SUPPLEMENTARY APPENDIX

Antibiotic treatments during infancy, changes in nasal microbiota, and asthma development: Population-based cohort study

Laura Toivonen, MD, PhD; Linnea Schuez-Havupalo, MD, PhD; Sinikka Karppinen, MD, PhD; Matti Waris, PhD; Kristi L. Hoffman, PhD; Carlos A. Camargo, Jr., MD, DrPH; Kohei Hasegawa, MD, MPH*; and Ville Peltola, MD, PhD*

*Equally contributed

Supplementary Materials

months in 697 children enrolled in the STEPS Study

Supplementary Table 6. Richness, alpha-diversity, and relative abundance by longitudinal nasal microbiota profile during age 2-24 months in 697 children enrolled in the STEPS Study

Supplementary Table 7. Association of antibiotic treatments during age 0-11 months with two longitudinal nasal microbiota profiles during age 2-24 months (n=697)

Supplementary Table 8. Antibiotic treatments during age 0-11 months and asthma at age 7 years (n=697)

Supplementary table 9. Sensitivity analysis for mediation analysis, using a cut-off of 0-2 vs. ≥3 antibiotic treatments during age 0-11 months as the exposure (n=623)

Supplementary Table 10. Sensitivity analysis for mediation analysis, with antibiotic treatments during age 0-2 months as the exposure (n=623)

Supplementary Table 11. Sensitivity analysis for mediation analysis, with the longitudinal nasal microbiota profiles being dichotomized into the profile with the lowest *Moraxella* abundance vs. other profiles (n=623)

SUPPLEMENTARY METHODS

Study Design, Setting, and Participants

In the prospective, population-based birth-cohort study—the Steps to the Healthy Development and Well-being of Children (STEPS Study), families of Finnish children are followed until early adulthood [1]. From all children born in the Hospital District of Southwest Finland from January 2008 through April 2010 to Finnish or Swedish-speaking mothers (eligible cohort—9811 mothers and 9936 infants), families of 1827 infants (30 pairs of twins) were recruited either during the first trimester of pregnancy or soon after birth. An intensive follow-up of acute respiratory infections (ARIs) from birth to age 24 months was offered to these families, and 923 children were enrolled [2, 3]. The children were followed for development of asthma until age 7.5 years [4]. No selection criteria other than language (Finnish or Swedish speaking family) were applied to recruiting the families in the STEPS Study or in the subcohort. All data were reviewed at the Turku Centre for Child and Youth Research. The Ministry of Social Affairs and Health and the Ethics Committee of the Hospital District of Southwest Finland approved the study. Parents of participating children gave their written, informed consent. The study complies with the Declaration of Helsinki.

Based on data from the Finnish National Birth Registry [1], the participating and nonparticipating children were similar in the baseline characteristics, such as sex, gestational age, birth weight, 5-minute Apgar-points, and maternal BMI (all $P>0.10$) while the participating children were more often first-borns.

Exposure

The primary exposure was exposure to systemic antibiotic use for any indication, including ARIs and non-ARIs, during infancy (age 0-11 months). Antibiotic treatments were classified in therapeutic classes as previously described by Poole and colleagues [5]. Narrow-spectrum penicillins (amoxicillin, phenoxymethylpenicillin, benzylpenicillin, and ampicillin), first-generation cephalosporins, sulfonamides, and nitrofurantoin were considered narrow-spectrum antibiotics [5]. All other antibiotics, including combination of β-lactam and β-lactamase inhibitors (e.g., amoxicillin-clavulanate) and macrolides, were considered broad-spectrum. Data of antibiotic use was captured through multiple sources. Parents were instructed to record all respiratory and other symptoms as well as physician visits with antibiotic treatments into a daily diary during age 0-11 months. Families were also instructed to visit the study clinic during ARIs at the Turku Centre for Child and Youth Research, Turku University Hospital and University of Turku (Turku, Finland), and children were examined by a study physician using a structured form. Data on emergency department visits, outpatient visits, and hospitalizations during age 0-11 months with antibiotic treatments were retrieved from medical and prescription records of the Hospital District of Southwest Finland [4].

Of 923 children in the STEPS respiratory cohort, 886 (96%) children had data of antibiotic treatments during age 0-11 months. Overall, 468 antibiotic treatments were identified through daily diaries and 739 antibiotic treatments (either new prescription or on-going use of antibiotics) were identified through physician visits (204 study clinic visits, 381 other outpatient clinic visits, and 154 emergency department visits or hospitalizations). With combining these data and filtering out duplicated treatments (e.g., multiple visits during the same antibiotic

treatment), a total of 754 antibiotic treatments were identified in 697 children during age 0-11 months.

Mediator

The mediator of interest was longitudinal changes in nasal airway microbiota during age 0-24 months. Using a standardized protocol [2, 3], nasal swab specimens were collected by study personnel using flocked nylon swabs (Copan, Brescia, Italy) at a scheduled participant visit at age 2, 13, and 24 months during healthy state.

16S rRNA Gene Sequencing of Nasal Airway Microbiota

The nasal swab samples were stored at -80°C after the collection. Swabs were suspended in phosphate buffered saline and tested by using 16S rRNA gene sequencing. 16S rRNA gene sequencing methods were adapted from the methods developed for the National Institutes of Health (NIH) Human Microbiome Project [6, 7]. Nasal swab samples were eluted in 500 µl of 1 x PBS by vortexing. An aliquot of 200 µl was used as a starting material for bacterial DNA extraction. The DNAs were isolated from nasal swab samples with automated MagNA Pure 96 System using MagNA Pure 96 DNA and Viral NA SV 2.0 kit (Cat. No 6543588001, Roche Diagnostics, Mannheim, Germany) with Pathogen Universal 200 3.1 protocol and an elution volume of 50 µl. ZymoBiomics Microb Community standard was used as a positive control (Cat. No. D6300, Zymo Research). DNA extractions were performed at Turku Centre for Biotechnology (Turku, Finland) and extracted DNAs were sent to Baylor College of Medicine (Houston, TX, USA) for microbiota testing.

