Causal Effects of Breastfeeding Promotion on Child Health: Understanding 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

May 2022

Abstract

Using data from the only large-scale randomized controlled trial promoting prolonged exclusive breastfeeding, we show that the intervention significantly and persistently increased weight-for-age. To explain this result, we provide novel evidence of changes in infant feeding patterns. The estimated increase in calories that treated infants consumed explains a major share of the weight gain in early infancy. Our results suggest that understanding the common alternatives to breast milk is key for designing optimal infant feeding policies.

*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, Institute of Labor Economics (IZA), and Center for Economic Behavior and Inequality (CEBI) at the University of Copenhagen. 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. We are grateful to Liaoliang Zhang, Maximilian Mähr, Claude Raisaro, and Yanzun Yang for excellent research assistance. We thank Pietro Biroli, Ana Costa-Ramon, David Dorn, Ernst Fehr, Nir Jaimovich, Henrik Kleven, Michael Kramer, Claudio Schilter, Hannes Schwandt, Xiaoyue Shan, Marcos Vera-Hernandez, Miriam Wüst, David Yanagizawa-Drott and seminar participants at the International Health Economics Association Conference 2019, the Importance of Early-Life Circumstances Workshop at the University of Copenhagen, the University of Zurich, and the University of Duisburg-Essen 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. Martin was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol, Weston NHS Foundation Trust, and the University of Bristol. 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.

JEL classification: I10, J13, J24
Keywords: Breastfeeding, infant feeding, child health, the Promotion of Breastfeeding Intervention Trial (PROBIT)

1 Introduction

The potential positive association between exclusive breastfeeding and beneficial child outcomes has extensive public health and economic consequences across the world.[1] 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 (e.g., Kramer and Kakuma (2012); Victora et al. (2016)), the causal evidence of the effects on child outcomes is scarce (Fitzsimons and Vera-Hernández, 2021). Arguably, the best 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 fewer than those found in observational settings, which are prone to issues of bias and reverse causation (Kramer, 2010).

A key limitation of the current causal literature is the lack of evidence on mechanisms. Any benefits of breastfeeding on child health and development could be driven by improved infant nutrition or caloric intake, the unique elements of breast milk (such as antibodies) that are not found in high-quality infant formula, or increased social stimulation through the physical act of breastfeeding. Understanding which of these mechanisms are most relevant is crucial for designing optimal infant feeding policies—whether it is to promote and support breastfeeding itself, enable breast pumping at work, regulate the composition of infant formula, or to promote programs to more broadly foster early childhood stimulation. In this paper, we provide novel evidence suggesting that improved nutrition is a key driver of the benefits of breastfeeding.

We use data from the PROBIT RCT to estimate the causal effects of promoting breastfeeding duration and exclusivity on child health, and provide novel insight into the role of nutrition as a plausible mechanism. We first show that the intervention significantly and persistently increased weight-for-age, but find little evidence of robust and persis-

[1] The World Health Organization (WHO) recommends that mothers worldwide should exclusively breastfeed infants during the child's first six months of life, and that children should continue to be breastfed in addition to receiving other nutritious foods until at least two years of age (WHO, 2011). Exclusive breastfeeding is defined as feeding no other foods or liquids, including water.

tent effects on several other dimensions of child health. As our main contribution, we next provide novel evidence on how the breastfeeding promotion intervention affected infant feeding patterns and nutrition, and how these changes mediated the effect on infant weight gain. We find that infants exposed to the breastfeeding intervention were breastfed more and consumed less water, juice, and other liquids throughout their first year of life. This resulted in a more calorie-dense and calorie-rich diet compared to infants in the control group.

These effects on infant feeding patterns are important for interpreting the mediating mechanisms of breastfeeding on child health. We employ a mediation analysis that shows that the increased caloric intake explains a large proportion of the effect on weight gain in early infancy. This suggests that improved nutrition during key periods of growth have contemporaneous and lasting effects on weight gain. The results also indicate that, at least in this setting where water and juice are common substitutes for breast milk, the primary benefit of breastfeeding on children's outcomes is improved nutrition. In contrast, we do not find any evidence that reduced incidence of illness can explain the effects on weight gain, ruling out the unique elements of breast milk not found in infant formula as an important mechanism of weight gain. Furthermore, there is no strong evidence of heterogeneity in the effects of the breastfeeding promotion intervention on child health or feeding patterns with respect to gender or socioeconomic status and, consistent with previous research, we do not find supportive evidence of effects of the intervention on socioemotional skills at age 6 (Kramer et al., 2008*a*) or cognitive development at age 16 (Yang et al., 2018).[2] These results suggest that improved nutrition in infancy is the main mechanism driving the effects on weight gain.

Our findings have important policy implications and underscore the need for continued research in this area. The results highlight the importance of understanding the local alternatives to breastfeeding when evaluating the potential benefits of breastfeeding interventions. Because mothers in Belarus did not use breast milk as a perfect substitute for infant formula, but instead replaced breastfeeding with less nutritious alternatives, the results on weight gain may represent an upper bound of the heath benefits of breastfeeding in other settings. 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, as is the case in the United States (Grummer-Strawn, Scanlon and Fein, 2008). Although more work is needed, evidence on the causal effects of breast-

[2]We report these results in Appendices A.5 and A.6.

feeding on early childhood health in high-income countries is extremely limited (Baker and Milligan, 2008; Del Bono and Rabe, 2012; Fitzsimons and Vera-Hernández, 2021).

We contribute to the literature in several important ways. Our key contribution is to bring new insights into the mechanisms of breastfeeding. Previous PROBIT papers have not reported the effect of the intervention on other infant feeding outcomes than breastfeeding exclusivity and duration. However, to interpret the effects of the intervention and for thinking about the generalizability of the results, it is important to understand what the alternatives to breast milk were and how the intervention affected infant feeding patterns. We further contribute to the existing literature analyzing the effect of the PROBIT intervention on child development by advancing the empirical framework. Overall, our findings on child outcomes are consistent with previous findings (Kramer et al., 2001, 2002, 2008*a*; Martin et al., 2017; Yang et al., 2018).

Our results also contribute to a literature on the effects of nutritional supplementation among infants and toddlers on contemporaneous weight gain and later cognitive development in developing countries (Schroeder et al., 1995; Walker et al., 1991, 2005; Maluccio et al., 2009). In contrast to these studies, it is worth noting that our findings are from a setting in which infants were not generally underweight or particularly unhealthy even in absence of the breastfeeding intervention. By providing the first causal evidence linking the positive effects of breastfeeding to improved nutrition among healthy infants, we highlight the importance of understanding mechanisms by which breastfeeding can improve child outcomes when designing cost-effective policies that support optimal infant feeding practices.

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.[3] Eligible infants were born weighing at least 2500 grams and at a gestational age of 37 weeks or greater. 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. (2000).

Across all hospitals, the postpartum stay after a routine vaginal delivery was 6–7 days (Kramer et al., 2001). Treatment hospitals received an intervention promoting breastfeeding, while the control hospitals continued the standard practices in effect at the time of randomization. The conventional practices at control sites included routine separation of mother and child, delayed onset of breastfeeding, scheduled (versus on demand) feedings, routine use of water, formula, and other liquids in newborn diets, and recommendation of early introduction of solid foods (Patel et al., 2013). In contrast, mothers at intervention hospitals received help initiating breastfeeding within 30 minutes of a normal birth, infants were supposed to remain with their mothers 24 hours a day during the postpartum hospital stay, and newborns were fed only breast milk on demand unless medically indicated. All pregnant women at intervention hospitals were also informed about the benefits and management of breastfeeding. The intervention required hospitals assigned to the active group to have a written breastfeeding policy that all staff had the skills to implement. The head obstetrician and pediatrician from each treatment hospital received an 18 hour lactation management training course, and then trial participants organized and implemented further training programs for midwives, nurses, physicians, and pediatricians working in the postpartum ward and pediatricians working in the associated polyclinic. The full implementation of the intervention required at least 12 months. Further details about the experimental context are provided in Appendix A.2.

As previously documented by Kramer et al. (2001), the breastfeeding promotion intervention substantially changed mothers' breastfeeding behavior. Mothers who gave birth at treated hospitals were considerably more likely to breastfeed exclusively for up to six months compared to mothers at control hospitals.[4] For example, treated moth-

[3] 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 (Appendix Figure A1), and the intervention had no effect on infant mortality (Kramer et al., 2001). Additionally, trial staff estimated that only 1–2 percent of eligible women declined participation in the study.

[4] Appendix Figure A2 shows the effects on exclusive and any breastfeeding.

ers 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 who had exclusive breastfeeding rates of only 27 percent at one month and seven percent at three months. The intervention also increased breastfeeding duration, with treated mothers significantly more likely to breastfeed for at least 12 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, locally made 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 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). Appendix Table A1 shows descriptive statistics at baseline. As expected, given the RCT design, mothers who gave birth at treated and control hospitals are very similar. Finally, θ_p are hospital pair fixed effects, which we include to account for the stratified design

(Duflo, Glennerster and Kremer, 2007).

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. This is in contrast with the clustering methods used in previous published PROBIT papers and may explain differences in the precision of the respective effect estimates.[5] 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.

4 Data and Outcome Variables

Participants were followed six times throughout their first year of life at regular health checkups (at 1, 2, 3, 6, 9, and 12 months), and again at ages 6.5, 11.5, and 16.[6] In each wave, a pediatrician conducted physical growth measurements. For the analysis of child health, we focus on weight-for-age, which we construct as a standardized age-specific z-score (Vidmar et al., 2004). Children in the control group had healthy weights during infancy and childhood (Appendix Table A3). Using a standardized weight-for-age measure yields an easy interpretation of the results on weight gain in terms of standard deviations. Particularly in early life, it also is highly correlated with recent changes in nutrition and health status.[7] This makes it a well-suited measure for studying the relationship between breastfeeding, nutrition, and physical growth.

Our infant feeding data are based on maternal reports of feedings during the 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). Less than one percent of breast milk feedings are expressed milk and less than 0.1 percent are donor milk, with the rest being breastfeedings at the breast. Thus, we refer to breast milk and breastfeeding interchangeably. Based on

[5] Previous published PROBIT papers have used mixed effects linear models and mixed logistic regression models.

[6] Appendix A.3 provides more details regarding the data and construction of the variables.

[7] The results are qualitatively similar if we use weight or standardize the weight within the sample instead of weight-for-age (Appendix Figure A3).

the reported frequency and volume of feedings from each category, we also construct measures of caloric intake. We assume breast milk, formula, and cow's milk contain 65 kcal per 100 ml, while juices and other liquids contain an average of 30 kcal per 100 ml.[8] For breast milk and solid foods, a measure of volume per feeding is not available. We follow Lupton et al. (2002) to estimate breast milk caloric intake and volume.[9] As for every type of self-reported measure, we cannot rule out potential reporting bias in the feeding data. However, mothers did not have any particular incentive to misreport.[10]

5 Results

5.1 Child Health

Figure 1 shows the effects of the breastfeeding promotion intervention on weight-for-age throughout childhood. Overall, we find a statistically significant and persistent effect of the intervention on weight-for-age. Infants exposed to the 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. These findings are consistent with those reported by Kramer et al. (2002) and Martin et al. (2017) from the PROBIT trial . Children were not statistically more likely to be either overweight or underweight at any of the considered ages (Appendix Figure A7), suggesting that the effect on weight-for-age was driven by weights within the healthy range.

Considering other dimensions of child health, we overall confirm previous findings from the PROBIT study (Appendix A.5). In line with Kramer et al. (2001), the breastfeeding intervention decreased the likelihood of infant illness, although our effects are imprecisely estimated (Appendix Figure A8). We also do not find meaningful effects

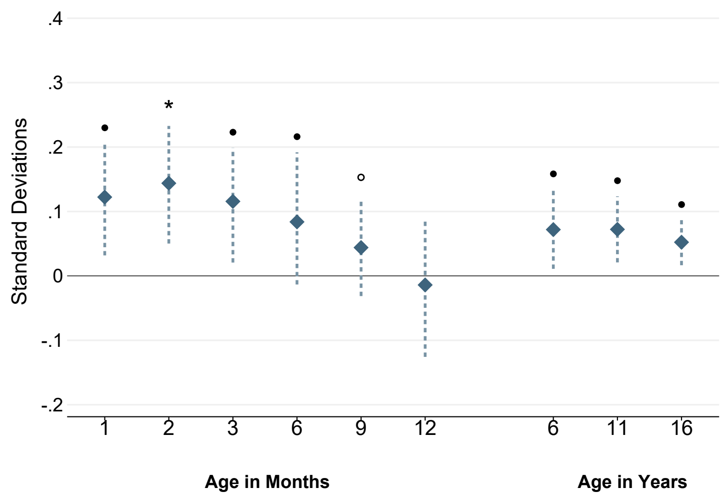
[8] Calories from juices and other liquids assumes two-thirds comes from apple juice (45 kcal per 100 ml) and one-third from black tea (0 kcal per 100 ml).

[9] For breastfed children at one to three months, we estimate breast milk intake based on the recommended daily calorie intake formula Calorie Intake $= 89 \times$ Weight in Kg $+ 75$, subtract their calorie intake from all other liquids, and calculate the volume. For breastfed children at 6 to 12 months, we assume that each breast milk feeding contains 175 ml (Whitehead, Paul and Cole, 1982). At these ages, we cannot use the daily calorie intake formula, as solid food makes up an increasing share of calorie intake and we do not have nutritional information about solid food.

