

MS. NIVITA D SHARMA (Orcid ID : 0000-0002-0974-9363)

Article type : Original

**Breastfeeding and the Risk of Childhood Asthma: A Two-Stage Instrumental Variable Analysis
to Address Endogeneity**

Nivita Divya Sharma

Moore College

108 Herty Drive

University of Georgia

Athens, GA 30602

USA

Running title: Breastfeeding and childhood asthma

Corresponding Address

Nivita D. Sharma

Moore College

108 Herty Drive

University of Georgia

Athens, GA 30602

USA

Email: nivita.sharma97@uga.edu

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pai.12750

This article is protected by copyright. All rights reserved.

Breastfeeding and the Risk of Childhood Asthma: A Two-Stage Instrumental Variable Analysis to Address Endogeneity

Pediatr Allergy Immunol

Abstract

Background: Several explanations for the inconsistent results on the effects of breastfeeding on childhood asthma have been suggested. The purpose of this study is to investigate one unexplored explanation, which is the presence of a potential endogenous relationship between breastfeeding and childhood asthma. Endogeneity exists when an explanatory variable is correlated with the error term for reasons such as selection bias, reverse causality, and unmeasured confounders. Unadjusted endogeneity will bias the effect of breastfeeding on childhood asthma.

Methods: To investigate potential endogeneity, a cross-sectional study of breastfeeding practices and incidence of childhood asthma in 87 pediatric patients in Georgia, USA was conducted using generalized linear modeling (GLM) and a two-stage instrumental variable analysis. First, the relationship between breastfeeding and childhood asthma was analysed without considering endogeneity. Second, tests for presence of endogeneity were performed and having detected endogeneity between breastfeeding and childhood asthma, a two-stage instrumental variable analysis was performed. The first stage of this analysis estimated the duration of breastfeeding and the second-stage estimated the risk of childhood asthma.

Results: When endogeneity was not taken into account, duration of breastfeeding was found to significantly increase the risk of childhood asthma (relative risk ratio (RR) = 2.020, 95% CI (1.143-3.570)). After adjusting for endogeneity, duration of breastfeeding significantly reduced the risk of childhood asthma (RR = 0.003, 95% CI (0.000-0.240)).

Conclusions: The findings suggest that researchers should consider evaluating how the presence of endogeneity could affect the relationship between duration of breastfeeding and the risk of childhood asthma.

Keywords: asthma; atopy; breastfeeding; causation; children; endogenous; epidemiology; instrumental variable; two-stage.

Introduction

Decades of research on the effect of duration of breastfeeding on childhood asthma, a chronic illness inflicting millions in developed countries, remains inconclusive (1-4). Although the World Health Organization (WHO), the United Nations Children's Fund (UNICEF) and the American Academy of Pediatrics (AAP) recommend six months of exclusive breastfeeding, many studies show that breastfeeding may actually increase the risk of asthma, some show breastfeeding has no effect, and others show a modest protective effect (1-4).

Several meta-analysis studies suggest the inconsistent and inconclusive results on breastfeeding and childhood asthma could be due to a number of research issues such as the inability to conduct controlled experiments, the sampling criteria used, and the analysis model (1-4). One limitation due to lack of control over the experiment and sampling criteria is that researchers often are unable to randomly select and randomly assign participants into treatment groups. This inability can lead to serious unintended consequences such as unmeasured confounders, selection bias, and reverse causality (1-4). If selection bias, reverse causality and unmeasured confounders are not appropriately addressed in the statistical analyses then the estimate of the coefficient, odds ratio or relative-risk ratio related to duration of breastfeeding will be biased (1).

In the literature, the most common approach to control for the consequences of selection bias is to identify, measure, and adjust for potential confounders (1-4). These confounders must be related to both the treatment and health outcome. However, because of

Accepted Article
sampling criteria and data limitations, researchers are not always able to statistically adjust for a comprehensive set of confounders (1-2). For example, prior research has not statistically controlled for often unquantifiable behavioral and psychological factors related to breastfeeding although these can affect a mother's decision to breastfeed, which in turn can affect childhood asthma (1-3, 5). If mothers of children with a family history of asthma are likely to breastfeed more believing that breastfeeding will protect their child against asthma or because highly educated mothers value breastfeeding, then a positive or no relation between breastfeeding and subsequent childhood asthma could be perceived as breastfeeding "causing" a diagnosis of childhood asthma or breastfeeding is not protective of childhood asthma (1-2). In contrast, mothers of children with a family history of asthma may decide to not breastfeed for the fear she may be "passing" asthma to the child (5-6), which can lead to the belief that not breastfeeding prevents asthma. As a further example, mothers suffering from postpartum depression may not breastfeed the child and engage in other unhealthy behaviors such as smoking that increases the risk of the child developing asthma (7). Non-breastfeeding mothers who engage in other unhealthy behaviors could heighten the risk of the child developing asthma (8).