The 16S rDNA V4 region was amplified by PCR and sequenced on the MiSeq platform (Illumina; SanDiego, CA) using the 2x250 bp paired-end protocol yielding pair-end reads that overlap almost completely. The primers used for amplification contain adapters for MiSeq sequencing and single-end barcodes allowing pooling and direct sequencing of PCR products [8, 9]. Sequencing read pairs were demultiplexed based on the unique molecular barcodes, and reads were merged using USEARCH v7.0.1090 [10], allowing zero mismatches and a minimum overlap of 50 bases. Samples with suboptimal amounts of sequencing reads were re-sequenced to ensure that the majority of bacterial taxa were encompassed in our analyses. 16S rRNA gene sequences were clustered into operational taxonomic units (OTUs) at a similarity cutoff value of 97% using the UPARSE algorithm [11]. The use of 97% cutoff value has been a widely used cut-off in the microbiota literature [7, 12, 13] because it offers a compromise between the potential inflation of the number of OTUs due to sequencing errors and the cutoff used for taxonomic classification. OTUs were determined by mapping the centroids to the SILVA database [14] version 128 containing only the 16S V4 region to determine taxonomies. Rarefaction curves of bacterial OTUs were constructed using sequence data for each sample to ensure coverage of the bacterial diversity present. A custom script constructed a rarefied OTU table (rarefaction was performed at only one sequence depth) from the output files generated in the previous two steps for downstream analyses. Analyses were conducted at the genus level using bacterial relative abundances. For clustering, relative abundances of zero were imputed with 1 / rarefaction cutoff and relative abundance data were $\log_2$-transformed.

Quality control

The processes involving microbial DNA extraction, 16S rRNA gene amplification, and amplicon sequencing included a set of controls that enabled us to evaluate the potential

introduction of contamination or off-target amplification. Nontemplate controls (extraction chemistries) were included in the microbial DNA extraction process and the resulting material was subsequently used for PCR amplification. In addition, at the step of amplification, another set of nontemplate controls (PCR-mix) was included to evaluate the potential introduction of contamination at this step. Similarly, a positive control composed of known and previously characterized microbial DNA was included at this step to evaluate the efficiency of the amplification process. Before samples (unknowns) were pooled together, sequencing controls were evaluated and the rejection criteria were the presence of amplicons in any of the nontemplate controls or the absence of amplicons in the positive control. In the present study, no amplicons were observed in the nontemplate controls and a negligible amount of raw reads was recovered after sequencing. A total of 46,441,397 high-quality merged sequences were obtained by 16S rRNA gene sequencing of the nasal airway samples, of which 45,854,654 (99%) were mapped to 16S reference data.

A total of 2,261 nasal swab samples were collected at age 2, 13, and 24 months, and 2,172 (96%) met the quality control requirements and had sufficient sequence depth for 16S rRNA gene sequencing; 89 (4%) samples did not meet the quality control requirements and 160 (7%) nasal samples were excluded because of missing follow-up or baseline samples. Longitudinally collected qualified samples and data on antibiotic use were available for 697 children who comprised the analytical cohort, with 1,923 nasal samples collected.

Longitudinal Clustering of Nasal Microbiota

To minimize model misspecification in the causal mediation analysis due to the high-dimensionality of microbiota data over time, the microbiota data were summarized into a

summarized variable (or longitudinal profiles). To identify profiles of longitudinal changes in the nasal microbiota during age 2-24 months, we applied an unsupervised clustering (longitudinal k-means clustering) approach [15] based on correlation distance [16] to the individual longitudinal trajectories based on $\log_2$-transformed relative abundances of the 100 most common genera which accounted for 99% of overall abundance. We used correlation-based distance—which was computed between the observed longitudinal patterns for each pair of observations (rather than between variables)—as the dissimilarity measure because it focuses on the shapes of longitudinal patterns rather than the abundance of individual bacteria at each time point [16]. This unsupervised clustering approach has advantages, such as effectively summarizing high dimension data, taking the characteristics of microbiota as dynamic ecology into account, and improving interpretability. We chose the number of profiles based on Calinski-Harabasz methods [15, 17]. To complement this approach, we also utilized *a priori* knowledge of the nasal microbiota during early childhood. Indeed, these derived microbiota clusters are biologically plausible because the four profiles (profiles A-D) are characterized by major airway bacteria: *Moraxella* (profile A); *Streptococcus* and *Moraxella* (profile B); *Dolosigranulum, Corynebacteriaceae* and *Staphylococcus* (profile C)*;* as well as *Haemophilus* and *Streptococcus* along with *Moraxella* sparsity (profile D); in addition to the fifth profile that is characterized by mixed pattern (profile E; **Figure 3a**). These profiles are consistent with earlier studies [3, 13, 18-23].

Outcome

The primary outcome was physician-diagnosed asthma defined as a diagnosis of asthma in the medical records at age 6.5-7.5 years (age 7 years) with or without a prescription of inhaled

corticosteroids for asthma at age 6.5-7.5 years. Physician-diagnosis of asthma was retrieved from the medical records using *ICD-10-CM* codes J45-J46 (**Supplementary Table 1**) or, if the code was missing, written physician-diagnosis of "asthma", and asthma medication use from nation-wide electronic prescription records. All asthma diagnoses and prescriptions were made by attending physicians. The electronic prescription was introduced in Finland in 2010, and all public health care providers had taken up its use by 2013 and private health care providers by 2015. All pharmacies have been able to deliver electronic prescriptions since 2011. Electronic prescription became the main form of prescription in the beginning of 2017, and paper or phone prescriptions have been allowed only in exceptional situations. Medical records and electronic prescription data were available for 910 (99%) of the cohort children.