[10] We also note that we added the infant feeding analysis as a secondary analysis to our original analysis plan to investigate the channels of our results on child growth (for more details, see Appendix A.4).

on a physical health index or persistent effects on height-for-age (Appendix Figure A7); these findings are consistent with those by Kramer et al. (2002, 2007); Martin et al. (2013, 2017). Thus, within the domain of child health, the only robust and persistent effect of the breastfeeding intervention is on weight gain. [11]

Figure 1
The Effect of Breastfeeding Promotion on Weight-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. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

5.2 Infant Feeding Patterns

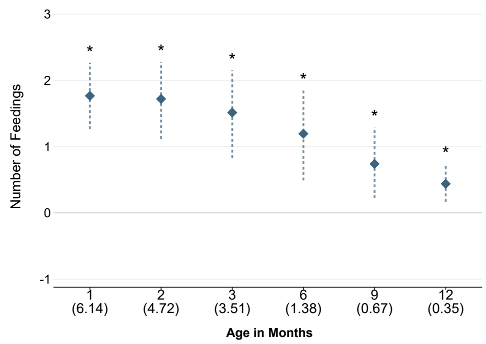
Studying child health and development, we only find robust statistically significant effects of the breastfeeding promotion intervention on weight-for-age. This effect persisted until at least age 16, but appears to be driven by early weight gain during months also associated with higher rates of exclusive 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 the number of feedings per day over the first 12 months of life by nutrition type (breast

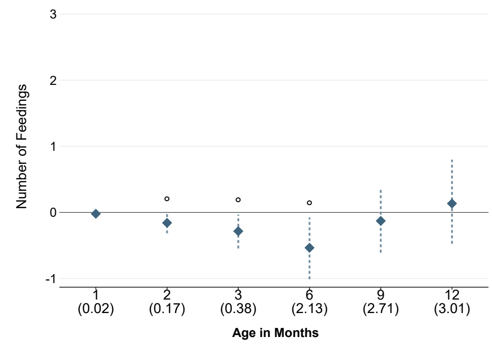
[11]We do not consider measures of eczema or asthma in our analysis. However, we note that Flohr et al. (2018) find a substantial reduction in flexural eczema on skin examination at age 16 by 54 percent, based on a rate in the control group of 0.7 percent.

Figure 2
The Effect of Breastfeeding Promotion on Frequency of Infant Feedings per Day

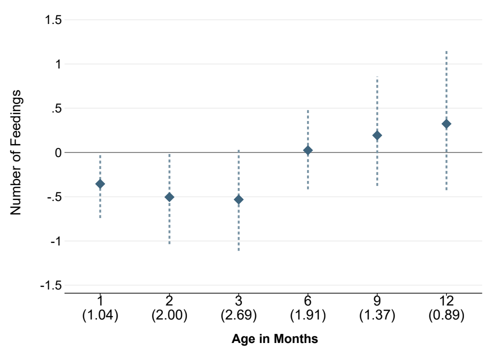
(a) Breastfeeding



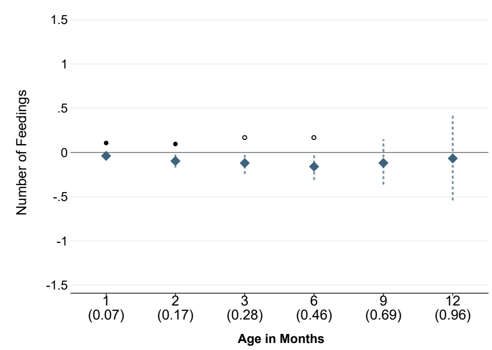
(b) Solid Food



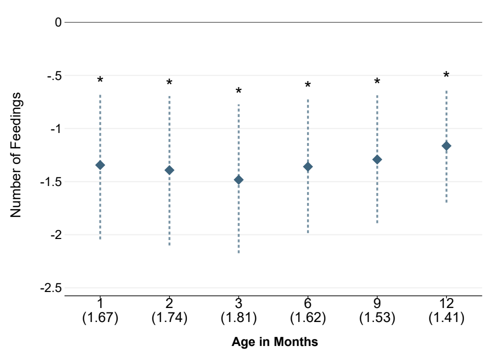
(c) Infant Formula



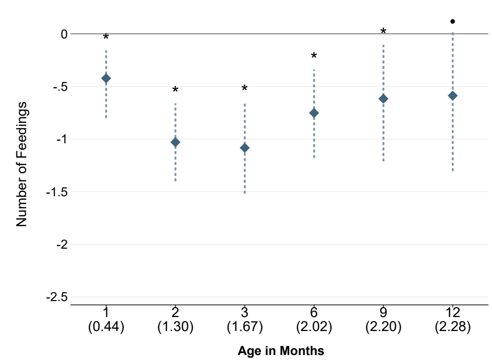
(d) Cow's Milk



(e) Water



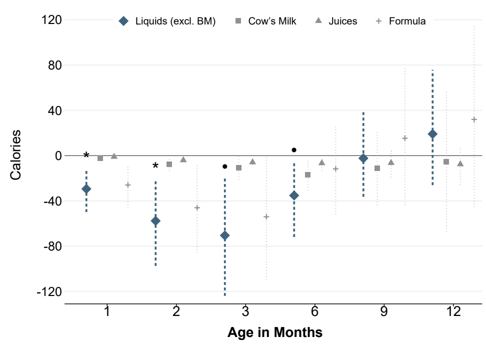
(f) Juice or Other Liquids



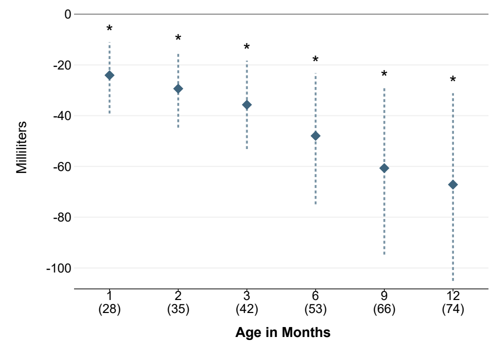
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on 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

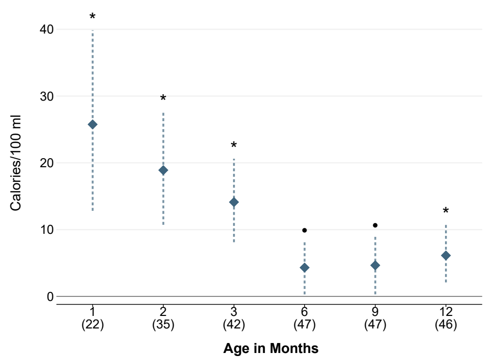
(a) Calories from Liquids (excl. BM)



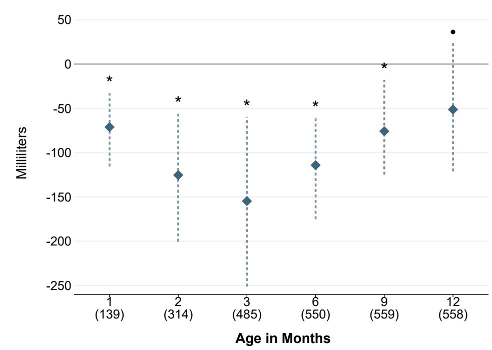
(b) Volume of Water



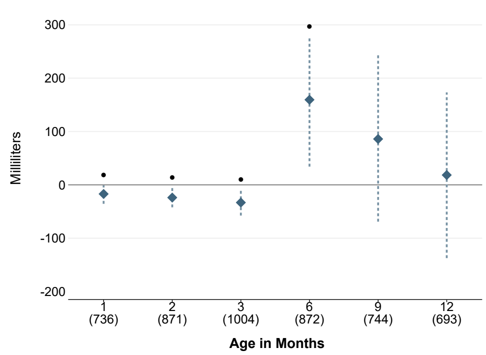
(c) Calories/100 ml Liquids (excl. BM)



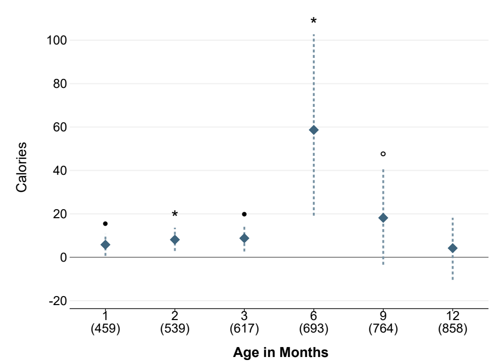
(d) Total Volume Liquids (excl. BM)



(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 Appendix Figure A4. 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 Appendix A.3 for the exact construction of the outcome variables.

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.[12] 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 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. 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. When including breast milk in the volume of all liquids, the negative effect is very small for ages one to three months and

[12] We do not have information on the caloric intake of solid foods. However, we note that it is quite limited at ages one to three months, as infants this young are usually not able to digest or swallow non-liquids.

becomes positive at age six months (Graph 3e).[13] Graph 3f shows the estimated total caloric intake from all liquids including breast milk. Treated infants received around eight calories more per day from liquids than control infants at ages one to three months and even more at later ages, representing almost a two percent increase in calories per day compared to the recommended nutritional guidelines for infants. 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.

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 and in line with Skugarevsky et al. (2014), we also find a robust decline in an index measuring problematic eating attitudes among treated children at age 11 (Appendix Figure A5).

5.3 The Role of Nutrition in Explaining Infant Weight Gain

Can the increase in calorie consumption due to the breastfeeding promotion intervention explain the effects on weight gain? Following Gelbach (2016), we decompose the treatment effect on weight-for-age into experimentally induced changes in caloric intake and changes in other (unmeasured) factors. We limit this mediation analysis to ages one to three 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.

[13] We note that the spike at age six months most likely is due to the delayed introduction of solid foods compared to control children; we therefore do not claim that this represents a net effect on total calorie intake at this age.

We calculate the mediated effect, ME, of the experimentally induced increase in the total caloric intake on weight-for-age as:

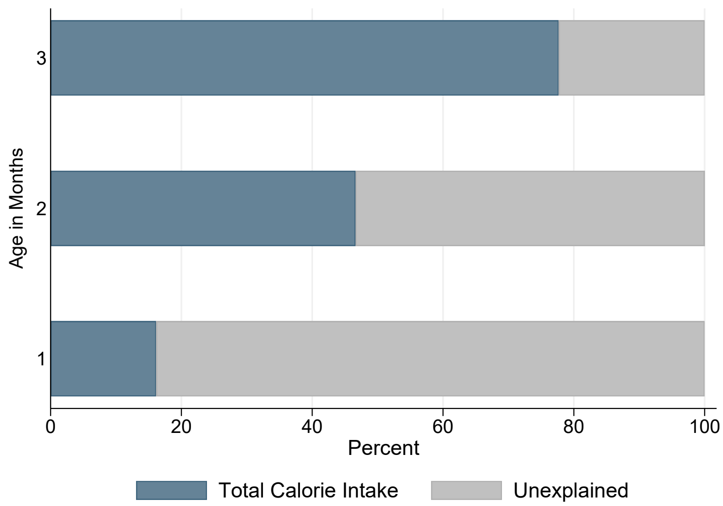
$$ME = \underbrace{\frac{\partial Y_{iph}}{\partial Calories_i}}_{\text{A}} \underbrace{\frac{\partial Calories_i}{\partial Treatment_h}}_{\text{B}}, \tag{2}$$

following the notation from equation (1). The term B of equation (2) represents the causal effect of the intervention on the calorie intake. The term A is the parameter ϕ from the following equation:

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

Under the assumption that equation (3) does not omit any variables that both affect the mediator (*Calories*) and the outcome, we can interpret the estimates from the mediation analysis as causal. Although we cannot think of any obvious omitted variables, we can naturally not rule out this possibility. We then calculate the mediated effect as the share of the total treatment effect of the intervention on weight-for-age. In the mediation analysis, we include as mediator the total cumulative calorie intake from all liquids including breast milk through the age at the weight measurement.

Figure 4
Mediated Effect of Total Caloric Intake on Weight-for-age



Note: This figure shows the results of the mediation analysis, decomposing the overall treatment effect on weight-for-age at age one, two, and three months into the experimentally induced changes in the total cumulated caloric intake from all liquids including breast milk and changes in other (unmeasured) factors. At one, two, and three months, respectively, the p-values of the estimate of the mediator are 0.068, < 0.001, and < 0.001. For the estimation, we use the Gelbach (2016) b1x2 Stata package.

Figure 4 presents the results from the mediation analysis. We estimate that a remarkable 78 percent of the effect of the breastfeeding intervention on weight-for-age is mediated by the experimentally induced increase in the total calorie intake at age three months. The effect of the intervention on caloric intake becomes increasingly important over the first few months of life, with 46 percent of the effect explained by increased calorie consumption at age two months and a third of this at age one month.

This analysis suggests that improved nutrition is the predominant mechanism for the effects of the breastfeeding intervention on weight gain. However, other channels might potentially also be relevant. In particular, the increase in breastfeeding also implies an increased exposure of infants to maternal antibodies and thereby a potential improvement in infants' immune system response (Moraes-Pinto, Suano-Souza and Aranda, 2021). Although we do not find statistically significant effects on the infant illness index, we cannot rule out medium-sized reductions. Appendix Table A2 includes the illness index as an additional mediator. Only a negligible part of the treatment effect of the intervention on weight-for-age is mediated through this index. The results are similar when including each of the components separately. In particular, gastroenteritis in early infancy does not explain any of the effect on weight gain which is worth highlighting as gastroenteritis typically involves diarrhea or vomiting and thereby likely reduces caloric intake.