Thus, a second limitation of observational studies on breastfeeding and childhood asthma is that prior research uses statistical techniques that assume breastfeeding is an exogenous variable when in fact breastfeeding could be correlated with unadjusted confounders related to childhood asthma. If this occurs, then the statistical relationship between duration of breastfeeding and childhood asthma is affected by the presence of endogeneity between breastfeeding and childhood asthma (9-12). Endogeneity is a relatively serious statistical problem because it violates the assumption that breastfeeding is exogenous and uncorrelated with the error term, and this violation can bias the statistical effect of breastfeeding on asthma (1, 10-12). If the exposure variable breastfeeding is correlated with

the error term of the asthma model, then it is an endogenous variable. Since the error term represents the effect of unadjusted confounders, the presence of endogeneity in this context suggests the presence of vital correlated omitted variables (10-12). However, it is important to note that there will always be potential omitted variables because researchers cannot adjust for all possible confounders. Therefore, addressing endogeneity can help minimize the effects of unadjusted confounders and enhance a study's internal validity (10-12).

A potential solution to deal with endogeneity is instrumental variable (IV) analysis, that involves executing a two-stage statistical analysis (10-12). A two-stage analysis is guided by theory and requires the identification of an IV. The two-stage IV analysis has a long history in econometrics originating in path analysis and structural equation modelling, and is becoming popular in epidemiological research (10-12). A two-stage analysis is also a powerful technique to address reverse causality (10-12), which some meta-analysis studies are beginning to recommend for addressing statistical limitations in prior breastfeeding and asthma studies (2).

The IV is a new independent variable that is correlated with the treatment or exposure (breastfeeding), uncorrelated with the entire set of unmeasured confounders (error term or residuals), and relatively independent of the outcome (childhood asthma risk). Once the IV is identified, a two-stage analysis is performed. In this study, the first-stage equation estimates duration of breastfeeding on mother's education, family history of asthma, patient characteristics, and the IV. Then in the second-stage, the predicted value of duration of breastfeeding from the first-stage equation is used to explain the risk of childhood asthma along with potential confounders (10-12).

Addressing the presence of potential endogeneity in the statistical relationship between breastfeeding and childhood asthma is an important objective of this study as it can help clarify the conflicting evidence on the beneficial effects of breastfeeding on childhood asthma.

Methods

Participants

Participants in this study were drawn from a population of pediatric patients of one physician at The Kaiser Permanente TownPark Comprehensive Medical Center in Kennesaw, Georgia, USA. Kaiser Permanente is one of the largest national not-for-profit health care provider with about 10.1 million members in 2015. However, the Kennesaw Medical Center is relatively small. Approximately 520 pediatric patients aged 3 to 17 years from English-speaking families who made office visits between 1pm and 4pm from Monday to Thursday in Fall 2013 were potential participants. The minimum age was set at 3 to minimize misdiagnosis of asthma that tends to occur in younger children (2). Only biological mothers bringing their children to the hospital were invited to voluntarily participate because other accompanying adults may not be able to answer the breastfeeding questions.

To minimize health condition confounding effects, participants were invited if they did not have a history or currently suffer from a serious health condition other than asthma, wheezing, atopic dermatitis, and respiratory allergies. A serious health condition included any one of the following top diseases affecting children in the United States: heart disease, cancer, diabetes, neoplasms, congenital abnormalities, cerebrovascular, influenza and pneumonia, septicemia, and HIV (13). Therefore, all participants made office visits for wellness/physical checks or a common health condition such as ear infection, sore throat, and cough. To avoid any further medical condition selection bias, participants were not pre-selected based on their history of asthma, breastfeeding, or any other health condition. Data were gathered using a detailed survey questionnaire and through electronic patient records. The primary purpose of the questionnaire was to gather information about the duration of breastfeeding and potential confounders.

This study was administratively reviewed and approved by Kaiser Permanente where the study was conducted. The study was exempt from formal Institutional Review Board (IRB) approval because the primary method of data collection was a survey administered to adult mothers of pediatric patients. The participants were also not identified or referred to in the study. The study was also not subject to any Food and Drug Administration (FDA) regulations and did not cause participants either physical or psychological discomfort or perceived as such based on a post-questionnaire interview. Verbal informed consent was obtained from mothers, which was confirmed upon their completion of the questionnaire.

The questionnaire used in this study (see Appendix) was adapted based on items reported in prior studies (14-15) and supplemental questions to measure potential confounders were added to address limitations of prior studies (2-3). The questionnaire was reviewed by Kaiser Permanente, which led to slight modifications in framing of questions to reduce sensitivity. A pilot test using healthcare professionals and five mothers led to cosmetic modifications.