Potential Confounders

Patient demographics, family history, pre-, peri-, and post-natal history, and environmental information (e.g., parental history of asthma, household siblings, and breastfeeding) were collected from the National Birth Registry and by structured questionnaires during the first trimester of pregnancy, at the time of birth, and at child's age 13 and 24 months, and with the diary. Children's sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and ARIs during infancy (age 0-11 months) were considered potential confounders (**Figure 1**). An ARI was defined as presence of rhinitis or cough (with or without fever or wheezing) documented in the symptom diary by the parents, or as any physician-diagnosed ARI [2]. The duration of 97% of ARIs was ≤30 days. To account for sequential infections, the length of an ARI was limited to 30 days; longer ARIs (3%) were considered as

separate episodes with a maximum duration of 30 days. If the symptom data were missing, repeated diagnoses of ARIs within 14 days were considered as one episode.

Statistical Analysis

The patient characteristics were compared by antibiotic exposure during age 0-11 months in **Table 1**. Relative abundances of most abundant genera at each sampling age were compared between the mediator groups (two longitudinal microbiota profiles) by using Welch's unequal variances t-test, adjusting for multiple comparisons with the use of the Benjamini-Hochberg false discovery rate (FDR) method [24].

A directed acyclic graph (DAG; **Figure 1**) was constructed to represent our proposed model linking the exposure (antibiotic exposures during age 0-11 months) to the outcome (asthma at age 7 years) with the mediator (longitudinal microbiota profiles during age 2-24 months) and potential confounders (sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and frequency of ARIs during age 0-11 months). The model was constructed based on clinical plausibility and *a priori* knowledge [21, 25-29]. Next, to examine the association between the frequency of antibiotic treatments and the derived longitudinal microbiome profiles, multinomial regression models adjusting for the confounders were constructed.

To examine the direct and indirect effects (i.e., estimands) in a counterfactual framework, the causal mediation analysis was performed [30-33]. This method enables us to examine the extent to which the effect of exposure on the outcome is direct (direct effect) and to what extent it is mediated by the mediator (indirect effect). More specifically, the natural direct effect represents how much asthma risk would change on average if patient were set to be exposed

versus to be unexposed but for each individual the longitudinal microbiota pattern were kept at the level it would have taken in the absence of the exposure [30-33]. The natural indirect effect represents how much asthma risk change if patient were set to be exposed, but the longitudinal microbiota pattern were changed from the level it would take if unexposed to the level it would take if exposed [30-33]. In the causal mediation analysis, the number of antibiotic exposures during age 0-11 months was dichotomized based on the empirical distribution of exposure—0-1 antibiotic treatment and ≥2 antibiotic treatments (the highest quartile), which also addresses the effect of multiple antibiotic exposures [25, 27, 28, 34]. Additionally, to improve the interpretability of inference, the longitudinal microbiota profiles were further consolidated into the profile with the highest *Moraxella* abundance (low-risk profile [with regard to asthma risk]) vs. other profiles (high-risk profile [with regard to asthma risk, **Supplementary Table 2**]). Stratification by *Moraxella* abundance was chosen based on its dominance of the nasal microbiota and the literature reporting the relations of *Moraxella* with ARIs, wheezing, and asthma [3, 13, 21, 23, 35]. In the mediation models, the data on exposure and mediator were available for all children in the analytic cohort, while part of covariate data were missing in 68 children and asthma outcome in 6 children, leaving 623 children.

The detailed notations of variables and definitions of estimated effects in the causal mediation analysis are following:

A: Exposure of interest (i.e., exposures to systemic antibiotic treatments during 0-11 months) for each individual.

M: Mediator (i.e., longitudinal patterns of the nasal airway microbiota during age 2-24 months) for each individual.

Y: Outcome of interest (i.e., asthma status at age 7 years) for each individual

$\boldsymbol{C}$: A set of covariates (sex [binary], parental history of asthma [binary], household siblings [binary], breastfeeding during age 0-2 months [binary], and acute respiratory infections [frequency] during age 0-11 months in the primary analysis) for each individual

Y_a: Counterfactual outcome Y for each individual when intervening to set A to a

Y_{am}: Counterfactual outcome Y for each individual when intervening to set A to a and M to m

M_a: Counterfactual mediator M for each individual when intervening to set A to a

Total effect: The total average effect comparing treatment level $A = 1$ to $A = 0$

$$TE = E[Y_1 - Y_0|\boldsymbol{C}]$$

Natural direct effect: The average natural direct effect comparing treatment level $A = 1$ to $A = 0$, with setting $M = M_0$

$$NDE = E[Y_{1M_0} - Y_{0M_0}|\boldsymbol{C}]$$

Natural indirect effect: The average natural indirect effect comparing the effect of $M = M_1$ vs. $M = M_0$, with setting $A = 1$

$$NIE = E[Y_{1M_1} - Y_{1M_0}|\boldsymbol{C}]$$

Controlled direct effects: The average controlled direct effect comparing treatment level $A = 1$ to $A = 0$ with setting $M = m$. Of note, in the absence of exposure-mediator interactions, the controlled direct effects coincide with the natural direct effects.

$$CDE(m) = E[Y_{1m} - Y_{0m}|\boldsymbol{C}]$$

Proportion mediated: $PM = \frac{NIE}{TE}$

To account for confounding, we used inverse probability weighting for marginal structural models [36, 37]. First, we estimated the individual-specific inverse probability weight by constructing a logistic regression model adjusting for the potential confounders (sex, parental

history of asthma, household siblings, breastfeeding, and acute respiratory infections), according to the assumed causal structure (**Figure 1**). Next, we constructed outcome and mediator logistic regression models to the pseudo-population which was simulated by inverse probability weighting—that is, fitting weighted outcome and mediator regression models in order to estimate the parameters of interest ($\theta_1, \theta_2, \beta_1$) in the following marginal structural models:

$$logit\{P(Y = 1|a, m\} = \theta_0 + \theta_1 + \theta_2 m$$

$$logit\{P(M = 1|a)\} = \beta_0 + \beta_1 a$$

Then, these parameter estimates were used to estimate the estimands of the analysis—the average natural direct and indirect effects—with the use of the R *mediation* package [38].