In addition to the increased exposure to biological components in breast milk, the intervention may also have reduced the exposure to non-sterile food or feeding equipment, which might have improved infant health beyond what our illness index captures. In Appendix Table A2, we therefore consider as mediators the number of the different types of feedings. None of the non-breast milk feeding types have any mediating power; but the number of breastfeedings does mediate a substantial fraction of the effect of the breastfeeding intervention on weight gain. Thus, there is no support of this potential channel, which is also in line with the generally good hygienic setting.

When considering the important factors generally affecting body weight among older children and adults, physical exercise is key in addition to nutrition. However, it seems unlikely that physical activity would be an important mechanism in our setting, as young infants typically sleep 13–16 hours per day (Iglowstein et al., 2003) and spend most of the remaining time eating. Finally, we also find no evidence of significant socioemotional or cognitive effects (Appendix Figure A9) that could be driven by the physical act of breastfeeding.

In conclusion, 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 children's physical growth in this setting.

6 Conclusion

WHO's recommendation of six months of 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 of breastfeeding on child health and development, or the specific mechanisms that might drive these effects. 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. Although the infant feeding data is self-reported and potentially measured with error (as is typically the case for this type of data, with the Health and Demographic Surveys being a prominent example), these findings suggest that the intervention resulted in a more calorie-dense and calorie-rich diet.

In explaining the mechanism driving the effects of breastfeeding on child health, our analysis suggests that nutrition is the most important channel. While the increased calorie consumption of breastfed infants can explain most of the effect on weight gain in the first three months of life, reductions in illness explains less than 0.2 percent. This highlights the importance of changes in nutrition as the primary channel through which breastfeeding increased child weight gain in Belarus. This finding is relevant for similar settings characterized by low-nutrition alternatives to breast milk and 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 for child health and resulting policy implications may be different in settings where higher-nutrition alternatives to breast milk are more common.

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.

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.

Cockburn, F. 2003. "Role of infant dietary long-chain polyunsaturated fatty acids, liposoluble vitamins, cholesterol and lecithin on psychomotor development." *Acta Paediatrica*, 92: 19–33.

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."

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using randomization in development economics research: A toolkit." *Handbook of development economics*, 4: 3895–3962.

Fitzsimons, Emla, and Marcos Vera-Hernández. 2021. "Breastfeeding and Child Development." *American Economic Journal: Applied Economics (Forthcoming)*.

Flohr, Carsten, A John Henderson, Michael S Kramer, Rita Patel, Jennifer Thompson, Sheryl L Rifas-Shiman, Seungmi Yang, Konstantin Vilchuck, Natalia Bogdanovich, Mikhail Hameza, et al. 2018. "Effect of an intervention to promote breastfeeding on asthma, lung function, and atopic eczema at age 16 years: Follow-up of the probit randomized trial." *JAMA pediatrics*, 172(1): e174064–e174064.

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

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.

Hoffman, Dennis R, Julia A Boettcher, and Deborah A Diersen-Schade. 2009. "Toward optimizing vision and cognition in term infants by dietary docosahexaenoic and arachidonic acid supplementation: a review of randomized controlled trials." *Prostaglandins, Leukotrienes and essential fatty acids*, 81(2-3): 151–158.

Iglowstein, Ivo, Oskar G Jenni, Luciano Molinari, and Remo H Largo. 2003. "Sleep duration from infancy to adolescence: reference values and generational trends." *Pediatrics*, 111(2): 302–307.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Kramer, Michael S. 2010. ""Breast is best": The evidence." *Early human development*, 86(11): 729–732.

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, George Shishko, Vyacheslav Zubovich, Dimitri Mknuik, Elena Gluchanina, Viktor Dombrovskiy, Anatoly Ustinovitch, Tamara Kot, Natalia Bogdanovich, Lydia Ovchinikova, and Elisabet Helsing. 2000. "PROMOTION OF BREASTFEEDING INTERVENTION TRIAL (PROBIT): A CLUSTER-RANDOMIZED TRIAL IN THE REPUBLIC OF BELARUS." *Short and Long Term Effects of Breast Feeding on Child Health*, 478: 327.

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, Lidia Matush, Irina Vanilovich, Robert W Platt, Natalia Bogdanovich, Zinaida Sevkovskaya, Irina Dzikovich, Gyorgy Shishko, Jean-Paul Collet, Richard M Martin, et al. 2007. "Effects of prolonged and exclusive breastfeeding on child height, weight, adiposity, and blood pressure at age 6.5 y: evidence from a large randomized trial." *The American journal of clinical nutrition*, 86(6): 1717–1721.

Kramer, Michael S, Richard M Martin, Jonathan A C Sterne, Stanley Shapiro, Mourad Dahhou, and Robert W Platt. 2009. "The double jeopardy of clustered measurement and cluster randomisation." *BMJ*, 339.

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. "Breastfeeding and infant growth: biology or bias?" *Pediatrics*, 110(2): 343–347.

Krol, Kathleen M, and Tobias Grossmann. 2018. "Psychological effects of breastfeeding on children and mothers." *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 61(8): 977–985.

Lien, EL, C Richard, and DR Hoffman. 2018. "DHA and ARA addition to infant formula: Current status and future research directions." *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 128: 26–40.

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.

Maluccio, John A, John Hoddinott, Jere R Behrman, Reynaldo Martorell, Agnes R Quisumbing, and Aryeh D Stein. 2009. "The impact of improving nutrition during early childhood on education among Guatemalan adults." *The Economic Journal*, 119(537): 734–763.

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.

Moraes-Pinto, Maria Isabel de, Fabíola Suano-Souza, and Carolina S Aranda. 2021. "Immune system: Development and acquisition of immunological competence." *Jornal de Pediatria*, 97: 59–66.

Oken, Emily, Rita Patel, Lauren B Guthrie, Konstantin Vilchuck, Natalia Bogdanovich, Natalia Sergeichick, Tom M Palmer, Michael S Kramer, and Richard M Martin. 2013. "Effects of an intervention to promote breastfeeding on maternal adiposity and blood pressure at 11.5 y postpartum: results from the Promotion of Breastfeeding Intervention Trial, a cluster-randomized controlled trial." *The American journal of clinical nutrition*, 98(4): 1048–1056.

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."

Owen, Christopher G, Emily Oken, Alicja R Rudnicka, Rita Patel, Jennifer Thompson, Sheryl L Rifas-Shiman, Konstatin Vilchuck, Natalia Bogdanovich, Mikhail Hameza, Michael S Kramer, et al. 2018. "The Effect of Longer-Term and Exclusive Breastfeeding Promotion on Visual Outcome in Adolescence." *Investigative ophthalmology & visual science*, 59(7): 2670–2678.

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.

Walker, Susan P, Susan M Chang, Christine A Powell, and Sally M Grantham-McGregor. 2005. "Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study." *The lancet*, 366(9499): 1804–1807.

Whitehead, RG, Alison A Paul, and TJ Cole. 1982. "How much breast milk do babies need?" *Acta Pædiatrica*, 71: 43–50.

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.

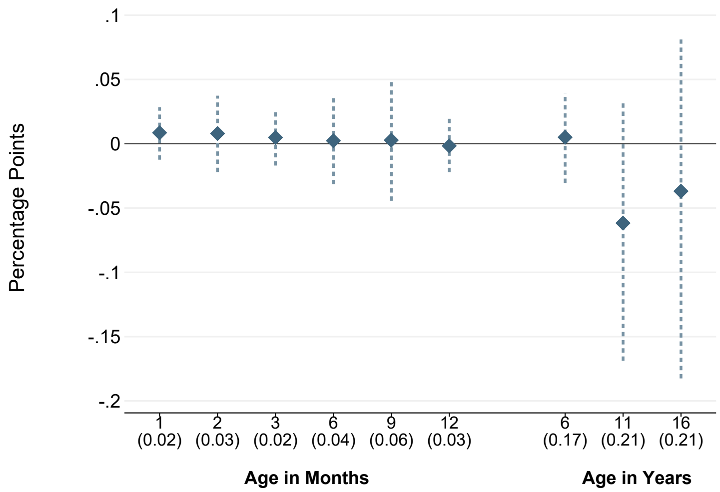
WHO, World Health Organization. 2020. "The WHO Child Growth Standards." URL: *https://www.who.int/childgrowth/standards/en/*, Accessed: 2020-07-16.

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 Online Appendix

A.1 Supplementary Figures and Tables

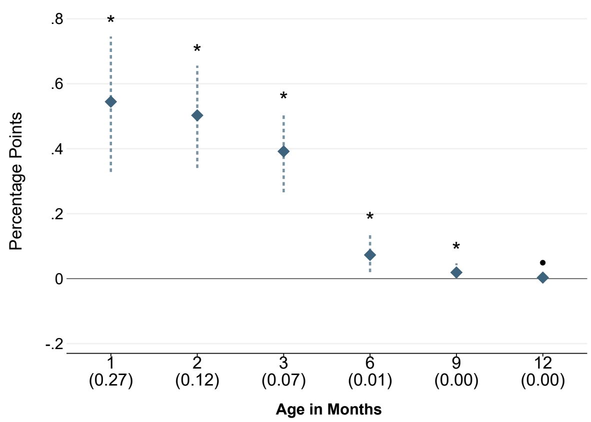
Figure A1
Attrition by Treatment Status



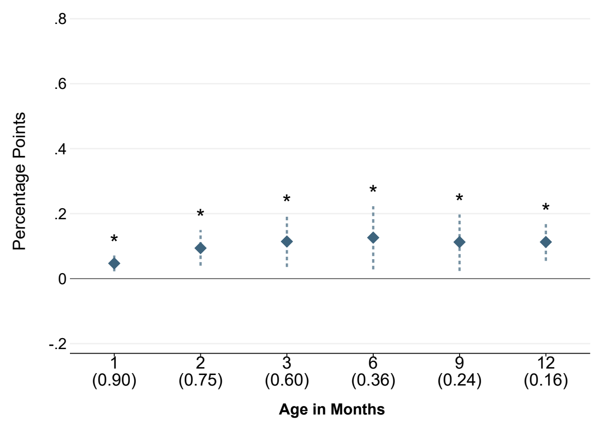
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on 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 share of missing observations in the control arm at each time point. 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



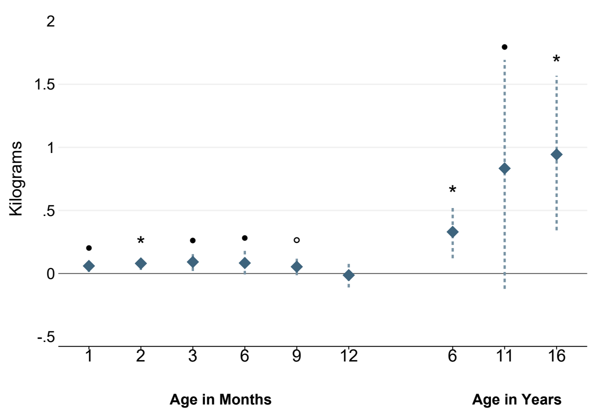
(b) Any Breastfeeding



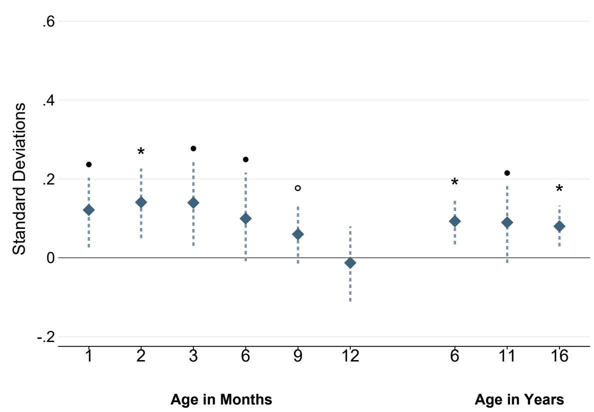
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on 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 Weight

(a) Weight in Kilograms



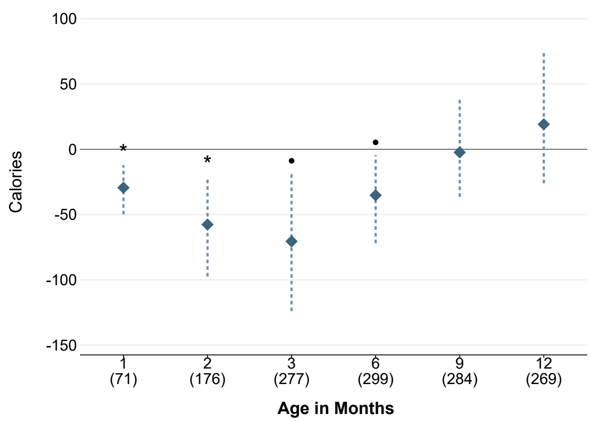
(b) Standardized Weight



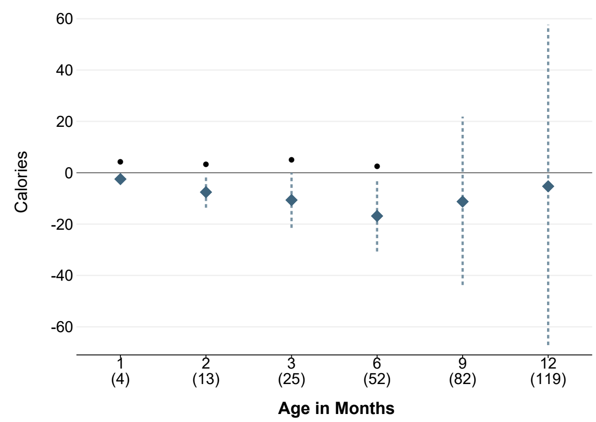
Note: The outcome in panel (b) is standardized with a mean of zero and standard deviation of one for the control group. Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard 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 A4
The Effect of Breastfeeding on Estimated Infant Liquid Calorie Intake

