ONLINE SUPPLEMENT

Early nasal microbiota and acute respiratory infections during the first years of life

Laura Toivonen, MD, PhD; Kohei Hasegawa, MD, MPH; Matti Waris, PhD; Nadim J. Ajami, PhD; Joseph F. Petrosino, PhD; Carlos A. Camargo, Jr., MD, DrPH; and Ville Peltola, MD, PhD

Supplement materials

Supplemental methods

Table E1. Comparison of infants with and without qualified nasal microbiota data at age 2 months

Table E2. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles in asymptomatic infants at age 2 months with rate of acute respiratory infections (ARI) and lower respiratory infections (LRTI) during age 2-24 months

Table E3. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of non-LRTIs (i.e., ARIs not including physician-diagnosed LRTIs) during age 2-24 months

Table E4. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of days with symptoms of acute respiratory infections (ARI) during age 2-24 months

Table E5. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of acute respiratory infections (ARI) during age 2-24 months

Table E6. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of lower respiratory tract infections (LRTI) during age 2-24 months

Table E7. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with risk of recurrent wheezing by age 24 months

Figure E1. The optimal number of nasal microbiota profiles

Figure E2. Enrolment and follow-up of the study children

Figure E3. Kaplan-Meier curve of nasal microbiota profiles at age 2 months and risk of recurrent wheezing by age 24 months

SUPPLEMENTAL METHODS

Study design, setting, and participants

We analysed data from the prospective, population-based, birth-cohort study called the Steps to the Healthy Development and Well-being of Children (STEPS study). From all infants born in the Hospital District of Southwest Finland from January 2008 through April 2010 to Finnish or Swedish-speaking mothers (eligible cohort, n = 9811 mothers; n = 9936 infants), families of 1827 infants (30 pairs of twins) were enrolled either during the first trimester of pregnancy or soon after birth and are followed until early adulthood.[1] Of these infants, 923 infants were enrolled in an intensive follow-up for acute respiratory infections (ARI) from birth to age 24 months.[2] No selection criteria other than language (Finnish or Swedish speaking family) were applied for recruiting the families for the STEPS study or for the ARI follow-up subcohort. Based on data from the Finnish National Birth Registry,[1] the participating and nonparticipating children were generally 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 and the age of their mothers were somewhat higher (mean age 30.8 (SD, 4.6) years in participating vs. 30.1 (SD, 5.2) years in non-participating mothers).

Data collection

Infants were followed for ARIs from birth to age 24 months with daily symptom diaries. Families were encouraged to visit the study clinic at the Turku Centre for Child and Youth Research, Turku University Hospital and University of Turku (Turku, Finland) during ARIs if they felt that an evaluation by a physician was needed. Nasal swab specimens were collected by study personnel using flocked nylon swabs (Copan, Brescia, Italy) at a scheduled participant visit at age 2 months. Demographics, family history, pre-, peri-, and post-natal history, and environmental information were collected from the National Birth Registry and by structured questionnaires. Data on outpatient and emergency department visits at hospitals and hospitalizations from birth to age 24 months were retrieved from medical records of

the Hospital District of Southwest Finland which comprises information from both hospitals providing inpatient pediatric care in the area (Turku University Hospital and Salo District Hospital). All data were reviewed at the Turku Centre for Child and Youth Research.

Respiratory Virus Detection

Nasal samples were analysed for respiratory viruses at the University of Turku (Turku, Finland). Nucleic acids were extracted by NucliSense easyMag (BioMerieux, Boxtel, Netherlands) or MagnaPure 96 (Roche, Penzberg, Germany) automated extractor. Extracted RNA was reverse transcribed and the cDNA was amplified using real-time, quantitative PCR for rhinovirus, human enteroviruses and respiratory syncytial virus (RSV) as described earlier.[3 4] All specimens collected during influenza seasons were analysed by reverse transcriptase PCR for influenza A and B viruses.[5] The first and last day of each influenza season were defined on the basis of the influenza antigen test results and data from the infectious disease surveillance registry of the National Institute for Health and Welfare, Finland.[6]

16S rRNA gene sequencing of nasal airway microbiota

16S rRNA gene sequencing methods were adapted from the methods developed for the National Institutes of Health (NIH) Human Microbiome Project.[7 8] 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 Life Science) 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 done at Turku Centre for Biotechnology (Turku, Finland) and extracted DNAs were sent for microbiota testing to Baylor College of Medicine (Houston, TX, USA).