Identifiability Assumptions of Causal Mediation Analysis

There are four identifiability assumptions in causal mediation [30]: 1) no unmeasured exposure-outcome confounders given measured confounders, 2) no unmeasured mediator-outcome confounders given both the measured confounders and exposure, 3) no unmeasured exposure-mediator confounders given the measured confounders, and 4) no mediator-outcome confounders affected by the exposure. It is plausible to assume that the first and third assumptions hold, by accounting for measured sex, household siblings, breastfeeding, and frequency of acute respiratory infections) in the analysis. However, the second assumption might not have hold. For example, the child's genetics may have served as an unmeasured confounder. Yet, this potential confounding has been mitigated, at least partially, by controlling the parental history of asthma because it is strongly correlated to the asthma-risk genetics of the parents thereby being correlated to child's genetics. The fourth assumption is difficult to verify. For example, a potential unmeasured mediator-outcome confounder affected by the exposure is

intestinal microbiota. Yet, within the sparse literature, it remains unclear how much the intestinal microbiota affects the airway microbiota in young children.

Sensitivity Analysis

In sensitivity analyses, the analysis was repeated with 1) any use of broad-spectrum antibiotics during age 0-11 months as the exposure, 2) use of a different cut-off for antibiotic exposure (0-2 vs. ≥3), 3) restriction of antibiotic use to age 0-2 months, and 4) use of a different mediator categorization (the lowest *Moraxella* abundance [profile D] vs. other profiles). This categorization was chosen because antibiotic exposures were associated with a higher probability of having a profile D (**Supplementary Tables 4** and **5**) and children with a profile D had the highest risk of developing asthma (**Supplementary Table 2**). Two-tailed P-values were reported, with $P<0.05$ considered statistically significant. Data were analyzed using R version 3.6.1.

SUPPLEMENTARY REFERENCES

1. Lagstrom H, Rautava P, Kaljonen A, et al. Cohort profile: Steps to the healthy development and well-being of children (the STEPS Study). Int J Epidemiol **2013**; 42: 1273-84.
2. Toivonen L, Schuez-Havupalo L, Karppinen S, et al. Rhinovirus infections in the first 2 years of life. Pediatrics **2016**; 138: e20161309.
3. Toivonen L, Hasegawa K, Waris M, et al. Early nasal microbiota and acute respiratory infections during the first years of life. Thorax **2019**; 74: 592-9.
4. Toivonen L, Forsstrom V, Waris M, Peltola V. Acute respiratory infections in early childhood and risk of asthma at 7 years of age. J Allergy Clin Immunol **2019**; 143: 407-10.e6.
5. Poole NM, Shapiro DJ, Fleming-Dutra KE, Hicks LA, Hersh AL, Kronman MP. Antibiotic prescribing for children in United States emergency departments: 2009-2014. Pediatrics **2019**; 143.
6. A framework for human microbiome research. Nature **2012**; 486: 215-21.
7. Structure, function and diversity of the healthy human microbiome. Nature **2012**; 486: 207-14.
8. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **2010**; 7: 335-6.
9. Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J **2012**; 6: 1621-4.
10. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **2010**; 26: 2460-1.

11. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods **2013**; 10: 996-8.

12. Stewart CJ, Ajami NJ, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature **2018**; 562: 583-8.

13. Man WH, van Houten MA, Mérelle ME, et al. Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study. Lancet Respir Med **2019**; 7: 417-26.

14. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **2013**; 41: D590-6.

15. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R packages to cluster longitudinal data. J Stat Softw **2015**; 65.

16. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer, **2017**.

17. Calinski T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics Theory and Methods **1974**; 3: 1-27.

18. Biesbroek G, Tsivtsivadze E, Sanders EA, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. Am J Respir Crit Care Med **2014**; 190: 1283-92.

19. Zhou Y, Jackson D, Bacharier LB, et al. The upper-airway microbiota and loss of asthma control among asthmatic children. Nat Commun **2019**; 10: 5714.

20. Bosch AA, de Steenhuijsen Piters WA, van Houten MA, et al. Maturation of the infant respiratory microbiota, environmental drivers and health consequences: a prospective cohort study. Am J Respir Crit Care Med **2017**; 196: 1582-90.

21. Teo SM, Mok D, Pham K, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe **2015**; 17: 704-15.

22. Hasegawa K, Mansbach JM, Ajami NJ, et al. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. Eur Respir J **2016**; 48: 1329-39.

23. Teo SM, Tang HHF, Mok D, et al. Airway microbiota dynamics uncover a critical window for interplay of pathogenic bacteria and allergy in childhood respiratory disease. Cell Host Microbe **2018**; 24: 341-52.e5.

24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc (Methodological) **1995**; 57: 289-300.

25. Donovan BM, Abreo A, Ding T, et al. Dose, timing, and type of infant antibiotic use and the risk of childhood asthma. Clin Infect Dis **2019**; May 31.

26. Abreo A, Gebretsadik T, Stone CA, Hartert TV. The impact of modifiable risk factor reduction on childhood asthma development. Clin Transl Med **2018**; 7: 15.

27. Metsala J, Lundqvist A, Virta LJ, Kaila M, Gissler M, Virtanen SM. Prenatal and post-natal exposure to antibiotics and risk of asthma in childhood. Clin Exp Allergy **2015**; 45: 137-45.

28. Marra F, Marra CA, Richardson K, et al. Antibiotic use in children is associated with increased risk of asthma. Pediatrics **2009**; 123: 1003-10.

29. Wypych TP, Wickramasinghe LC, Marsland BJ. The influence of the microbiome on respiratory health. Nat Immunol **2019**; 20: 1279-90.

30. VanderWeele T. Explanation in causal inference: methods for mediation and interaction: Oxford University Press, **2015**.

31. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology **1992**; 3: 143-55.

32. Pearl J. Direct and indirect effects. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence. San Francisco, CA, United States: Morgan Kaufmann Publishers Inc, **2001**:411-20.