Breastfeeding duration and exclusiveness

For the purposes of this study, breastfeeding variables following best practice recommendations of the WHO, CDC and AAP were included. Consistent with the prior literature (14-15), the mother was asked to select one of eight categories of the duration of breastfeeding ranging from “never” to “more than 12 months.” According to the CDC (16), breastfeeding rates are relatively low after 12 months, therefore, the last category was “more than 12 months.” To avoid formula milk confounding the measure of duration of exclusive breastfeeding (DEBF) (1-2), the questionnaire also asked the mother to indicate on a similar eight category scale ranging from “less than one month” to “never,” when the infant was introduced to formula milk. The responses on both the breastfeeding question and the formula milk question were compared to derive DEBF. For example, if a mother checked the 6-9 months breastfeeding category and also checked the 3-6 months category for formula

milk, then DEBF was conservatively determined to be 3 months. A further advantage of providing mothers a pre-determined set of breastfeeding and formula feeding options to select from is that it reduces recall bias (15).

Asthma outcomes

Incident childhood asthma cases were identified from electronic patient records indicating physician diagnosed presence of asthma identified by the *International Classification of Diseases, Ninth Revision* (ICD-9). To ensure reliable measurement of the outcome, records of non-asthma patients were screened to ensure they were not developing asthma (1).

Covariates

Information on various demographic variables such as gestational age, date of birth, sex, race/ethnicity, exposure to smoke including perinatal smoking, whether child lives with a single parent, and residential zip code were obtained from electronic patient records. Some of these questions were also asked of mothers to cross-validate with patient records; no differences were found between the mothers' responses and patients' records. Information on biological family history of asthma and allergies were obtained from questionnaire responses (14-15). Parental demographics included mother's age at the time of patient's birth and education level. To control for household income, the residential zip code income level from county records was used (17). Data on potential lifestyle confounders obtained through the questionnaire included child's physical activity and exposure to pets/animals (18-19).

Statistical analysis

Demographic and clinical characteristics were summarized using frequencies or means and standard deviations. The presence of multicollinearity in the independent variables was analysed using Spearman and Pearson correlations and variance-inflation-factors (VIFs). The highest correlation coefficient was 0.528 and all others were below 0.332 with most being less than 0.10. The VIFs confirmed the absence of multicollinearity as none was above 3.0. In this study, a two-stage

multivariate analysis was performed due to potential endogeneity between DEBF and risk of asthma.

The first-stage involved predicting the DEBF. Potential predictors of DEBF were identified from prior studies that affected both DEBF and risk of childhood asthma, and were restricted to the following to maintain at least 10 observations per predictor (20): education of biological mother, premature birth of patient, whether the biological mother or older siblings were ever physician diagnosed with asthma, and whether the biological mother or older siblings were ever physician diagnosed with an allergy.

It is quite challenging, if not impossible, to identify an IV that affects DEBF but not the risk of childhood asthma because of the inconclusive evidence on the determinants of DEBF and childhood asthma. To qualify as a suitable IV, the IV should meet three conditions: (1) it should be correlated with the explanatory (exposure) variable, which in this case is DEBF, (2) it should not be correlated with unadjusted confounders of childhood asthma (i.e., the error term), and (3) it should be relatively independent of the outcome (i.e., childhood asthma) (10-12). As the IV must be selected based on theory and subject-matter knowledge, and not statistical testing, the inconclusive results in the literature make this quite challenging (11). After reviewing the vast literature on breastfeeding and asthma, it was determined that neighborhood income may serve as an IV because DEBF varies with income level (21), and there is little reason to believe that income level directly causes childhood asthma (22). There is some evidence that non-Hispanic African American children from households with income significantly below Federal Poverty levels may be more likely to exhibit symptoms of asthma (23). In the sample of this study, none of the income levels by zip code is below the 2013 Federal Poverty levels and all the participants in this study had health insurance to visit Kaiser Permanente. Hence, it was determined that income level might serve as an appropriate IV. To provide some ex-post validation for the use of income as the IV for the sample of this study, the Pearson correlation coefficient between DEBF and income was 0.16, significant at one-tailed $p = 0.07$. To evaluate that the IV was not correlated with unadjusted confounders, the IV was correlated with the residuals from the second-stage analysis and found to be uncorrelated ($r = 0.000$).

In the second-stage, a Poisson generalized linear model (GLM) with a log-link function was estimated based on the predicted value of DEBF obtained from the first-stage. A GLM was estimated because the proportion of childhood asthma in the sample was high; using odds ratios would bias the estimate of relative risk (RR) (24). A Poisson log-link GLM with robust variance estimator was performed to address convergence problems encountered in a GLM binomial log-link function (25). Potential confounders included all covariates from the breastfeeding regression excluding the IV. Two additional confounders, hours of physical activity patients engaged in and exposure to pets in the first 12 months of life, were included to ensure a minimum set of appropriate confounders are included (2). Data were analysed using STATA 13.0 for Windows (StataCorp LP, College Station, Texas, USA).