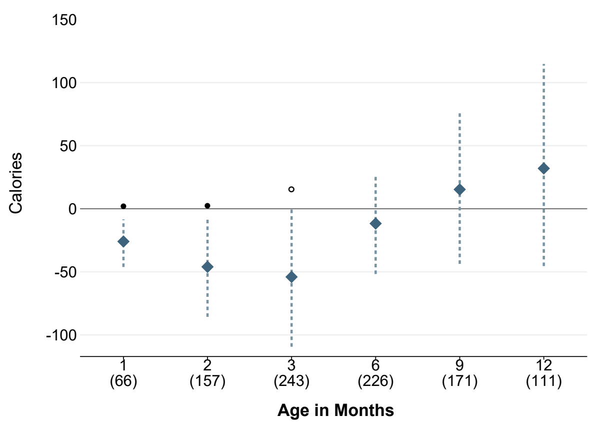
(a) Calories from Liquids (excl. BM)



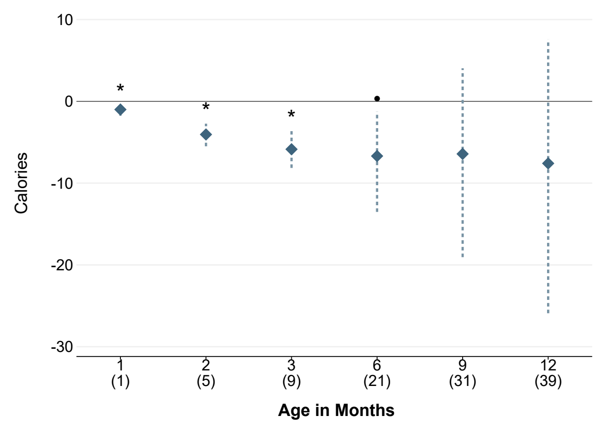
(b) Calories from Cow's Milk



(c) Calories from Formula

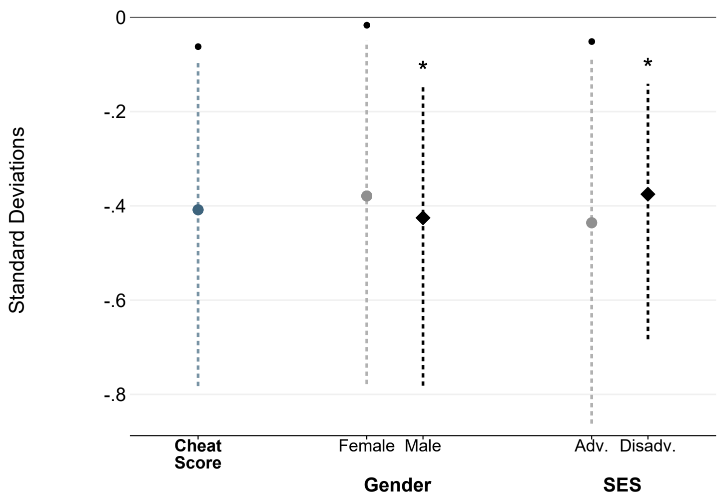


(d) Calories from Juices



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: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$. The numbers reported in parenthesis on the horizontal axis indicate the control mean of the respective outcome variable.

Figure A5
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 (4); 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 Tests

	Control Mean (1)	Standard Deviation (2)	Difference (3)	P-value (4)
Pregnancy and infant characteristics				
Male	0.52	0.50	-0.00	0.85
Birth weight (g)	3437.91	420.98	3.21	0.78
Birth length (cm)	52.02	2.18	-0.21	0.49
Head circumference at birth (cm)	34.79	1.64	0.39	0.27
Delivery complications	0.08	0.28	-0.00	0.87
Caesarean section	0.10	0.31	0.02	0.17
Gestational age at birth (weeks)	39.33	0.98	0.16	0.20
Firstborn	0.56	0.50	0.04	0.11
Smoking during pregnancy	0.02	0.13	0.01	0.48
Alcohol during pregnancy	0.02	0.13	0.03	0.17
Household characteristics				
Married at birth	0.91	0.29	-0.04	0.00
Absent father	0.04	0.19	-0.01	0.52
Mother's Age (years)	24.44	4.91	-0.08	0.65
Father's Age (years)	27.34	5.15	-0.04	0.79
Number of other children in HH	0.58	0.86	-0.05	0.28
Previous exclusive breastfeeding $\geq$ 3 months	0.25	0.43	-0.00	0.96
Socio-economic disadvantage	0.36	0.48	-0.10	0.05
Mother's education				
University degree	0.13	0.34	0.01	0.59
Adv. secondary or partial university	0.53	0.50	-0.06	0.12
Secondary degree	0.30	0.46	0.04	0.36
Incomplete secondary	0.03	0.17	0.01	0.05
Father's education				
University degree	0.13	0.34	0.01	0.52
Adv. secondary or partial university	0.50	0.50	-0.10	0.08
Secondary degree	0.32	0.47	0.09	0.14
Incomplete secondary	0.06	0.23	0.00	0.97
Mother's occupation				
Agriculture or industry	0.32	0.47	0.03	0.65
Services	0.45	0.50	-0.05	0.33
Housewife	0.13	0.34	-0.00	0.74
Unemployed	0.07	0.26	0.01	0.36
Student	0.03	0.17	0.01	0.06
Father's occupation				
Agriculture or industry	0.50	0.50	0.05	0.46
Services	0.31	0.46	-0.06	0.37
Unemployed	0.14	0.35	-0.01	0.44
Student	0.01	0.10	0.01	0.00
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 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 3.

Table A2

Mediation Analysis: Additional Specifications

	1 Month				2 Month				3 Month			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Calorie Intake	0.020	0.020	0.020		0.071	0.071	0.071		0.094	0.095	0.095	
	[0.068]	[0.069]	[0.069]		[0.000]	[0.000]	[0.000]		[0.000]	[0.000]	[0.000]	
Infant Health Index		0.002				0.002				0.005		
		[0.059]				[0.110]				[0.031]		
Any Rash			0.000				0.000				0.001	
			[0.386]				[0.308]				[0.179]	
Any Gastroentiritis			0.000				-0.001				0.000	
			[0.770]				[0.290]				[0.670]	
Any Respiratory			0.000				-0.000				0.000	
			[0.822]				[0.662]				[0.366]	
Any Other Illness			0.000				-0.000				-0.000	
			[0.773]				[0.896]				[0.738]	
Any Hospitalization			-0.000				-0.000				-0.000	
			[0.702]				[0.778]				[0.793]	
Freq. of Breast Milk				0.031				0.050				0.038
				[0.004]				[0.001]				[0.039]
Freq. of Formula				0.001				0.001				0.002
and Cow's Milk				[0.104]				[0.191]				[0.243]
Freq. of Juices and				-0.006				0.003				0.006
Other Liquids				[0.451]				[0.876]				[0.840]
Freq. of Water				0.022				0.024				0.009
				[0.030]				[0.166]				[0.542]
Est. eff. of interv. on WAZ at X months		0.122				0.152				0.121		
Mediated effect of	0.020	0.022	0.020	0.048	0.071	0.072	0.069	0.078	0.094	0.099	0.096	0.055
intervention	[0.068]	[0.031]	[0.053]	[0.010]	[0.000]	[0.000]	[0.000]	[0.007]	[0.000]	[0.000]	[0.000]	[0.085]
Mediated effect of interv. in percent	16.1	18.3	16.4	39.3	46.6	47.8	45.8	51.7	77.7	81.6	78.7	44.9
Est. eff. of interv.	0.103	0.100	0.102	0.074	0.081	0.079	0.082	0.073	0.027	0.022	0.026	0.067
on WAZ at X months with mediators	[0.000]	[0.000]	[0.000]	[0.007]	[0.000]	[0.000]	[0.000]	[0.013]	[0.236]	[0.333]	[0.263]	[0.043]

Note: This table reports the mediated effects based on a Gelbach (2016) decomposition of the intervention, using her b1x2 Stata package. The sample size is 10,152 for all specifications. We include the controls as specified in equation (1). The mediated effect of the intervention in percent is calculated using the ratio of the mediated effect of the intervention to the estimated effect of the intervention. The p-values generated by b1x2 packaged are reported below the estimates.

A.2 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).

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.3 Outcome Measures

This section provides additional details on the data and construction of the outcome measures in the main text as well as child outcomes shown in Appendix A.5. 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 in the appendix, 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). Following Anderson (2008), 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. Following Kling, Liebman and Katz (2007), if an individual has a valid response to at least one component of a particular index, we impute any missing values for other component measures of that index with the mean of individuals from the same hospital, maternal education (low/high), and birth order (firstborn/non-firstborn). In addition to the indices, we also show the results for the individual components; these we do not impute. 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.

Physical Growth 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 a physical growth 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. Thus, the physical growth index is constructed so that larger values reflect more beneficial outcomes. 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). Appendix Table A3 provides descriptive statistics on the physical growth measures for the control group.

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. 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

Table A3
Descriptive Statistics for the Control Group: Physical Growth Measures

	Weight-for-Age (1)	**Height-for-Age** (2)	**BMI-for-Age** (3)	**Underweight** (4)	**Overweight** (5)
1 Month	−0.10	0.10	−0.63	0.20	0.11
	(0.80)	(0.84)	(1.11)	(0.40)	(0.32)
2 Months	0.07	−0.00	−0.26	0.15	0.16
	(0.81)	(0.89)	(1.10)	(0.35)	(0.37)
3 Months	0.25	0.14	−0.06	0.12	0.17
	(0.83)	(0.94)	(1.08)	(0.33)	(0.38)
6 Months	0.45	0.23	0.54	0.03	0.34
	(0.87)	(1.00)	(0.98)	(0.17)	(0.47)
9 Months	0.47	0.28	0.81	0.01	0.50
	(0.84)	(0.95)	(0.90)	(0.09)	(0.50)
12 Months	0.51	0.34	0.93	0.00	0.59
	(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 measures 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).

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)* indicates the total volume of liquids the child received including breast milk. For breastfed children at one to three months, we estimate breast milk intake based on the recommended daily calorie intake formula Calorie Intake $= 89 \times$ 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 to 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.

For the mediation analysis in Section 5.3, we construct measures of the total cumulated calorie intake by respectively age one, two, and three months. To be precise, for the calculation of the total cumulated calorie intake at one month, we assume a calorie intake during the first month of life equal to the estimated calorie intake at the first month health checkup multiplied by the age in days at this checkup. Similarly, we assume that the calorie intake during the second month is equal to the estimated calorie intake at the two months checkup multiplied by the age difference in days between the first and the second checkups. The total cumulated calorie intake at two months is thus the sum of the calorie intakes during the first and the second months. At age three months, we perform a similar calculation.

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. The rates of illnesses during the first year of life were similar or slightly lower in our setting compared to reports from Canada, the United Kingdom, and the United States (Baker and Milligan, 2008; Fitzsimons and Vera-Hernández, 2021; Haider et al., 2014) (Appendix Tables A4).

Table A4
Descriptive Statistics for the Control Group: Infant Illnesses

	Rash	**Gastroenteritis Illness**	**Respiratory Illness**	**Other Illness**	**Hospitalization**
	(1)	(2)	(3)	(4)	(5)
Any	0.20	0.13	0.67	0.37	0.25
	(0.40)	(0.33)	(0.47)	(0.48)	(0.43)
Multiple	0.07	0.04	0.44	0.18	0.09
	(0.26)	(0.20)	(0.50)	(0.39)	(0.28)

Note: The table shows means and standard deviations (in parenthesis) for indicators of any and multiple episodes of the specific infant illness measure during the first year.

Cognitive Development We examine three dimensions of cognitive development. First, the academic index at age six includes teachers' ratings of each child's performance in reading, writing, and mathematics. Previous PROBIT research has shown large improvements in IQ at age 6.5, using the pediatrician administered Wechsler Abbreviated Scales of Intelligence (WASI) (Kramer et al., 2008*b*). However, given pediatricians were not blind to treatment, we caution that the WASI measures may be associated with a potential measurement bias, as also pointed out by (Kramer et al., 2008*b*).[14] We therefore choose not to consider the WASI measures at age 6.5. Second, the IQ index at age 16 includes scores from the NeuroTrax cognitive tests for seven domains, measuring memory, executive function, visual–spatial perception, verbal function, attention, information processing, and fine motor skills. The tests were self-administered and computerized to minimize potential measurement bias caused by non-blinding of the pediatricians and within polyclinic correlations. Third, we consider whether the child is enrolled in the academic educational track at age 16. Finally, the cognitive index includes all eleven components of these three dimensions.