33. Hernán M, Robins J. Causal inference: What if. Boca Raton: Chapman & Hall/CRC. **2020**.

34. Hoskin-Parr L, Teyhan A, Blocker A, Henderson AJ. Antibiotic exposure in the first two years of life and development of asthma and other allergic diseases by 7.5 yr: a dose-dependent relationship. Pediatr Allergy Immunol **2013**; 24: 762-71.

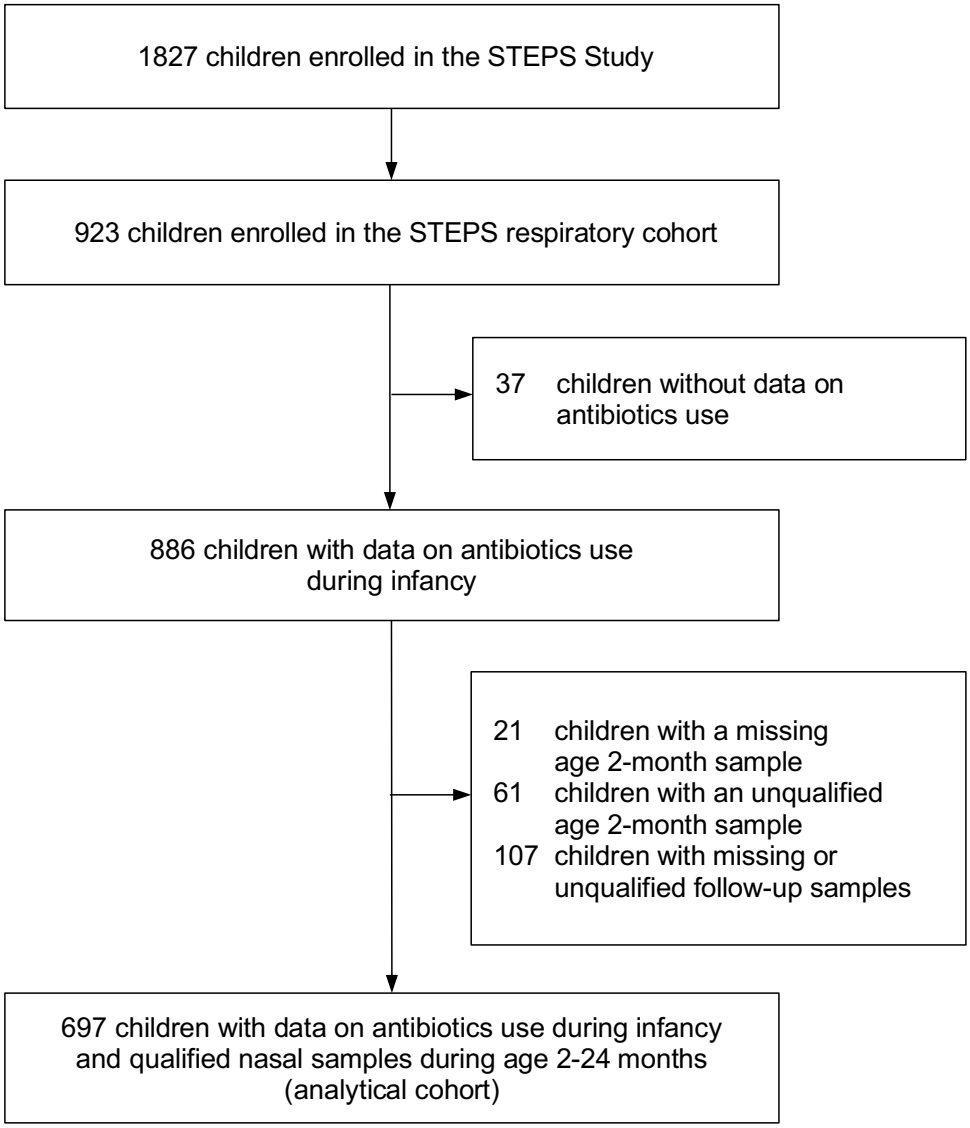
35. Bisgaard H, Hermansen MN, Buchvald F, et al. Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med **2007**; 357: 1487-95.

36. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology **2000**; 11: 550-60.

37. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology **2000**; 11: 561-70.

38. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. J Stat Softw **2014**; **59**: 1-38.

Supplementary Figure 1. Enrolment and follow-up of children in the STEPS Study



Supplementary Table 1. *ICD-10-CM* codes used for retrieving physician-diagnosis of asthma from the medical records[a]

***ICD-10-CM* code**
J45.0 *Asthma praecipue allergicum*
J45.1 *Asthma non allergicum*
J45.8 *Asthma mixtum*
J45.9 *Asthma non specificatum*
J46 *Status asthmaticus*

[a] Physician-diagnosis of asthma was retrieved from the medical records using *ICD-10-CM* codes J45-J46 or, if the code was missing, written physician-diagnosis of "asthma". Asthma medication use was retrieved from nation-wide electronic prescription records.

Supplementary Table 2. Longitudinal nasal microbiota profiles during age 2-24 months and asthma at age 7 years (n=697)

Longitudinal nasal microbiota profiles[a]	**Children with asthma,** n (%)[b]
Profile A with persistent *Moraxella* dominance (n=279)	18 (6.5%)
Profile B with *Streptococcus*-to-*Moraxella* transition (n=84)	5 (6.1%)
Profile C with early *Dolosigranulum/Corynebacteriaceae* dominance (n=139)	11 (7.8%)
Profile D with early *Moraxella* sparsity with its subsequent increase (n=100)	15 (15.2%)
Profile E with mixed longitudinal patterns (n=92)	6 (6.6%)

[a] Longitudinal clustering of nasal microbiota during age 2-24 months identified 6 distinct profiles. Of these, the profile F included only 3 children and is not shown in the table. For the mediation analysis, longitudinal nasal microbiota profiles were dichotomized to 1) profile with persistent *Moraxella* dominance (profile A) and 2) profile with early *Moraxella* sparsity (profiles B-F, **Figure 3b**).
[b] Medical records data for asthma were missing from 6 children.

