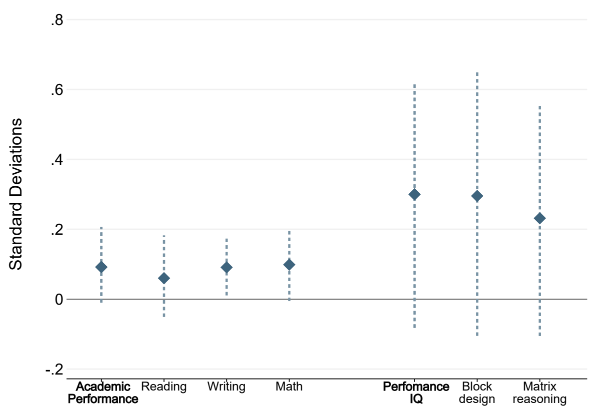
(a) Parent: Socioemotional Skills



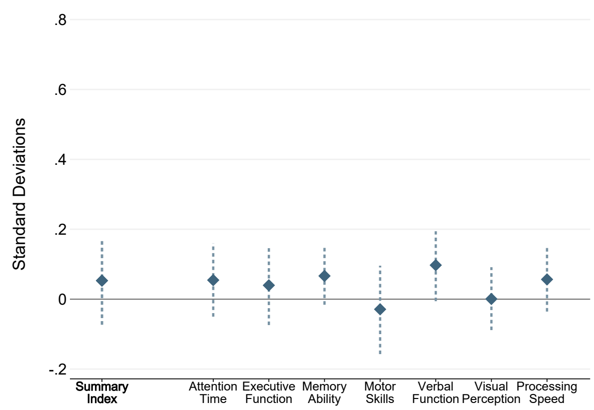
(b) Teacher: Socioemotional Skills



(c) Cognitive Skills at Age 6



(d) Cognitive Skills at Age 16

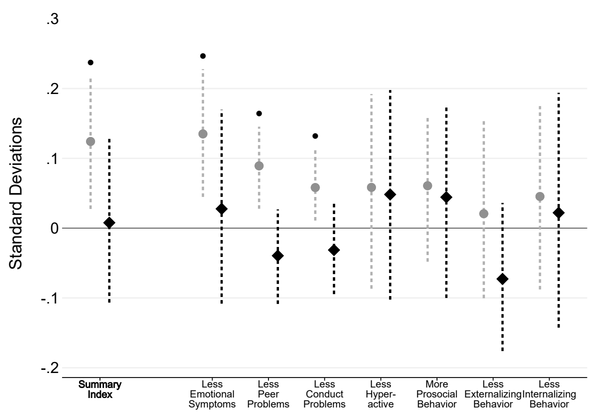


Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

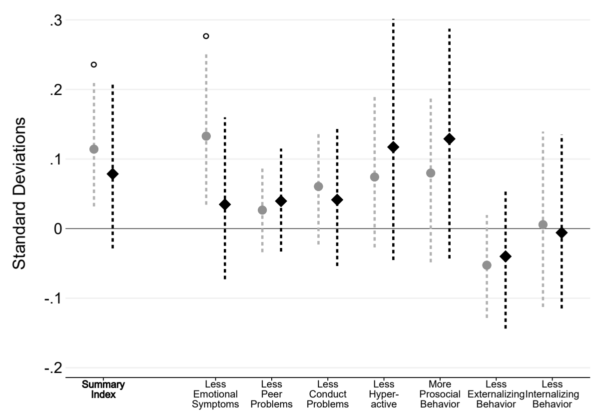
corrections. SI Figures A21 and A22 further show that there is no noteworthy heterogeneity in the effects on scocioemotional or cognitive skills by child sex or socioeconomic status.

Figure A21
The Effect of Breastfeeding Promotion on Socioemotional and Cognitive Skills: Heterogeneity by Gender