Academic Index At age six, teachers rated each child's academic performance in reading, writing, and mathematics on a five-point Likert scale as far below, somewhat below, at, somewhat above, or far above his or her grade level. We observe teacher ratings for about 60 percent of the children who are observed in the six year follow up; children without a teacher rating had not started school yet for the most part. Because of potential selection into the teacher sample as a consequence of the intervention, we choose to restrict the analysis sample to children who were supposed to have started school at the time they participated in the follow up assessment at the polyclinic. Children who had turned six years by September 1^{st} were supposed to have started school that year. Within this restricted sample of children, 15.1 percent do not have a teacher assessment. As the main reason for not having a teacher assessment for this restricted sample is still not hav-

[14] At age 6, study pediatricians assessed children's cognitive development using the Wechsler Abbreviated Scales of Intelligence (WASI) —the best available IQ test at the time. 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. The pediatricians were not blind to the treatment status, as they were involved in implementing the intervention. It would have been ideal to have had psychologists blind to the treatment status assess each child. However, that was practically impossible. Nowadays a second best solutions would have been to use computerized tests. But at the time of the assessment, neither the children nor the pediatricians knew how to use computers, and validated computer-assisted cognitive tests were not available. If they had used computerized tests, it could clearly have led to attenuation bias due to measurement error. Therefore, this was also not a possibility. Primarily due to the subjective nature of rating the child's responses, the WASI measures are associated with a high intraclass correlation coefficient (ICC) for cognitive scores of 0.31 (Kramer et al., 2008*b*). In the assessment, some pediatricians were very strict, while others were quite lenient. This led to a high variability in mean scores between polyclinics, even within each treatment group (Kramer et al., 2009). Kramer et al. (2008*b*) conducted an audit of 190 children. They 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).

ing started school, we assume that the reason for not being in school is that they were not school ready. We therefore impute their teacher rating with "below" their grade level, where we combine the teacher ratings "far below" and "somewhat below" to reduce the skewness of the distribution, as less than three percent of the graded children have a grade of far below. Appendix Figure A6 shows that our results are robust to alternative constructions of the academic index, both in terms of the restriction of the sample and the imputation. Teachers were unaware of the children's participation in the study (and therefore also of their treatment status) and had only three to four PROBIT children on average (Kramer et al., 2008*a*). Consequently, any potential bias in the teacher reports should be unrelated to the treatment status.

IQ Index 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 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.

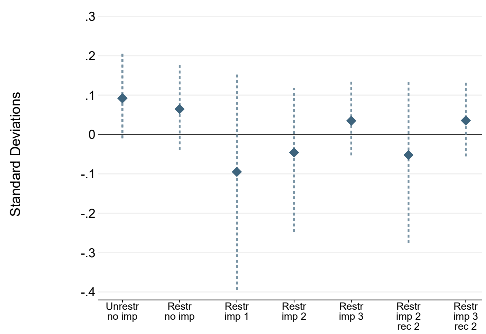
Academic Track The academic track indicated whether the child is enrolled in the academic educational track at age 16 (1) or not (0).

Cognitive Index The cognitive index includes the three components from the academic index, the seven components from the IQ index, and the academic track.

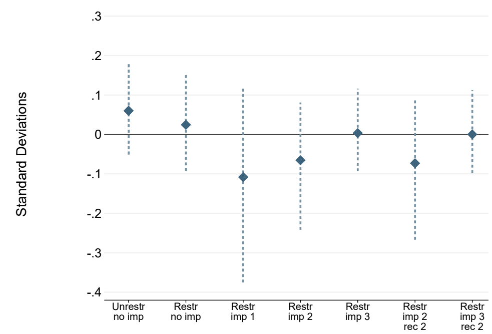
Socioemotional Skills Index The socioemotional skills summary index includes seven individual components. The measures are constructed based on parent reports at child age six to the Strengths and Difficulties Questionnaire (SDQ) and to supplemental behavioral questions taken from the Canadian National 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; we reverse the sign of the first four subscales so that

Figure A6
The Effect of Breastfeeding Promotion on Alternative Measures of Teacher's Academic Assessment

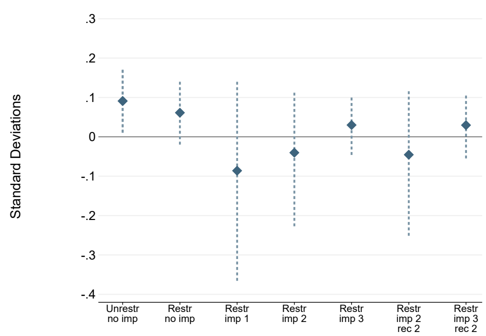
(a) Academic Index



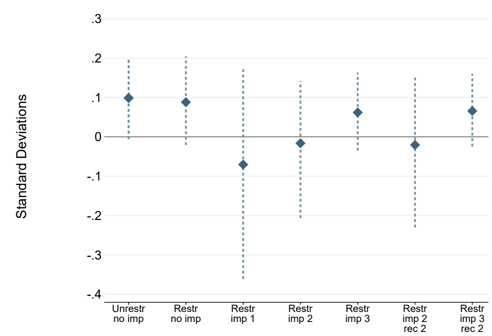
(b) Math



(c) Reading



(d) Writing



Note: Teachers rated each child's academic performance in reading, writing, and mathematics on a five-point Likert scale as far below (1), somewhat below (2), at (3), somewhat above (4), or far above (5) his or her grade level. All outcome variables are stanradized so the control group has a mean of zero and a standard deviation of one. The sample is restricted to all children with at least one measurement in wave 2. The restricted sample only includes children who were supposed to be in school. The first of the seven estimates is without this restriction. The second estimate comes from the restricted sample without any imputation. From the third to the fifth estimate, missing values in this restricted sample are imputed with far below grade (1), somewhat below grade (2), or at grade level (3), respectively. The sixth and seventh estimates additionally recode far below grade (1) with somewhat below grade (2) and impute missing values in this restricted sample with somewhat below grade (2) and at grade level (3) respectively. 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: $\circ p < 0.10$, $\bullet p < 0.05$, $*p < 0.01$.

a higher score is more favorable. Based on the questions from the NLSCY, we construct two summary indices capturing respectively externalizing (five questions) and internalizing (ten questions) behavioral problems; we reverse the sign of these measures so that a higher score is more favorable. The summary index includes the seven components: the five SDQ subscales and the two NLSCY behavioral problem indices.

A.4 Original Analysis Plan

When applying for access to the PROBIT data, we included a study protocol outlining our plan for the empirical analysis. Our work commenced with three original aims that revealed null or no novel results. For transparency, we here describe the original three aims and include the results in the Appendix.

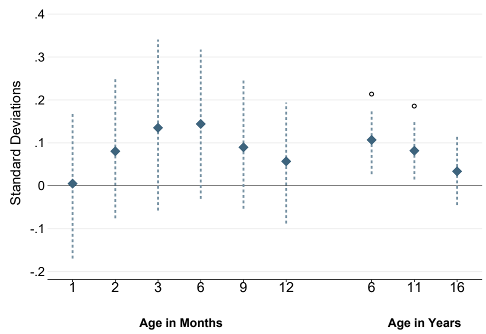
First, we planned to construct summary indices that aggregate information over multiple treatment effect estimates —in particular, summary indices of health (Appendix Figures A7e and A8a) and human capital (Appendix Figures A9a and A9d). We had also planned to consider a parental investment index; however, the data proved not to provide sufficient variation on parental inputs. Second, we proposed exploring heterogeneity in the effects of breastfeeding by socioeconomic status and child gender (see Appendix Section A.6). Third, to convince (particularly) economists about the validity of a causal interpretation of the intervention, we planned to provide extensive randomization tests (Appendix Table A1). In terms of evaluating the breastfeeding intervention, our primary plan was to estimate the intention to treat (ITT) effects, as done in the main analysis (see Section 3). We also mentioned that we would scale the magnitude of the effects of exclusive (or any) breastfeeding for at least 3 and 6 months respectively by using the random assignment of treatment as an instrumental variable (IV). However, this analysis did not provide any additional insights for two reasons. First, the point estimates using IV are even more noisily estimated than for ITT, as shown in existing PROBIT papers (Oken et al., 2013; Owen et al., 2018; Martin et al., 2013, 2017; Yang et al., 2018). Second, given the results from our infant feeding analysis, we are not convinced that the exclusion restriction is valid in this particular context.

A.5 Effects on Child Health and Development

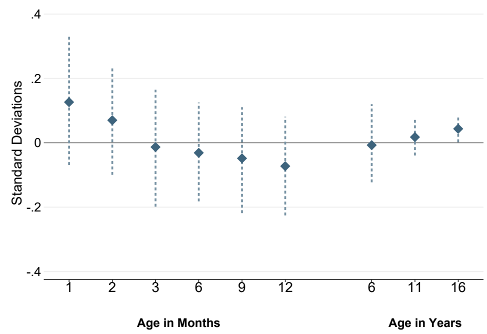
In this appendix section, we report results on child health and development. Overall, all results are in line with previous PROBIT research and we therefore chose not to include

Figure A7
The Effect of Breastfeeding Promotion on Physical Growth

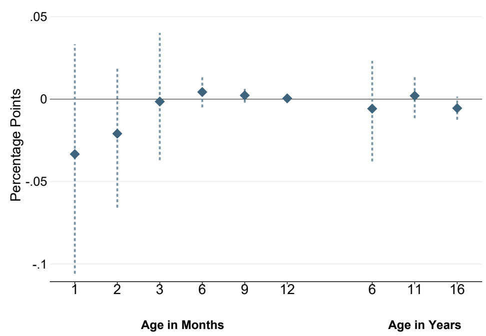
(a) Height-for-age



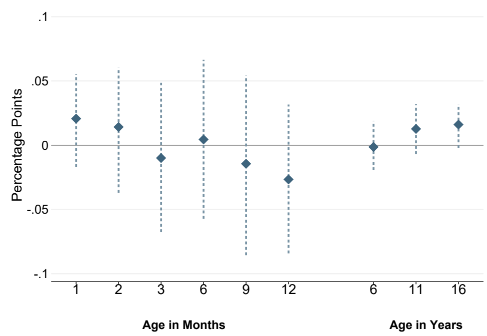
(b) Weight-for-Length / BMI



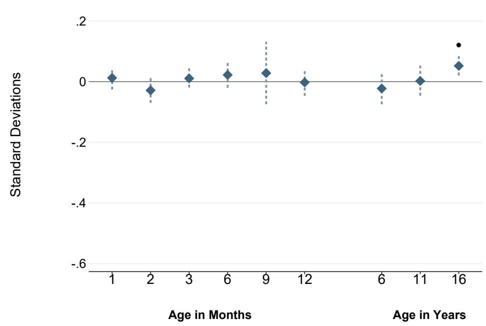
(c) Probability of Underweight



(d) Probability of Overweight



(e) Physical Growth Index



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: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

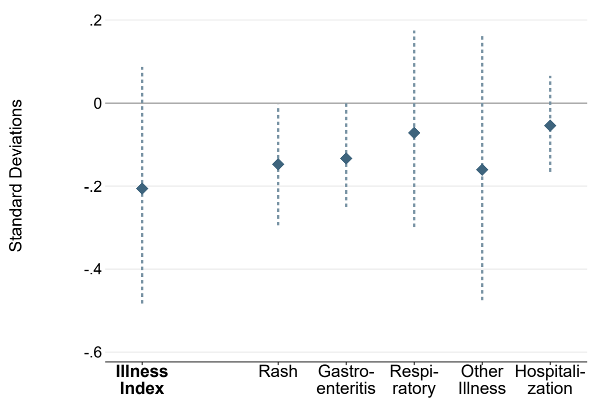
them in the main paper for brevity.

Appendix Figure A7 shows the effects of the breastfeeding promotion intervention on physical growth from age one month throughout childhood. Overall, we do not find consistent or persistent effects on height-for-age, weight-for-length/BMI, the probability of being overweight or underweight, or the physical growth index. These results are overall consistent with previous findings from the PROBIT study (Kramer et al., 2002, 2007; 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. The estimate of the intervention on the probability of being overweight at age 16 is 1.61 percentage points with a 95 percent confidence interval [-0.20; 3.21] (see Appendix Figure A7d). This corresponds to an increased risk of being overweight by 11 percent compared to the control group, which is comparable to the estimate of 14 percent (CI [2; 28]) found in Martin et al. (2017). The main reason for us finding a statistically insignificant effect compared to Martin et al. (2017) is the difference in the way of clustering the standard errors. In the main text, we find a statistically significant and persistent effect of the intervention on weight-for-age (Figure 1). In infancy, this is in line with Kramer et al. (2002) who find positive effects on weight at ages 3, 6, and 9 months, but not at age 12 months as we also do not find. During childhood, Martin et al. (2017) find positive effects on weight gain from age 2.8 to 14.5 years (Table 3) and on weight at ages 6.5, 11.5, and 16.2 (Figure 2A) —again in line with our findings.