Supplementary Table 3. Comparison of children between the analytical and non-analytical cohorts

Characteristic	Analytical cohort[a] n=697 (76%)	Non-analytical cohort n=226 (24%)	P-value
Male sex	369 (53)	119 (53)	0.99
Household sibling	302 (43)	74 (33)	0.006
Maternal history of asthma	52 (7)	19 (8)	0.75
Parental history of asthma	86 (12)	34 (15)	0.35
Maternal smoking during pregnancy	32 (5)	18 (8)	0.08
Birth by Caesarean section	90 (13)	34 (15)	0.48
Prematurity (< 37 weeks)	30 (4)	8 (4)	0.76
Low birth weight (< 2500 g)	21 (3)	4 (2)	0.45
Small for gestational age	14 (2)	4 (2)	0.99
Breastfed during age 0-2 months[b]	555 (80)	72 (32)	0.33
Parental smoking[c]	88 (13)	16 (7)	0.37
Eczema by age 13 months	108 (15)	24 (11)	0.22
Outside home day care at age 13 months	154 (22)	31 (14)	0.50
Asthma at age 7 years	56 (8)	19 (8)	0.90

Data are no. (%) of children unless otherwise indicated.
[a] Analytical cohort comprised children with data on antibiotic use during age 0-11 months and microbiota data during age 2-24 months.
[b] Data on breastfeeding available for 716 (78%) children.
[c] Data on parental smoking available for 635 (69%) children.

Supplementary Table 4. Unadjusted and adjusted associations of antibiotic treatments during age 0-11 months with longitudinal nasal microbiota profiles during age 2-24 months[a] in 697 children enrolled in the STEPS Study

	Antibiotic treatments during age 0-11 months (exposure)			
	Unadjusted analysis		**Multivariable-adjusted analysis[b]**	
Longitudinal microbiota profiles (dependent variable)	RRR (95% CI), per each antibiotic treatment	P-value	RRR (95% CI), per each antibiotic treatment	P-value
Profile A with persistent *Moraxella* dominance (n=279, 40%)	reference		Reference	
Profile B with *Streptococcus*-to-*Moraxella* transition (n=84, 12%)	0.98 (0.82-1.18)	0.82	1.16 (0.92-1.45)	0.21
Profile C with early *Dolosigranulum/ Corynebacteriaceae* dominance (n=139, 20%)	1.09 (0.95-1.25)	0.24	1.20 (1.01-1.43)	0.04
Profile D with early *Moraxella* sparsity with its subsequent increase (n=100, 14%)	**1.18 (1.02-1.37)**	**0.03**	**1.38 (1.15-1.66)**	**<0.001**
Profile E with mixed longitudinal patterns (n=92, 13%)	1.05 (0.89-1.24)	0.56	1.20 (0.98-1.48)	0.08

Abbreviations: CI, confidence interval; RRR, relative rate ratio

[a] Longitudinal clustering of nasal microbiota during age 2-24 months identified 6 distinct profiles. Of these, the profile F included only 3 children and was excluded from the analysis. To examine the association between frequency of antibiotic treatments and derived longitudinal microbiota profiles, multinomial logistic regression models were constructed. Profile A with persistent *Moraxella* dominance (low-risk profile) was used as the reference.
[b] Multinomial logistic regression model adjusting for potential confounders (sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and ARIs during age 0-11 months).

Supplementary Table 5. Unadjusted and adjusted associations of broad-spectrum antibiotic treatments during age 0-11 months[a] with longitudinal nasal microbiota profiles during age 2-24 months[b] in 697 children enrolled in the STEPS Study

	Broad-spectrum antibiotic treatments during age 0-11 months (exposure)			
	Unadjusted analysis		**Multivariable-adjusted analysis[c]**	
Longitudinal nasal microbiota profiles (dependent variable)	RRR (95% CI), per each antibiotic treatment	P-value	RRR (95% CI), per each antibiotic treatment	P-value
Profile A with persistent *Moraxella* dominance (n=279, 40%)	reference		reference	
Profile B with *Streptococcus*-to-*Moraxella* transition (n=84, 12%)	1.02 (0.74-1.40)	0.90	1.16 (0.80-1.67)	0.44
Profile C with early *Dolosigranulum/ Corynebacteriaceae* dominance (n=139, 20%)	1.12 (0.87-1.44)	0.39	1.16 (0.87-1.55)	0.32
Profile D with early *Moraxella* sparsity with its subsequent increase (n=100, 14%)	**1.47 (1.16-1.88)**	**<0.001**	**1.74 (1.31-2.30)**	**<0.001**

Profile E with mixed longitudinal patterns (n=92, 13%)	1.14 (0.86-1.52)	0.37	1.30 (0.94-1.79)	0.12

Abbreviations: CI, confidence interval; RRR, relative rate ratio

[a] Broad-spectrum antibiotics included broad-spectrum penicillins (e.g., amoxicillin-clavulanate), second and third generation cephalosporins, macrolides, and aminoglycosides.

[b] Longitudinal clustering of nasal microbiota during age 2-24 months identified 6 distinct profiles. Of these, the profile F included only 3 children and was excluded from the analysis. To examine the association between frequency of antibiotic treatments and derived longitudinal microbiota profiles, multinomial logistic regression models were constructed. Profile A with persistent *Moraxella* dominance (low-risk profile) was used as the reference.

[c] Multinomial logistic regression model adjusting for potential confounders (sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and ARIs during age 0-11 months).