Results

Breastfeeding, asthma and demographic characteristics

Ninety mothers, yielding a 17.3% response rate, voluntarily completed the questionnaires. Three questionnaires were excluded because the mothers did not answer all the questions. As stated earlier, the surveys were conducted during office visits between 1pm and 4pm, and consequently, most non-participating mothers excused themselves stating they did not have time as they had to return to work, pick up another child from school, do grocery shopping, or head home to prepare a meal. Since there is no data for non-participating mothers, it is difficult to adjust for the effect of non-response bias. To evaluate whether non-response bias is present when there is no pertinent data for non-responders, comparisons on the incidence of asthma and DEBF between “earlier” and “later” participants was conducted (26). The surveys were partitioned into two groups using the mid-point date between the start and completion of the survey administration dates. Statistical comparisons between the two groups for incidence of asthma ($\chi^2 = 1.765$, $p = 0.184$) and DEBF ($t = -0.0898$, $p = 0.928$) did not show significance.

Table 1 reports the frequencies of categorical variables. The majority of the patients were breastfed for up to 1 month and less than half were breastfed for at least 6 months, which is below the 6-month recommendation of the WHO, CDC, and AAP. Of the 87 patients, 31 were either never

breastfed or were given formula milk together with human milk since birth. Thirty-one or 35.6 percent of patients in this study were physician diagnosed with asthma. Other demographic characteristics are reported in Table 1.

Multivariate model

First, a Hausman test was performed to assess potential endogeneity between DEBF and asthma (15). If there were an endogenous relationship between DEBF and asthma, then a two-stage analysis would be called for. Otherwise, the estimation of a single equation would be more efficient. Results of the Hausman test indicated strong presence of endogeneity ($z = 3.12$, $p = 0.002$). Results of a single equation Poisson GLM log-link of asthma in Table 2 indicates DEBF was positive and significantly related to childhood asthma (relative risk ratio (RR) = 2.020; 95% Confidence Interval (CI), 1.143 - 3.570, $p = 0.016$).

The results of the two-stage analysis are shown in Table 3. The upper portion of the table shows the results of estimating the first-stage ordinary least squares (OLS) regression model explaining DEBF. To satisfy the OLS assumptions, the DEBF was transformed using the natural logarithm (12). The overall regression model was significant ($F = 3.44$, $p = 0.003$) with R^2 of 23% and an adjusted- R^2 of 17%. The explanatory power compares well to other studies (27). The biological mother's education level and older biological siblings diagnosed with allergy were the two most positive and significant ($p < 0.01$) predictors of DEBF. For every biological mother with a bachelor's degree and higher, compared to mothers without a bachelor's degree, the DEBF was greater by about 34%. Similarly, if an older biological sibling had allergy compared to siblings without allergy, the DEBF was greater by about 59%.

The lower part of Table 3 presents results for the second-stage Poisson GLM log-link estimating the risk of asthma. Independent variables included the exogenous variables from the DEBF model and two additional confounders, physical activity and exposure to pets in first 12 months of life. After adjusting for confounders, Table 3 shows the predicted value of DEBF was significantly negatively associated with childhood asthma. The RR of 0.003 (95% CI, 0.000-0.240, $p = 0.009$)

suggests a one-unit increase in DEBF or close to at least 3 months (2.72 months) since the DEBF measure was transformed using the natural logarithm, reduces the risk of childhood asthma by more than 99%. Of the confounders, biological mother's education level (RR = 11.214; CI, 2.384 - 52.741, $p = 0.002$) and older biological sibling physician diagnosed allergy (RR = 14.370; CI, 1.051 - 196.428, $p = 0.046$) were significantly associated with childhood asthma.

To maintain at least 10 observations per predictor, not all the demographic variables were included in the models (20). When additional demographics such as mother's age at birth of child, father diagnosed with asthma or allergy, daycare attendance in first 12 months, race (White, African American, Asian), child living with mother as single parent, and age of the child were included in the asthma model, none of them were significant while the predicted value of DEBF remained significantly negatively associated with childhood asthma (RR = 0.000; CI, 0.000 - 0.025, $p \leq 0.001$).

Since mothers' ability to recall breastfeeding behavior may reduce with the mother's age and as the child grows older (1-2), the first-stage regression was modified to adjust for child age greater than 7 and mothers' age. The new predicted DEBF was used in the second-stage asthma model. Additionally, the second-stage asthma model was modified to adjust for child age greater than 7 and mother's age because asthma diagnosis could differ by these factors (1-2), and atopic dermatitis; 11 patients had indication of atopic dermatitis. When these adjustments were made, the effect of DEBF remained negative and significantly associated with childhood asthma (RR = 0.008; CI, 0.000 - 0.333, $p = 0.011$). The age of the child and mother did not have a significant effect on either DEBF or asthma risk. However, atopic dermatitis was positive and significant in the asthma model (RR = 2.339; CI, 1.122 - 4.876, $p < 0.05$).