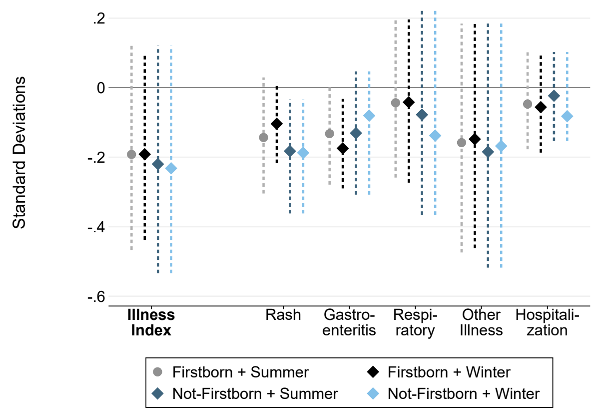
Appendix Figure A8 shows the effects of the breastfeeding promotion intervention on infant illness. Although our summary index and its individual components are not statistically significant, the estimates suggest that the breastfeeding intervention decreased the likelihood of infant illness. The point estimates are all negative and we cannot rule out medium-sized reductions in illness of up to 0.17-0.48 standard deviations. For example, the point estimate for the illness index is -0.21 (p-value$=$ 0.16) and its 95 percent confidence interval ranges from -0.48 to 0.09 standard deviations. The point estimates for rash and gastroenteritis are -0.15 (95 percent confidence interval not adjusted for multiple hypothesis testing $[-0.30; -0.01]$) and -0.13 ($[-0.25; 0.01]$) standard deviations, respectively. We do not find any heterogeneity in the effects on dimensions proxying for the infectious environment; Appendix Figure A8b shows no meaningful differences across subgroups split by parity and season of birth. Overall, the results are consistent with Kramer et al. (2001), finding negative effects on the risk of any gastrointestinal tract infection, atopic eczema, and rash in the first 12 months. Despite

Figure A8
The Effect of Breastfeeding Promotion on Infant Illness

(a) Main effects



(b) Heterogeneity by Infectious Environment



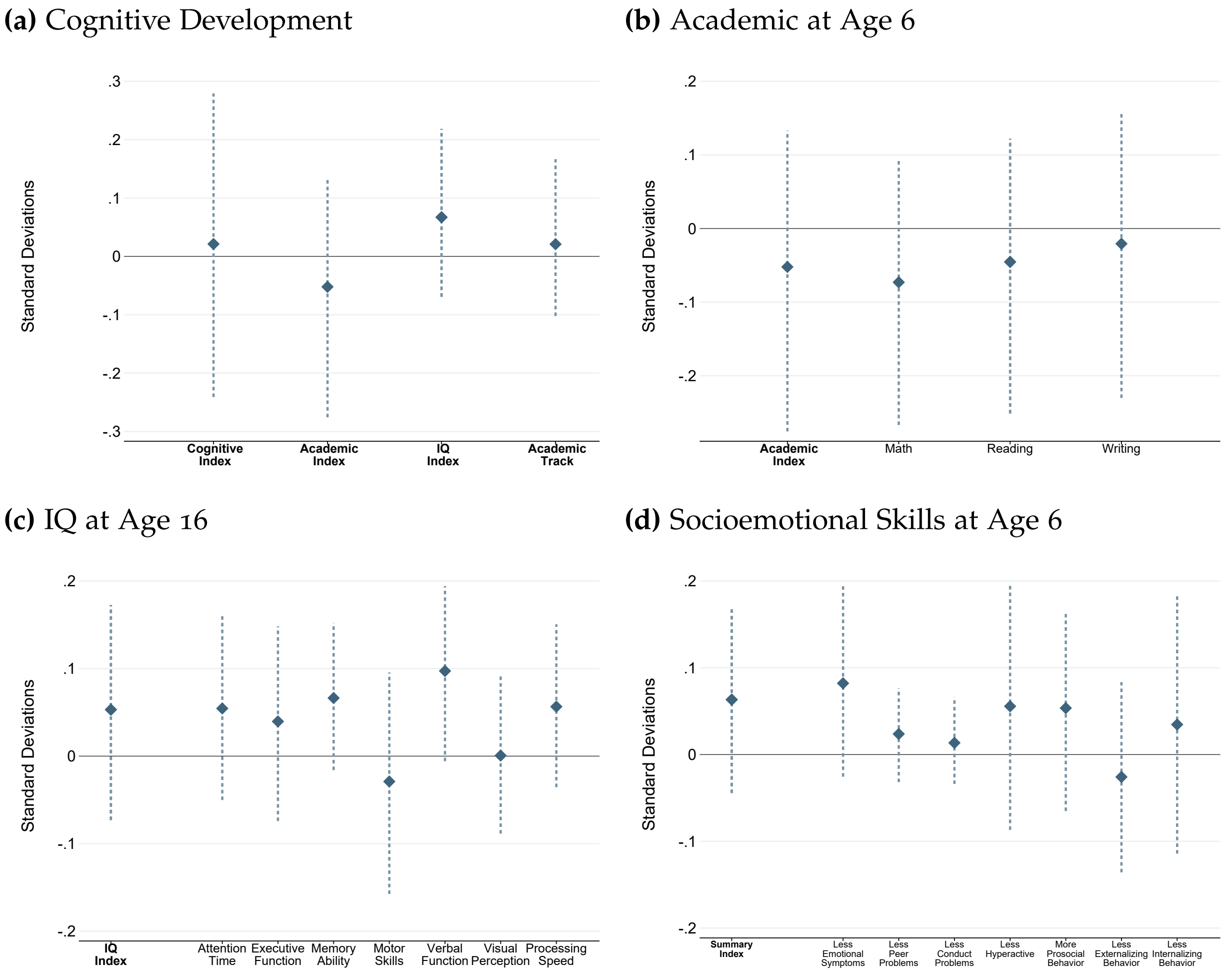
Note: Each estimate comes from a separate regression as specified in equation (1); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. The models in graph A8a also control for indicators of age at last visit in months 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: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

differences in precision, Kramer et al. (2001)'s and our effect sizes for rash (-36 vs. -29 percent) and gastroenteritis (-38 vs. -34 percent) are relatively similar when expressed as percent changes.

Consistent with (Kramer et al., 2008*b*), we do not find statistically significant effects on teachers' assessment of children's academic performance at age 6 (Appendix Figure A9b). Consistent with (Yang et al., 2018), we do not find statistically significant effects on IQ at age 16 (Appendix Figure A9c). The overall cognitive index also shows little statistical evidence of effects (Graph A9a) with a point estimate of 0.02 standard deviation—though with a wide 95 percent confidence interval ranging from -0.24 to 0.28. Consistent with (Kramer et al., 2008*a*), we also do not find statistically significant effects on children's socioemotional skills at age 6 (Appendix Figure A9d).

We find little evidence that this breastfeeding intervention has significant effects on our measures of cognitive development in Belarus, although such effects are theoretically possible. Any effect of breastfeeding on cognitive development is typically hypothesized to be due to changes in the social stimulation of the child or due to the particular composition of breast milk (Fitzsimons and Vera-Hernández, 2021). The most important nutritional difference between formula and breast milk is breast milk's content of two long-chain polyunsaturated fatty acids—docosahexaenoic acid (DHA) and arachi-

Figure A9
The Effect of Breastfeeding Promotion on Cognitive and Socioemotional Development



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: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

donic acid (ARA)—that are crucial for infants' neurocognitive development (Krol and Grossmann, 2018). However, infants produce insufficient amounts of these fatty acids (Cockburn, 2003). This is the reason for adding DHA and ARA to infant formula, which is standard in high-quality infant formula today, but was not the case in our setting nor in the one of Fitzsimons and Vera-Hernández (2021)—the only other published causal study of breastfeeding on cognitive development outside of the PROBIT RCT.[15]

[15]Since 1996 (2002), the European Union (United States) has allowed the addition of DHA and ARA to

But in contrast to our results, Fitzsimons and Vera-Hernández (2021) find beneficial effects on cognitive skills among children of low educated mothers at ages 3–7 in the United Kingdom. In their investigation of mechanisms, they do not find robust effects on child health, socioemotional skills, or any factors related to maternal investments or attachment. This suggests that the positive effect of breastfeeding on cognitive skills in their setting could be due to the nutritional composition of breast milk rather than changes in the social environment. One key difference between these two settings is the prevalence of any breastfeeding among the control group, and thereby the intake of some quantity of DHA and ARA. Most mothers in Fitzsimons and Vera-Hernández (2021) only breastfed (partially) for a few weeks or did not initiate breastfeeding at all, instead relying on almost exclusive use of infant formula without added DHA and ARA. In contrast, most control mothers in Belarus partially breastfed for at least three months. As seen in Figure 2, mothers in the control group still breastfed an average of 3.5 times per day at three months, so infants still ingested some amount of these fatty acids through breast milk.

Our results suggest that increasing the frequency of breastfeeding and the duration of breastfeeding exclusivity does not improve cognitive development. This result is important for policy discussions about the optimal duration of maternity leave and lactation policies at the workplace. The results also suggest that adding DHA and ARA to infant formula might remove or at least substantially reduce any beneficial effect of breastfeeding on cognitive development among healthy infants (Cockburn, 2003; Hoffman, Boettcher and Diersen-Schade, 2009; Lien, Richard and Hoffman, 2018). While more causal work is needed before drawing strong conclusions, this would have important policy implications because governments regulate the composition of infant formula. To the extent that these additives are much more commonly found in high-quality infant formula today, our estimates of the benefits of breastfeeding may reasonably represent an upper bound of any potential beneficial effect of breastfeeding on cognitive development in settings where high-quality infant formula is widely used.

formula and made the addition of DHA mandatory in 2020.

A.6 Heterogeneity

In this appendix, we report all the 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.

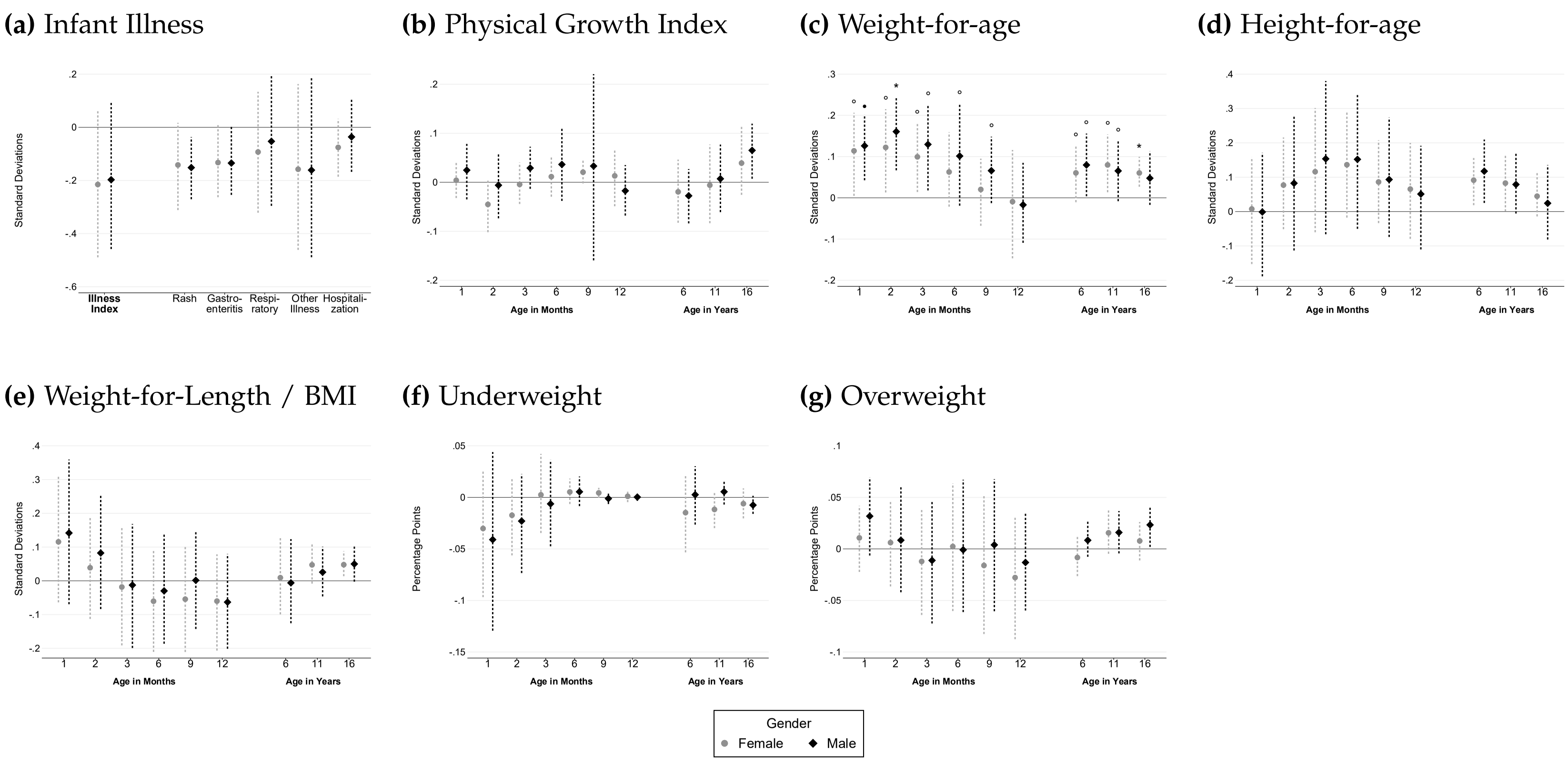
To estimate heterogeneous effects, we modify the empirical specification given in equation (1). We estimate fully interacted models, in which we interact all control variables with the particular dimensions of interest. For the estimation of heterogeneity with respect to gender, the specification is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

$$
\begin{aligned}
Y_{iph} = &\gamma_0^{female} Female + \gamma_1^{female} Treatment_h \times Female + \gamma_1^{male} Treatment_h \times Male + \\
&Female \times Z_i' \delta^{female} + Male \times Z_i' \delta^{male} + \theta_p^{female} + \theta_p^{male} + \varepsilon_{iph},
\end{aligned} \tag{4}
$$

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 A10
The Effect of Breastfeeding Promotion on Infant Illness and Childhood Physical Growth: Heterogeneity by Gender

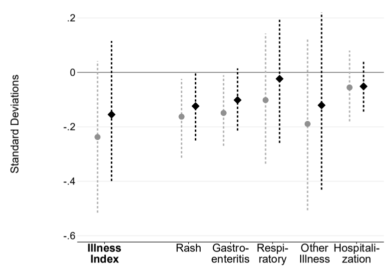


Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

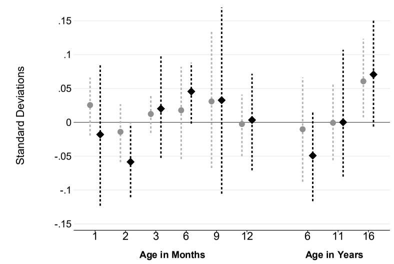
Figure A11

The Effect of Breastfeeding Promotion on Infant Illness and Childhood Physical Growth: Heterogeneity by Socioeconomic Status