Supplementary Table 6. Richness, alpha-diversity, and relative abundance by longitudinal nasal microbiota profile during age 2-24 months in 697 children enrolled in the STEPS Study

		Longitudinal nasal microbiota profile (age 2-24 months)		
	Age, months	**Low-risk profile with persistent *Moraxella* dominance** n=279 (40%)	**High-risk profile with early *Moraxella* sparsity** n=418 (60%)	**P-value**
Richness				
Number of OTUs, median (IQR)	2	15 (9-23)	24 (15-37)	<0.001
	13	15 (8-24)	46 (31-62)	<0.001
	24	12 (6-21)	28 (11-58)	<0.001
Alpha-diversity				
Shannon index, median (IQR)	2	0.73 (0.36-1.05)	1.15 (0.71-1.64)	<0.001
	13	0.51 (0.19-0.78)	1.51 (0.82-2.35)	<0.001
	24	0.58 (0.26-0.93)	0.85 (0.38-1.89)	<0.001
Relative abundance of 20 most abundant genera, mean (SD)				
Moraxella	2	0.38 (0.43)	0.18 (0.34)	<0.001*
	13	0.72 (0.33)	0.31 (0.38)	<0.001*
	24	0.70 (0.33)	0.48 (0.41)	<0.001*
Dolosigranulum	2	0.23 (0.30)	0.13 (0.24)	<0.001*
	13	0.14 (0.23)	0.19 (0.27)	0.009*
	24	0.12 (0.20)	0.15 (0.25)	0.09*
Streptococcus	2	0.09 (0.18)	0.19 (0.21)	<0.001*
	13	0.06 (0.16)	0.12 (0.16)	<0.001*
	24	0.09 (0.20)	0.11 (0.19)	0.19*
Staphylococcus	2	0.11 (0.25)	0.20 (0.28)	<0.001*
	13	0.01 (0.06)	0.03 (0.09)	<0.001*

	24	0.00 (0.02)	0.03 (0.10)	<0.001*
Corynebacteriaceae genus 1	2	0.08 (0.16)	0.07 (0.16)	0.44*
	13	0.02 (0.05)	0.05 (0.11)	<0.001*
	24	0.02 (0.05)	0.03 (0.08)	0.04*
Haemophilus	2	0.01 (0.03)	0.01 (0.07)	0.05*
	13	0.03 (0.14)	0.05 (0.12)	0.09*
	24	0.04 (0.15)	0.04 (0.10)	0.73*
Corynebacteriaceae genus 2	2	0.04 (0.13)	0.06 (0.13)	0.09*
	13	0.00 (0.00)	0.01 (0.04)	<0.001*
	24	0.00 (0.00)	0.00 (0.03)	0.02*
Neisseriaceae genus 1	2	0.02 (0.09)	0.04 (0.13)	0.03*
	13	0.01 (0.03)	0.02 (0.07)	<0.001*
	24	0.01 (0.03)	0.01 (0.05)	0.04*
Neisseria	2	0.00 (0.01)	0.00 (0.01)	0.17*
	13	0.00 (0.00)	0.03 (0.04)	<0.001*
	24	0.00 (0.02)	0.02 (0.04)	<0.001*
Gemella	2	0.01 (0.03)	0.02 (0.04)	<0.001*
	13	0.00 (0.00)	0.02 (0.03)	<0.001*
	24	0.00 (0.00)	0.01 (0.02)	<0.001*
Veillonella	2	0.00 (0.01)	0.01 (0.03)	<0.001*
	13	0.00 (0.00)	0.01 (0.02)	<0.001*
	24	0.00 (0.00)	0.01 (0.01)	<0.001*
Alloprevotella	2	0.00 (0.00)	0.00 (0.01)	0.15*
	13	0.00 (0.00)	0.02 (0.03)	<0.001*
	24	0.00 (0.00)	0.01 (0.02)	<0.001*
Granulicatella	2	0.00 (0.00)	0.00 (0.01)	<0.001*
	13	0.00 (0.00)	0.02 (0.02)	<0.001*
	24	0.00 (0.00)	0.01 (0.01)	<0.001*
Lactococcus	2	0.00 (0.02)	0.00 (0.00)	0.49*

	13	0.00 (0.00)	0.01 (0.03)	<0.001*
	24	0.00 (0.00)	0.01 (0.05)	0.002*
Acinetobacter	2	0.01 (0.06)	0.00 (0.02)	0.73*
	13	0.00 (0.00)	0.00 (0.01)	<0.001*
	24	0.00 (0.00)	0.01 (0.03)	0.001*
Lactobacillus	2	0.01 (0.06)	0.01 (0.04)	0.44*
	13	0.00 (0.00)	0.00 (0.01)	<0.001*
	24	0.00 (0.00)	0.00 (0.01)	0.004*
Rothia	2	0.00 (0.01)	0.01 (0.01)	<0.001*
	13	0.00 (0.00)	0.00 (0.01)	<0.001*
	24	0.00 (0.00)	0.00 (0.01)	<0.001*
Prevotellaceae genus 1	2	0.00 (0.00)	0.00 (0.02)	0.004*
	13	0.00 (0.00)	0.01 (0.02)	<0.001*
	24	0.00 (0.00)	0.00 (0.01)	<0.001*
Enhydrobacter	2	0.00 (0.00)	0.00 (0.01)	<0.001*
	13	0.00 (0.00)	0.00 (0.01)	<0.001*
	24	0.00 (0.00)	0.00 (0.01)	<0.001*
Porphyromonas	2	0.00 (0.01)	0.00 (0.00)	0.44*
	13	0.00 (0.00)	0.01 (0.01)	<0.001*
	24	0.00 (0.00)	0.00 (0.01)	<0.001*

Abbreviations: IQR, interquartile range; OTU, operational taxonomic unit; SD, standard deviation

* Benjamini-Hochberg false-discovery rate adjusted P-value accounting for multiple comparisons

Supplementary Table 7. Association of antibiotic treatments during age 0-11 months with two longitudinal nasal microbiota profiles during age 2-24 months (n=697)[a]