Discussion

This observational study illustrates that the presence of an endogenous relationship between breastfeeding and childhood asthma affects the conclusion whether DEBF is protective or not of childhood asthma. An undetected endogenous association could exist when researchers do not implement appropriate statistical analysis to address selection bias, unmeasured confounders, and

reverse causality, which have long been recognised as significant issues affecting research on breastfeeding and asthma (1-4). Failure to appropriately address the presence of an endogenous association will underestimate the effect of breastfeeding on childhood asthma, which is one of the reasons for the reported positive effect of breastfeeding on the risk of childhood asthma in the literature (1-2).

When breastfeeding and childhood asthma researchers attempt to address potential endogeneity, their two main design choices are to include potential confounders and perform interaction analysis (1-2). However, these two approaches are insufficient to statistically address endogeneity, especially when unadjusted confounders are unknown or unmeasurable (2-3, 10-11). An alternative approach is to use a two-stage IV analysis, which has its origins in path analysis and structural equation modelling (2, 10-12). The IV approach is well suited to breastfeeding and childhood asthma research because it provides a causal pathway whilst simultaneously adjusting for effects of unmeasured confounders, reverse causality, and selection bias (10-12).

The main strength of this study is that it illustrates how researchers can use the IV approach and adjust for endogeneity between DEBF and childhood asthma. This study first explains the concept of endogeneity including how it may arise in research on breastfeeding and childhood asthma. Second, it demonstrates that failing to statistically address endogeneity can affect the estimated effect of DEBF on childhood asthma. Third, it documents the statistical analysis and considerations for implementing the IV approach. Finally, the study compares the results between not using the IV approach to using the IV approach to address endogeneity in the association between DEBF and childhood asthma. This study has important implications because the results show that failure to consider and appropriately address the presence of endogeneity between breastfeeding and childhood asthma can lead to results indicating that breastfeeding is not protective from the risk of childhood asthma. However, when endogeneity was addressed statistically, the evidence indicates breastfeeding is protective against the risk of childhood asthma.

The study's main limitations stem from its convenience sample. First, the small sample size has a relatively higher proportion of children with asthma and most of the mothers are well educated, which may be due to the sampling criteria based on pediatric hospital visits. Because highly educated mothers utilize hospital services much more than less educated mothers (28), the diagnosis of childhood asthma is likely to be detected earlier and more asthma cases included in the sample. Future researchers could verify the findings here using a larger sample. Second, the small sample is not conducive to performing further IV analysis across a spectrum of different sub-cohorts such as by categories of asthma severity, asthma aetiology, immunological asthma manifestation, and age group. In particular, the child's age is an important factor as studies are known to differ across different age groups (1-2). However, when the age of the child (1-2), mother's age (1), and incidence of atopic dermatitis were included as additional confounders, the effect of DEBF on childhood asthma remained strong. There are other factors that can influence the incidence of childhood asthma and this study, like others, cannot control for all of these factors (1-2). Some of these factors include presence of humidity and mold in the home, rhinitis, allergic rhinoconjunctivitis (2, 4), and mode of human milk feeding (breast or expressing and bottle feeding) (22). Despite the presence of inherent heterogeneity in the sample and unmeasured confounders, this study provides strong evidence that DEBF is protective of childhood asthma risk. This study supports the robustness of the IV approach over standard multivariate outcome models in breastfeeding and childhood asthma research especially when there are unmeasured confounders and sample heterogeneity (2, 10-12). It should be recognised that although the IV approach utilizing a two-stage analysis is robust to unmeasured confounders, it does not completely eliminate such confounding.

Fourth, in common with retrospective studies on breastfeeding and childhood asthma, this study required mothers to recall information, which could affect the reliability of the survey responses. However, research shows that mothers are good at recalling duration of breastfeeding behavior even after 20 years (2), and to reduce recall bias, mothers were required to select one option from a set of pre-determined breastfeeding and formula feeding patterns (14-15). Adjusting for mothers' age in the analyses did not alter the findings. Fifth, the bias associated with mothers

declining to participate in this voluntary study could affect the results as the differences in the measured variables between responders and non-responders is unknown. Although non-response bias tests were performed and the IV statistical approach addresses selection bias, the presence of a bias of breastfeeding on the development of asthma cannot be completely ruled out.

The final limitation relates to the IV two-stage analysis to address endogeneity between breastfeeding and childhood asthma. It is very difficult to identify an IV and the results here could vary if a different IV is used, regardless of whether the IV is strong or weak. However, using a weaker IV that might be the best available at the time should not discourage researchers as the IV approach is robust and can provide important insights rather than not addressing endogeneity at all (11). Generic to the use of an IV is that it can never be proven that the IV is not correlated with unobserved covariates (10-12). However, this limitation should not discourage researchers from attempting to understand the complex causal pathways between breastfeeding and childhood asthma. Further research can benefit from addressing the limitations of this study. More specifically, research on the link between breastfeeding and childhood asthma can be enhanced through the application of an IV two-stage analysis so that the complex causal pathways can be clinically better understood for developing public health policies.