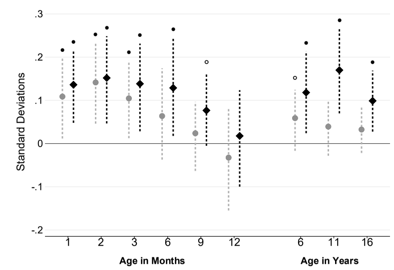
(a) Infant Illness



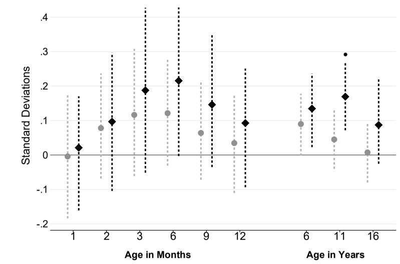
(b) Anthropometric Index



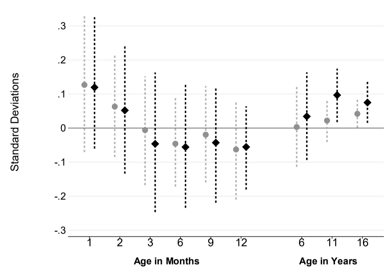
(c) Weight-for-age



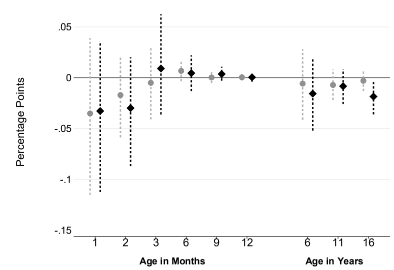
(d) Height-for-age



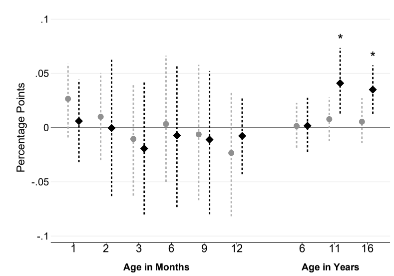
(e) Weight-for-Length / BMI



(f) Underweight



(g) Overweight

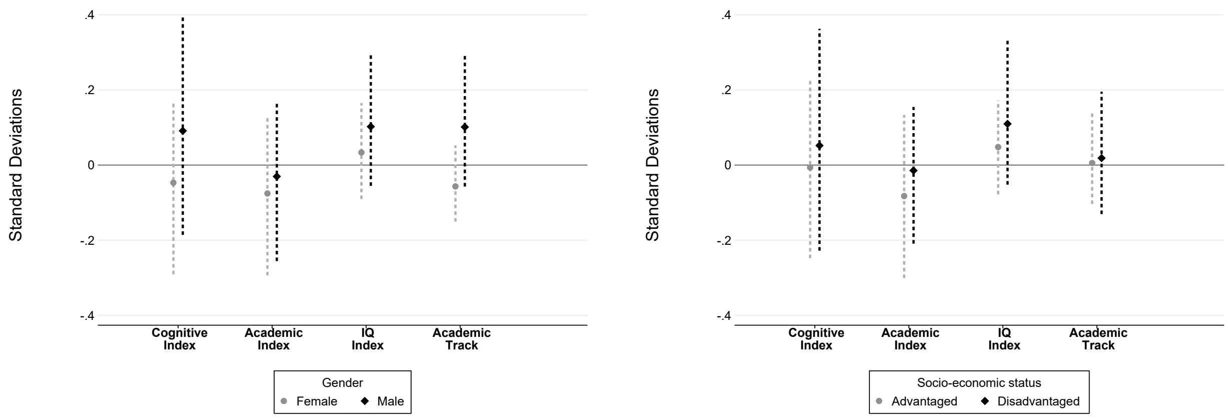


Socio-economic status
● Advantaged ◆ Disadvantaged

Note: Each set of estimates comes from a separate regression as specified in equation (4); 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

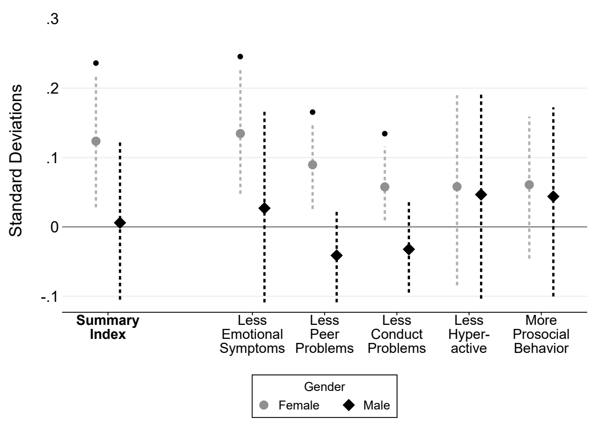
The Effect of Breastfeeding Promotion on Cognitive Development:
Heterogeneity by Gender and Socioeconomic Status



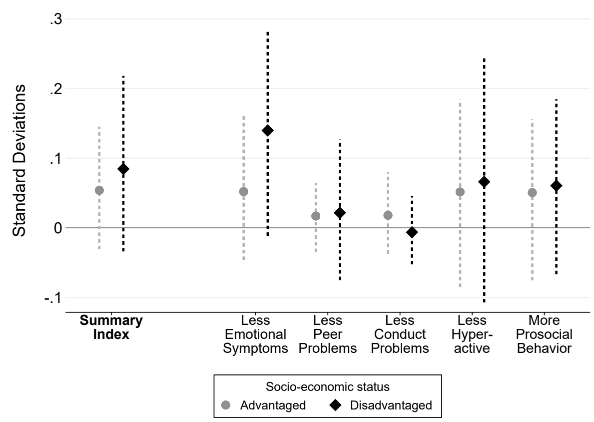
Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed 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 Socialemotional Skills:
Heterogeneity by Gender and Socioeconomic Status

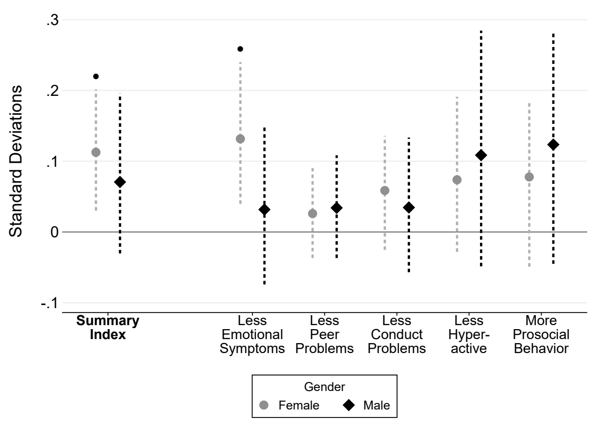
(a) Parent-reported, by Gender



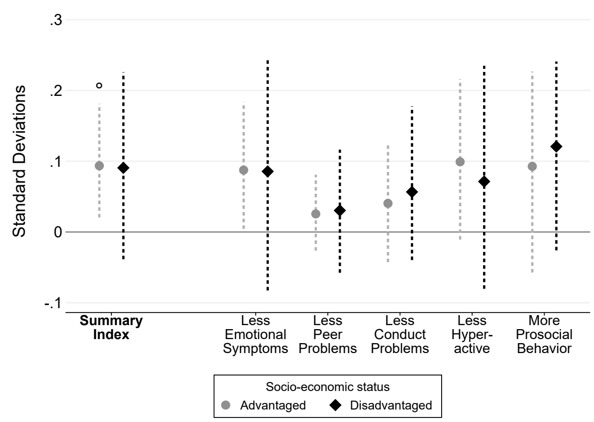
(b) Parent-reported, by SES



(c) Teacher-reported, by Gender



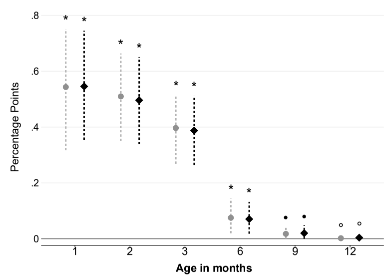
(d) Teacher-reported, by SES



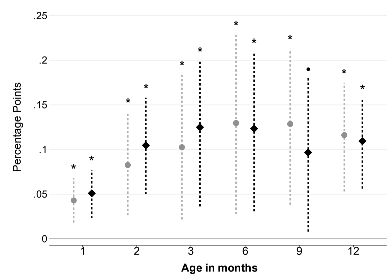
Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed 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 A14
The Effect of Breastfeeding Promotion on Infant Feedings: Heterogeneity by Gender

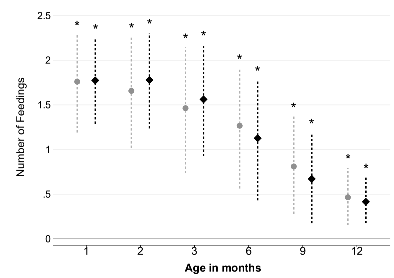
(a) Exclusive Breastfeeding



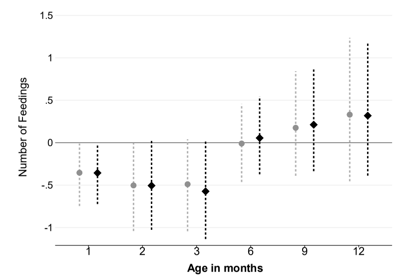
(b) Any Breastfeeding



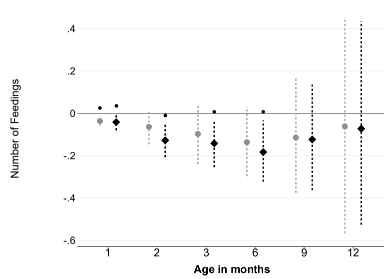
(c) Breastfeeding



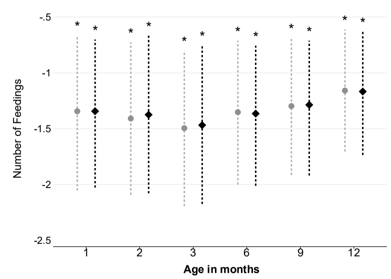
(d) Infant Formula



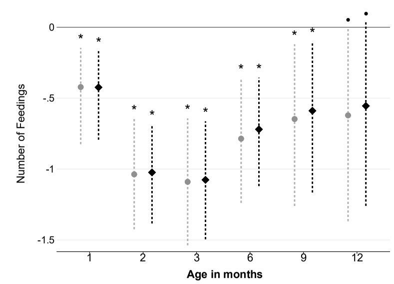
(e) Cow's Milk



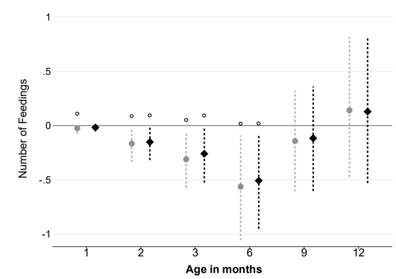
(f) Water

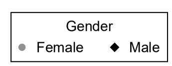


(g) Juice or Other Liquids



(h) Solid food

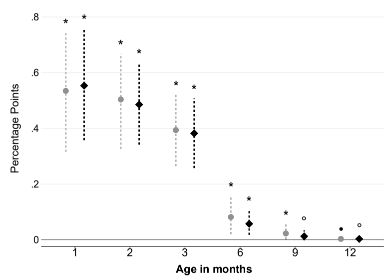




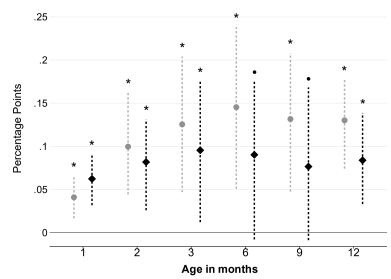
Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed 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 Infant Feedings: Heterogeneity by Socioeconomic Status

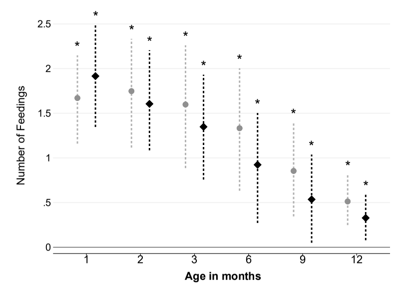
(a) Exclusive Breastfeeding



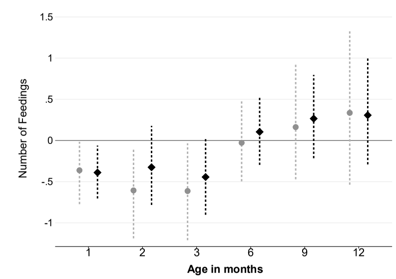
(b) Any Breastfeeding



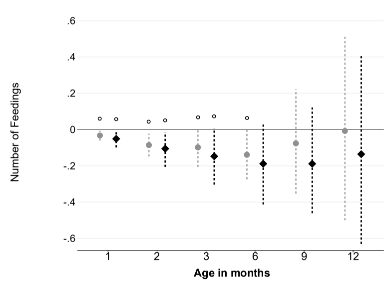
(c) Breastfeeding



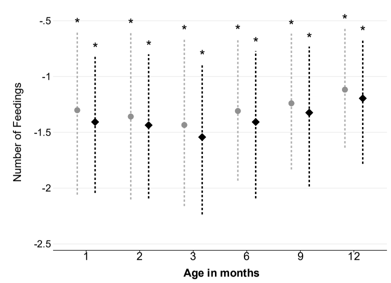
(d) Infant Formula



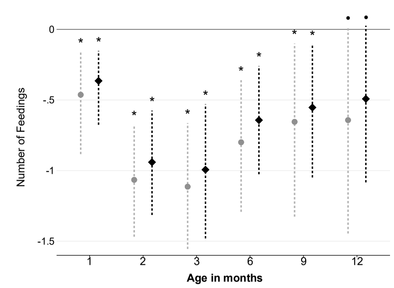
(e) Cow's Milk



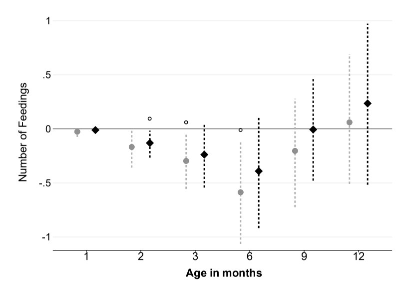
(f) Water



(g) Juice or Other Liquids



(h) Solid food

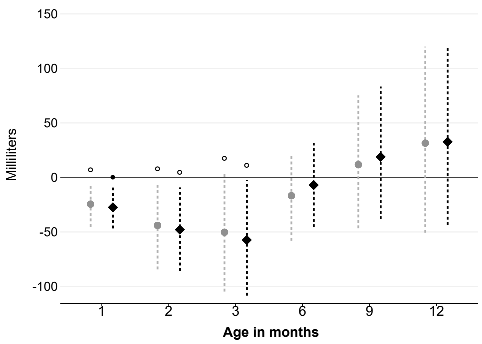


Socio-economic status
● Advantaged ◆ Disadvantaged

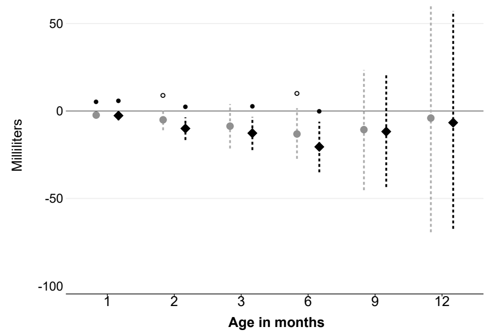
Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed 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 A16
The Effect of Breastfeeding Promotion on Estimated Infant Liquid Calorie Intake, excluding Breast Milk (BM): Heterogeneity by Gender