	Antibiotic treatments during age 0-11 months (exposure)		**Broad-spectrum antibiotic treatments during age 0-11 months[b] (exposure)**	
Dichotomized longitudinal nasal microbiota profiles (dependent variable)	RRR (95% CI), per each antibiotic treatment	P-value	RRR (95% CI), per each antibiotic treatment	P-value
Low-risk profile with persistent *Moraxella* dominance (profile A) n=279 (40%)	reference		reference	
High-risk profile with early *Moraxella* sparsity (profiles B-F), n=418 (60%)	**1.24 (1.09-1.42)**	**0.001**	**1.35 (1.09-1.67)**	**0.006**

Abbreviations: CI, confidence interval; OR, odds ratio

[a] Longitudinal clustering of nasal microbiota during age 2-24 months identified 6 distinct profiles. For the mediation analysis, longitudinal nasal microbiota profiles were dichotomized to 1) low-risk profile with persistent *Moraxella* dominance (profile A) and 2) high-risk profile with early *Moraxella* sparsity (profiles B-F, **Figure 3b**). Logistic regression model adjusting for potential confounders (sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and ARIs during age 0-11 months). Low-risk profile with persistent *Moraxella* dominance was used as the reference.

[b] Narrow-spectrum antibiotics were defined as narrow-spectrum penicillins (amoxicillin, phenoxymethylpenicillin, benzylpenicillin, and ampicillin), first generation cephalosporins, and sulfonamides. All other antibiotics were defined as broad-spectrum antibiotics, including broad-spectrum penicillins (e.g., amoxicillin-clavulanate), second and third generation cephalosporins, macrolides, and aminoglycosides.

Supplementary Table 8. Antibiotic treatments during age 0-11 months and asthma at age 7 years (n=697)

Number of antibiotic treatments during age 0-11 months	**Children with asthma, n (%)**[a]
0 (n=338)	21 (6.2%)
1 (n=163)	13 (8.0%)
≥2 (n=196)	22 (11.2%)

[a] Medical records data for asthma were missing from 6 children.

Supplementary Table 9. Sensitivity analysis for mediation analysis, using a cut-off of 0-2 vs. ≥3 antibiotic treatments during age 0-11 months as the exposure (n=623)[a]

	Antibiotic treatments (≥3) during age 0-11 months	
	Absolute risk difference (95% CI)	**P-value**
Total effect	4.8% (1.5% - 8.3%)	<0.001
Natural direct effect	4.2% (1.1% - 7.5%)	0.008
Natural indirect effect	0.6% (0.1% - 1.3%)[b]	0.03

Abbreviation: CI, confidence interval.

[a] Causal mediation analysis estimating the total and direct effects of antibiotic exposure (0-2 vs. ≥3) during age 0-11 months on risk of developing asthma by age 7 years as well as indirect effect by longitudinal changes in nasal microbiota during age 2-24 months (low-risk profile with persistent *Moraxella* dominance vs. high-risk profile with early *Moraxella* sparsity). Inverse probability weighting with marginal structural models was used in the mediation analysis to account for potential confounders (i.e., sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and acute respiratory infections during age 0-11 months).

[b] Proportion of indirect effect by antibiotic exposure was 11.4% (0.9-40.9%)

Supplementary Table 10. Sensitivity analysis for mediation analysis, with antibiotic treatments during age 0-2 months as the exposure (n=623)[a]

	Antibiotic treatment (≥1) during age 0-2 months		Broad-spectrum antibiotic treatment (≥1) during age 0-2 months	
	Absolute risk difference (95% CI)	**P-value**	**Absolute risk difference (95% CI)**	**P-value**
Total effect	2.9 (−0.3 - 6.2)	0.08	6.2 (2.9 - 9.5)	<0.001
Natural direct effect	2.4 (−0.7 - 5.5)	0.14	5.8 (2.7 - 9.1)	<0.001
Natural indirect effect	0.5 (0.0 - 1.2)	0.06	0.3 (−0.3 - 1.1)	0.36

Abbreviation: CI, confidence interval.

[a] Causal mediation analysis estimating the total and direct effects of antibiotic exposure during age 0-2 months on risk of developing asthma by age 7 years as well as indirect effect by longitudinal changes in nasal microbiota during age 2-24 months (low-risk profile with persistent *Moraxella* dominance vs. high-risk profile with early *Moraxella* sparsity). Inverse probability weighting with marginal structural models was used in the mediation analysis to account for potential confounders (i.e., sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and acute respiratory infections during age 0-2 months).

Supplementary Table 11. Sensitivity analysis for mediation analysis, with the longitudinal nasal microbiota profiles being dichotomized into the profile with the lowest *Moraxella* abundance vs. other profiles (n=623)[a]

	Antibiotic treatments (≥2) during age 0-11 months		Broad-spectrum antibiotic treatment (≥1) during age 0-11 months	
	Absolute risk difference (95% CI)	**P-value**	**Absolute risk difference (95% CI)**	**P-value**
Total effect	4.0% (0.9% - 7.1%)	0.01	3.6% (0.5% - 6.6%)	0.02
Natural direct effect	3.6% (0.5% - 6.7%)	0.02	3.0% (−0.0% - 6.0%)	0.05
Natural indirect effect	0.4% (−0.0% - 1.2%)[b]	0.08	0.6% (0.1% - 1.4%)[c]	0.02

Abbreviation: CI, confidence interval.

[a] Causal mediation analysis estimating the total and direct effects of antibiotic exposure during age 0-11 months on risk of developing asthma by age 7 years as well as indirect effect by longitudinal changes in nasal microbiota during age 2-24 months. The longitudinal nasal microbiota profiles were dichotomized into the profile with the lowest *Moraxella* abundance (profile D) vs. other profiles (profiles A, B, C, E, F) as the mediator. Inverse probability weighting with marginal structural models was used in the mediation analysis to account for potential confounders (i.e., sex, parental history of asthma, household siblings, breastfeeding during age 0-2 months, and acute respiratory infections during age 0-11 months).

[b] Proportion of indirect effect by antibiotic exposure was 10.1% (95% CI, −1.0% - 43.8%).

[c] Proportion of indirect effect by broad-spectrum antibiotic exposure was 16.6% (95% CI, 1.1% - 76.1%)