To conclude, this is one of the first studies to propose and test the possible presence of endogeneity in the relationship between duration of exclusive breastfeeding and childhood asthma so as to provide an explanation for the controversial results in the literature. After addressing for endogeneity using an IV approach, results showed duration of exclusive breastfeeding is protective against childhood asthma. In contrast, when endogeneity was not statistically addressed, the results indicated that duration of exclusive breastfeeding “causes” childhood asthma. The primary implication is that future researchers should consider and address potential endogeneity when analysing the relationship between breastfeeding and childhood asthma.

Acknowledgements: This study was supported by Kaiser Permanente in Kennesaw, Georgia, USA.

The participation of the consenting mothers and children for generously giving their time is greatly appreciated. The helpful comments and suggestions from the two anonymous reviewers are much appreciated.

References

1. Lodge CJ, Tan DJ, Lau MXZ, et al. Breastfeeding and asthma and allergies: a systematic review and meta-analysis. *Acta Paediatrica* 2015; 104(S467): 38-53.
2. Dogaru CM, Nyffenegger D, Pescatore AM, Spycher BD, Kuehni CE. Breastfeeding and childhood asthma: systematic review and meta-analysis. *Am J Epidemiol* 2014; 179: 1153-1167.
3. Kramer MS. Does breast feeding help protect against atopic disease? Biology, methodology, and a golden jubilee of controversy. *J Pediatr* 1988; 112: 181-190.
4. Brew BK, Allen CW, Toelle BG, Marks GB. Systematic review and meta-analysis investigating breast feeding and childhood wheezing illness. *Paediatr and Perinat Epidemiol* 2011; 25: 507-518.
5. Scott JA, Binns CW. Factors associated with the initiation and duration of breastfeeding: a review of the literature. *Breastfeed Rev* 1999; 7: 5-16.
6. Dennis CL. Breastfeeding initiation and duration: a 1990–2000 literature review. *JOGNN – Journal of Obstetric, Gynecologic, and Neonatal Nursing* 2002; 31: 12-32.
7. Amir LH, Donath SM. Does maternal smoking have a negative physiological effect on breastfeeding? The epidemiological evidence. *Birth* 2002; 29: 112-123.
8. Turcksin R, Bel S, Galjaard S, Devlieger R. Maternal obesity and breastfeeding intention, initiation, intensity and duration: a systematic review. *Matern Child Nutr* 2014; 10: 166-183.

9. Kusunoki T, Morimoto T, Nishikomori R, et al. Breastfeeding and the prevalence of allergic diseases in school children: does reverse causation matter? *Pediatr Allergy Immunol* 2010; 21: 60-66.
10. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29: 722-729.
11. Johnson KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statist Med* 2008; 27: 1539-56.
12. Greene WH. *Econometric Analysis*. New York: Prentice Hall 2008.
13. National Center for Injury Prevention and Control, CDC. 10 leading causes of death by age group, United States – 2012.
https://www.cdc.gov/injury/wisqars/pdf/leading_causes_of_death_by_age_group_2012-a.pdf. Accessed 21 November, 2013.
14. Silvers KM, Frampton CM, Wickens K, et al. Breastfeeding protects against current asthma up to 6 years of age. *J Pediatr* 2012; 160: 991-996.
15. Takemura Y, Sakurai Y, Honjo S, Kusakari A, Hara T, Gibo M, Kugai N. Relation between breastfeeding and the prevalence of asthma. *Am J Epidemiol* 2001; 154: 115-119.
16. Center for Disease Control and Prevention (CDC). (Graph and table representation of percentage of children who were breastfed). Breastfeeding among U.S. children born 2000-2010, CDC national immunization survey. 2013. http://www.cdc.gov/breastfeeding/data/NIS_data/index.htm. Accessed 21 November, 2013.
17. Black MH, Zhou H, Takayanagi M, Jacobsen SJ, Koebnick C. Increased asthma risk and asthma-related health care complications associated with childhood obesity. *Am J Epidemiol* 2013; 178: 1120-8.

18. Ownby DR, Johnson CC, Peterson EL. Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. *JAMA* 2002; 288: 963-972.
19. Williams B, Powell A, Hoskins G, Neville R. Exploring and explaining low participation in physical activity among children and young people with asthma: a review. *BMC Fam Pract* 2008; 9: 40 doi:10.1186/1471-2296-9-40.
20. Linares AM, Rayens MK, Dozier A, Wiggins A, Dignan MB. Factors influencing exclusive breastfeeding at 4 months postpartum in a sample of urban Hispanic mothers in Kentucky. *J Hum Lact* 2015; 31: 307-314.
21. Stuebe AM, Bonuck K. What predicts intent to breastfeed exclusively? Breastfeeding knowledge, attitudes, and beliefs in a diverse urban population. *Breastfeed Med* 2011; 6: 413-420.
22. Akinbami LJ, LaFleur BJ, Schoendorf KC. Racial and income disparities in childhood asthma in the United States. *Ambulatory Pediatrics* 2002; 2: 382-387.
23. Smith LA, Hatcher-Ross JL, Wertheimer R, Kahn RS. Rethinking race/ethnicity, income, and childhood asthma: racial/ethnic disparities concentrated among the very poor. *Public Health Reports* 2005; 120: 109-116.
24. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998; 280: 1690-1691.
25. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004; 159: 702-706.
26. de Winter, Andrea F, et al. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. *Eur J Epidemiol* 2005; 20: 173-181.
27. Merewood A, Patel B, Newton KN, et al. Breastfeeding duration rates and factors affecting continued breastfeeding among infants born at an inner-city US baby-friendly hospital. *J Hum Lact* 2007; 23: 157-64.