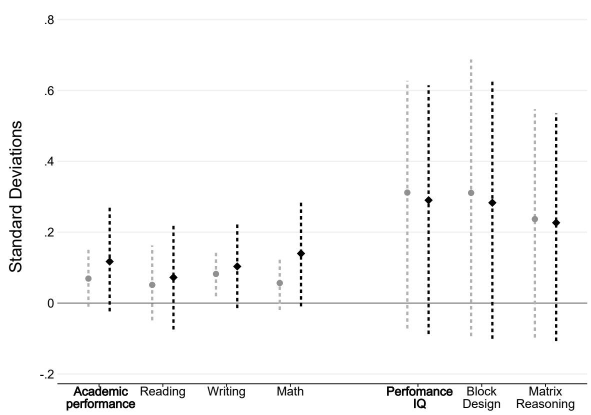
(a) Parent: Socioemotional Skills



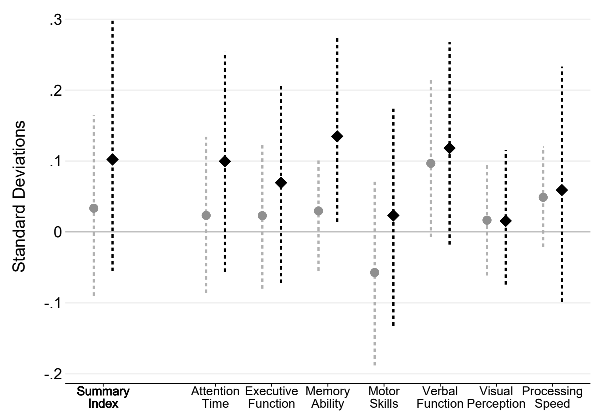
(b) Teacher: Socioemotional Skills



(c) Cognitive Skills at Age 6



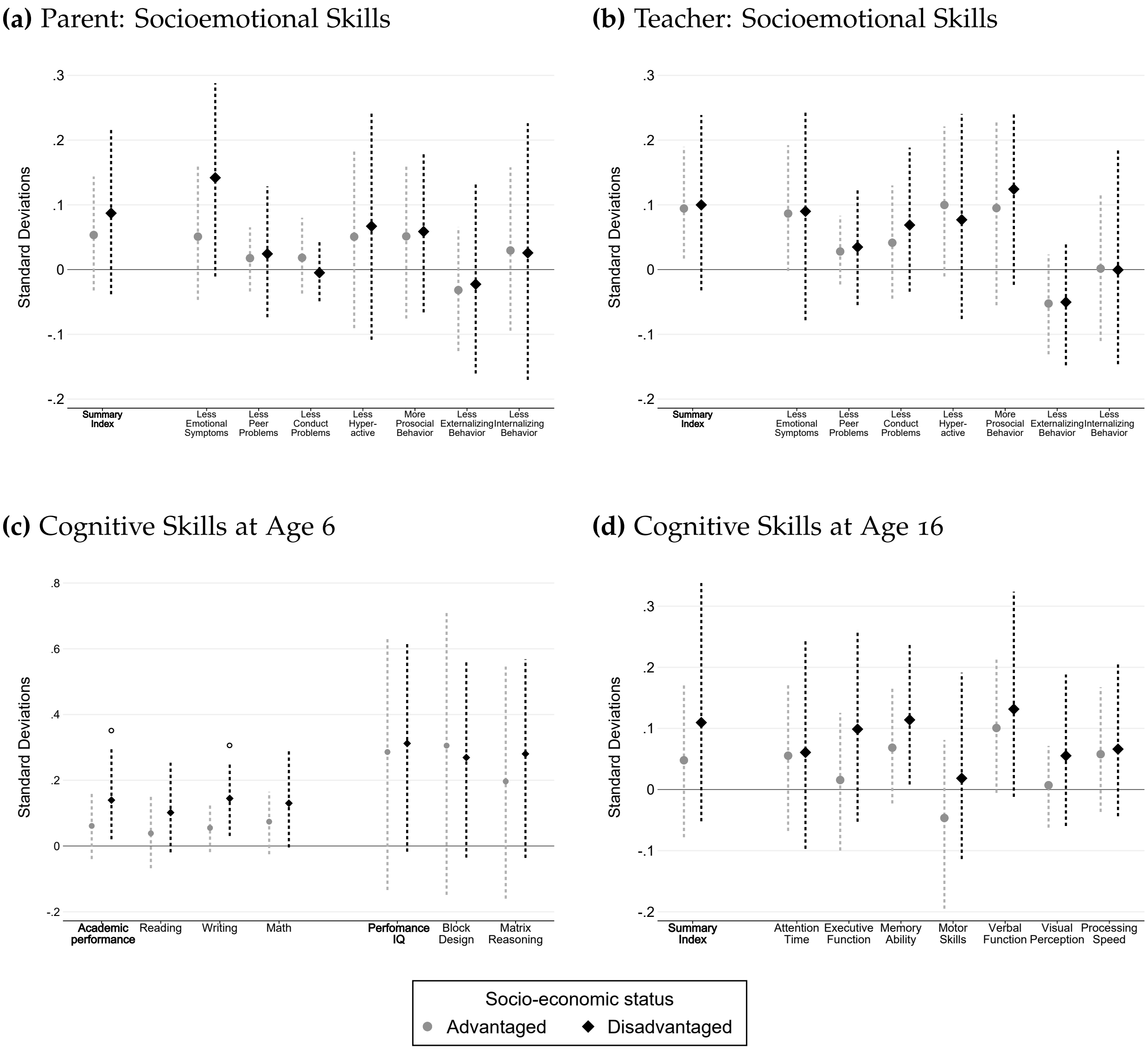
(d) Cognitive Skills at Age 16



Gender
● Female ◆ Male

Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ◦$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

Figure A22
The Effect of Breastfeeding Promotion on Socioemotional and Cognitive Skills: Heterogeneity by Socioeconomic Status



Note: Each set of estimates comes from a separate regression as specified in equation (5); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

A.6 Institutional review board approval

PROBIT I and II were approved by the Belarusian Ministry of Health and received ethical approval from the McGill University Health Centre Research Ethics Board; PROBIT III and IV were approved by the Belarusian Ministry of Health and received ethical approval from the McGill University Health Centre Research Ethics Board, the Human Subjects Committee at Harvard Pilgrim Health Care, and the Avon Longitudinal Study of Parents and Children (ALSPAC) Law and Ethics Committee. A parent or legal guardian provided written informed consent in Russian at enrollment and at the follow-up visits, and all children provided written assent at the 11.5-year and 16-year visit.