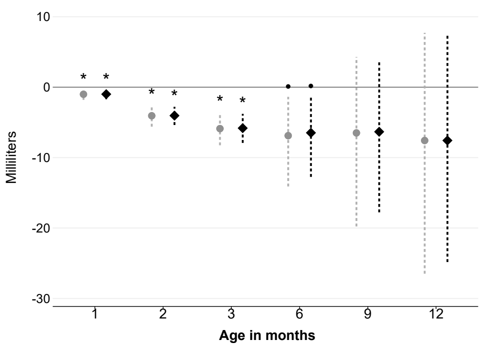
(a) Calories from Formula



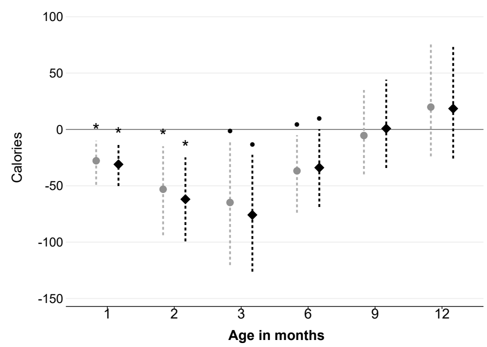
(b) Calories from Cow's Milk



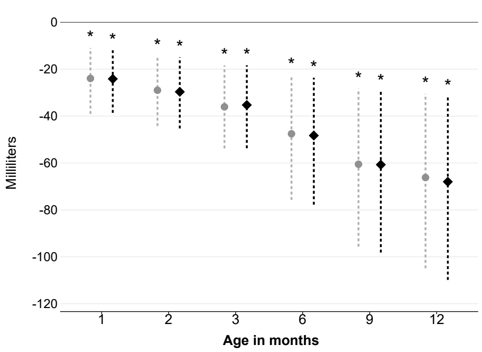
(c) Calories from Juices



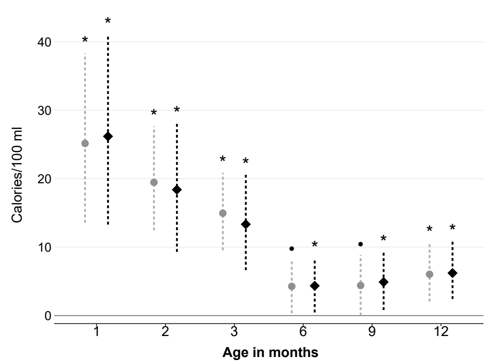
(d) Calories from Liquids (excl. BM)

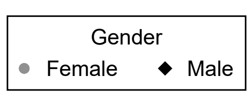


(e) Volume of Water



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



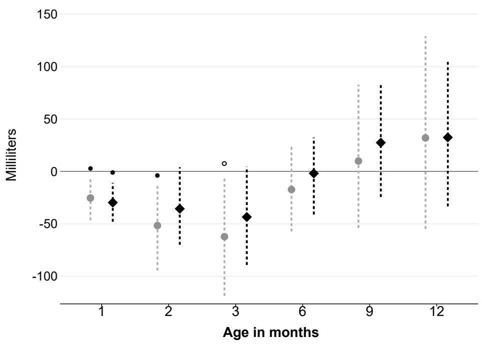


Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: $\circ p < 0.10$, $\bullet p < 0.05$, $^{*}p < 0.01$.

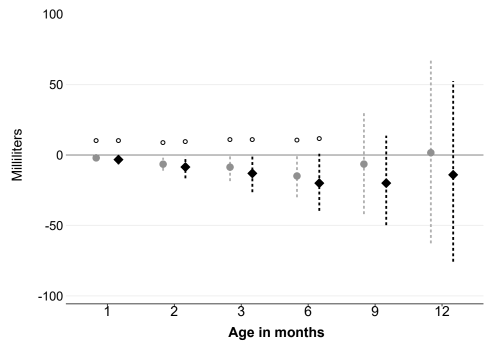
Figure A17

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

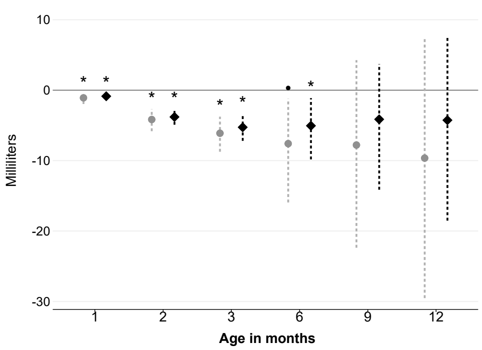
(a) Calories from Formula



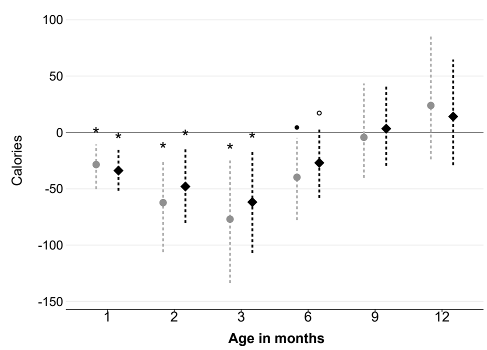
(b) Calories from Cow's Milk



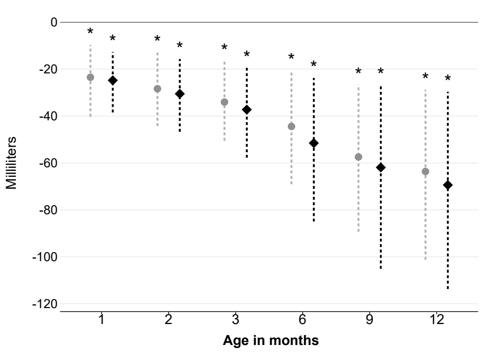
(c) Calories from Juices



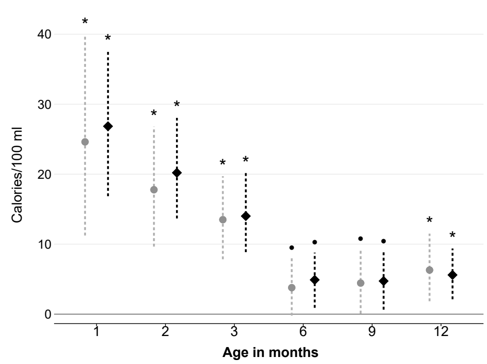
(d) Calories from Liquids (excl. BM)



(e) Volume of Water



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

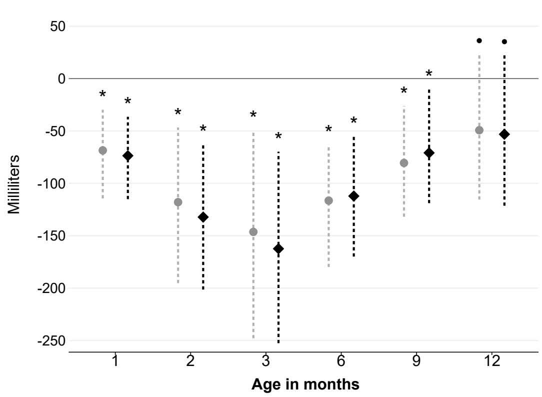


Socio-economic status
● Advantaged ◆ Disadvantaged

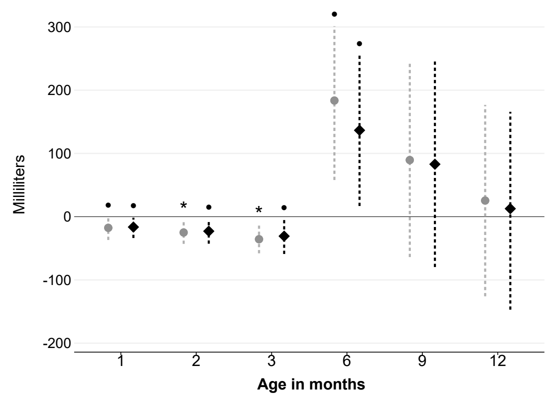
Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed 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 Gender

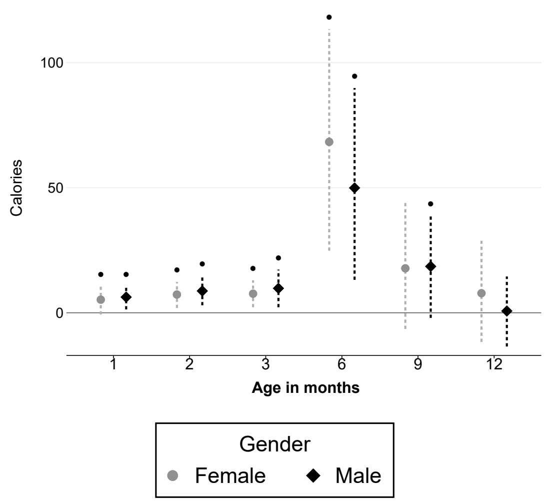
(a) Total Volume Liquids (excl. BM)



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



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

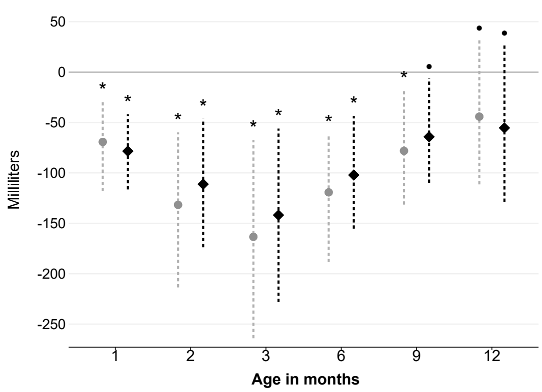


Note: Each set of estimates comes from a separate regression as specified in equation (4); the dashed lines show the 95 percent confidence interval based on wild cluster bootstrapped (WCB) standard errors clustered at the hospital level. Multiple hypothesis testing using the *krieger* method is performed on all estimates within the same graph. Significance levels after testing for multiple hypothesis are indicated as follows: ∘$p < 0.10$, •$p < 0.05$, *$p < 0.01$.

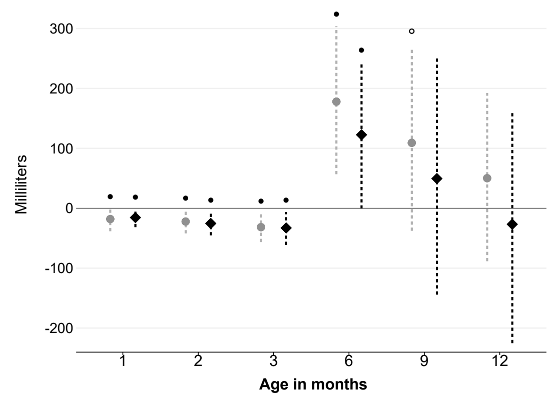
Figure A19

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

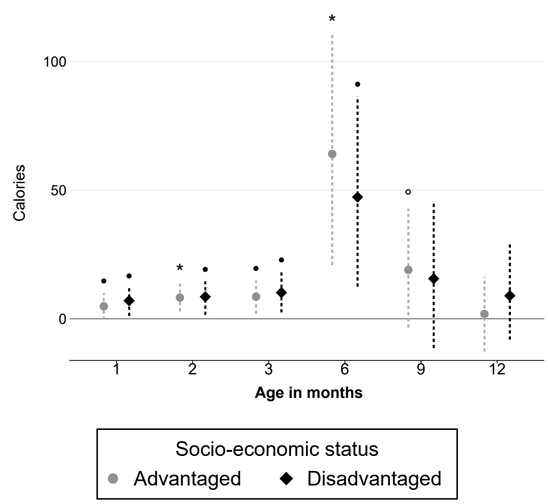
(a) Total Volume Liquids (excl. BM)



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



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



Note: Each set of estimates comes from a separate regression as specified in equation (4); 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.7 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.