28. Flores, G, Bauchner H, Feinstein AR, Nguyen UDT. The impact of ethnicity, family income, and parental education on children's health and use of health services. *Am J Public Health* 1999; 89: 1066-1071.

Table 1: Demographic Characteristics

Panel A: Demographic Characteristics for Categorical Variables^a	
Variable	Total n (%)
Asthma	
Yes	31 (35.6)
No	56 (64.4)
Duration breastfeeding	
Never breastfed	10 (11.5)
< 1 month	18 (20.7)
1-2 months	13 (14.9)
2-3 months	8 (9.2)
3-6 months	10 (11.5)
6-9 months	9 (10.3)
9-12 months	8 (9.2)
> 12 months	11 (12.6)
Introduction of formula milk	
< 1 month	31 (35.6)
1-2 months	11 (12.6)
2-3 months	10 (11.5)
3-6 months	9 (10.3)
6-9 months	10 (11.5)
9-12 months	3 (3.4)

Panel A: Demographic Characteristics for Categorical Variables^a	
Variable	Total n (%)
> 12 months	1 (1.1)
Never formula milk	12 (13.8)
Sex	
Male	38 (43.7)
Female	49 (56.3)
Race	
White	56 (64.4)
African American	15 (17.2)
Hispanic	5 (5.7)
Asian	4 (4.6)
Other	7 (8.0)
Premature	
Yes	13 (14.9)
No	74 (85.1)
Mother asthma	
Yes	8 (9.2)
No	77 (90.8)
Older sibling asthma	
Yes	13 (14.9)
No	74 (85.1)
Mother allergy	
Yes	38 (43.7)
No	49 (56.3)
Older sibling allergy	
Yes	11 (12.6)
No	76 (87.4)

Panel A: Demographic Characteristics for Categorical Variables ^a		
Variable	Total n (%)	
Mother’s highest education		
Grade 1 to 8	0 (0)	
High school	10 (11.5)	
Some college	38 (43.7)	
Bachelor’s degree	31 (35.6)	
Master’s degree and above	8 (9.2)	
Pet exposure first 12 months		
Yes	44 (50.6)	
No	43 (49.4)	
Day care attendance first 12 months		
Yes	24 (27.6)	
No	63 (72.4)	
Number of children		
1	12 (13.79)	
2	42 (48.28)	
3	21 (24.14)	
4	8 (9.20)	
6	4 (4.60)	
Number of siblings		
0	12 (13.79)	
1	42 (48.28)	
2	21 (24.14)	
3	8 (9.20)	
5	4 (4.60)	
Panel B: Demographic Characteristics for Continuous Variables		
Characteristic	Mean	SD

Panel A: Demographic Characteristics for Categorical Variables^a

Variable	Total n (%)	
Age (years)	8.55	4.15
Mother's age at child's birth	28.92	6.18
Family Income (\$) ^b	53,707	9,509
Physical activity (hours)	2.05	0.71

^aThe total percentage for each category may not add up to 100% due to rounding.

^bFamily income is average neighborhood income based on zip code.

Table 2: Model for Risk of Asthma Prior to Accounting for Endogeneity in Pediatric Patients^a

Independent variables	Risk Ratio		95% CI
DEBF	2.020*	1.143	3.570
Mother's education (bachelor's degree or higher = 1, otherwise 0)	1.361	0.742	2.492
Biomother asthma (yes = 1, no = 0)	2.686**	1.522	4.743
Biomother allergy (yes = 1, no = 0)	1.071	0.620	1.850
Older biosibling asthma (yes = 1, no = 0)	2.221*	1.087	4.538
Older biosibling allergy (yes = 1, no = 0)	0.333*	0.136	0.814
Premature (yes = 1, no = 0)	1.768*	0.997	3.134
Physical activity (hours per day)	1.044	0.677	1.610
Pets 12 months (yes = 1, no = 0)	0.666	0.357	1.242
Constant	0.098	0.020	0.478

^aTable reports GLM Poisson log-link function to estimate the relative risk of asthma with (1/df) Deviance = 0.686, AIC = 1.538, and BIC = -292.080. Confidence intervals are based on patient-clustered robust standard errors. DEBF = duration of exclusive breastfeeding.

*p < 0.05

**p < 0.01

Table 3: Multivariate Two-Stage IV Model for Risk of Asthma in Pediatric Patients^a*First-Stage (OLS Regression) Dependent Variable: Duration of Exclusive Breastfeeding*

Independent variables	Coefficient	95% CI	
Mother's education (bachelor's degree or higher = 1, otherwise 0)	0.340**	0.122	0.558
Biomother asthma (yes = 1, no = 0)	-0.273	-0.670	0.123
Biomother allergy (yes = 1, no = 0)	-0.012	-0.232	0.209
Older biosibling asthma (yes = 1, no = 0)	-0.247	-0.583	0.090
Older biosibling allergy (yes = 1, no = 0)	0.592**	0.264	0.920
Premature (yes = 1, no = 0)	-0.094	-0.384	0.196
Income (natural logarithm of median zip code income)	0.371	-0.240	0.983
Constant	-2.910	-9.498	3.678

Second-Stage (GLM Poisson Log-Link) Dependent Variable: Asthma

Independent variables	Risk Ratio	95% CI	
Predicted DEBF	0.003**	0.000	0.240
Mother's education (bachelor's degree or higher = 1, otherwise 0)	11.214**	2.384	52.741
Biomother asthma (yes = 1, no = 0)	0.538	0.166	1.737
Biomother allergy (yes = 1, no = 0)	1.242	0.713	2.166
Older biosibling asthma (yes = 1, no = 0)	0.457	0.162	1.292
Older biosibling allergy (yes = 1, no = 0)	14.370*	1.051	196.428
Premature (yes = 1, no = 0)	0.940	0.448	1.971
Physical activity (hours per day)	1.017	0.659	1.569
Pets 12 months (yes = 1, no = 0)	0.622	0.334	1.156
Constant	128.038	1.183	13863.49

^aThe first stage regression corresponds to the OLS model in which the dependent variable is the natural logarithm of the duration of exclusive breastfeeding, where duration of breastfeeding was measured in months using an eight category scale. The adjusted R^2 (R^2) of first-stage model = 0.17 (0.23), $F = 3.44$, $p = 0.003$. The second stage regression corresponds to the GLM Poisson log-link function to estimate the relative risk of asthma with (1/df)

Deviance = 0.630, AIC = 1.530, and BIC = -292.801. Confidence intervals are based on patient-clustered robust standard errors. DEBF = duration of exclusive breastfeeding.

* $p < 0.05$

** $p < 0.01$

Appendix A: Survey Instrument

Thank you for taking your time to complete this survey to help us better understand and provide treatment for your child/children. Please answer all questions below by checking the bubble, filling in the blanks, and circling Y or N (Yes or No) as appropriate. Please think back to the period surrounding the child's birth.

When your child was born what was the biological:

Mother's age? _____ Father's age? _____

What is the child's age? _____

Was the child premature? Y N If yes, how many weeks? _____

Did the biological mother have: Asthma? Y N Allergies? Y N

Did the biological father have: Asthma? Y N Allergies? Y N

Did any biological older siblings have: Asthma? Y N Allergies? Y N

Did the paternal grandmother have: Asthma? Y N Allergies? Y N

Did the paternal grandfather have: Asthma? Y N Allergies? Y N

Did the maternal grandmother have: Asthma? Y N Allergies? Y N

Did the maternal grandfather have: Asthma? Y N Allergies? Y N

How long did the mother breastfeed the child?

- ☐ Less than 1 month
 ☐ 1 to 2 months
 ☐ 2 to 3 months
☐ 3 to 6 months
 ☐ 6 to 9 months
 ☐ 9 to 12 months
☐ More than 12 months
 ☐ Never

When did the child begin formula milk?

- ☐ Less than 1 month ☐ 1 to 2 months ☐ 2 to 3 months
☐ 3 to 6 months ☐ 6 to 9 months ☐ 9 to 12 months
☐ More than 12 months ☐ Never

On average, how long does the child engage in physical activity (swimming, soccer, basketball, football, running, etc.) daily?

- ☐ Less than 1 hour daily ☐ 1-2 hours daily ☐ 3-4 hours daily
☐ More than 4 hours daily

Did the child grow up with pets (cat, dog, etc.) in the first 12 months? Y N

Did the child grow up with pets after the first 12 months? Y N

Did the child attend a day care in the first 12 months of life? Y N

Is the child regularly exposed to smoke? Y N

Does the father smoke? Y N

Does the mother smoke? Y N

Did the biological mother smoke while she was pregnant with the child? Y N

What is mother's highest level of education?

- ☐ Primary school ☐ High School ☐ Some college
☐ Bachelor's degree ☐ Master's degree ☐ Professional Degree

What is father's highest level of education?

- ☐ Primary school ☐ High School ☐ Some college
☐ Bachelor's degree ☐ Master's degree ☐ Professional Degree

How many biological children do you have?

- ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

Thank you for participating in this survey. Your responses will remain anonymous and confidential.