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.[9 10] Sequencing read pairs were demultiplexed based on the unique molecular barcodes, and reads were merged using USEARCH v7.0.1090,[11] 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.[12] OTUs were determined by mapping the centroids to the SILVA database[13] 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 of alpha-diversity (e.g., Shannon index) and beta-diversity (e.g., Bray-Curtis).[14]

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.

Outcome definition

The primary outcome was the incidence rate of all ARIs (including upper ARIs and LRTIs). The secondary outcomes were 1) the rate of LRTIs during age 2-24 months and recurrent wheezing by age 24 months. 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 ARI diagnosed by a physician.[2] The duration of 97.2% of ARIs was ≤30 days. To account for overlapping infections, the length of an ARI was limited to 30 days; longer ARIs (2.8%) were calculated as separate episodes with a maximum duration of 30 days. LRTIs (e.g., bronchiolitis, pneumonia) were diagnosed by study physicians or attending physicians at other clinics or hospitals. Repeated diagnoses of LRTIs within 14 days were calculated as one episode. Recurrent wheezing was defined as ≥3 wheezing episodes (with or without other respiratory symptoms) during age 2-24 months reported by the parents in the diary or diagnosed by a physician. In the sensitivity analyses, we repeated the analysis by fitting a negative binomial regression model with 1) excluding infants with respiratory symptoms at the time of age 2-month nasal sampling, 2) modelling the rate of non-LRTIs (i.e., ARIs not including physician-diagnosed LRTIs) during age 2-24 months as the outcome, and 3) modelling the number of days with ARI symptoms during age 2-24 months as the outcome.

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. Osterback R, Tevaluoto T, Ylinen T, *et al*. Simultaneous detection and differentiation of human rhino- and enteroviruses in clinical specimens by real-time PCR with locked nucleic Acid probes. *J Clin Microbiol* 2013;51:3960-7.

4. Toivonen L, Schuez-Havupalo L, Rulli M, *et al*. Blood MxA protein as a marker for respiratory virus infections in young children. *J Clin Virol* 2015;62:8-13.

5. Jokela P, Vuorinen T, Waris M, *et al*. Performance of the Alere i influenza A&B assay and mariPOC test for the rapid detection of influenza A and B viruses. *J Clin Virol* 2015;70:72-6.

6. Teros-Jaakkola T, Toivonen L, Schuez-Havupalo L, *et al*. Influenza virus infections from 0 to 2 years of age: A birth cohort study. *J Microbiol Immunol Infect* 2017; Dec 7. doi: 10.1016/j.jmii.2017.10.007.

7. A framework for human microbiome research. *Nature* 2012;486:215-21.

8. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207-14.

9. Caporaso JG, Kuczynski J, Stombaugh J, *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335-6.

10. 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.

11. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460-1.

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

13. 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.

14. Bray J, Curtis J. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* 1957;27:326-49.

Table E1. Comparison of infants with and without qualified nasal microbiota data at age 2 months

Characteristic	Nasal microbiota data at age 2 months: Infants with qualified data, n=839 (90.1%)	Infants without qualified data,[a] n=84 (9.1%)	P-value
Age at 2-month scheduled visit, months, median (IQR)	2.6 (2.4-2.7)	2.5 (2.3-2.7)	0.26
Male sex	443 (52.8)	45 (53.6)	0.98
Household sibling	350 (41.7)	26 (31.0)	0.07
Maternal asthma	64 (7.6)	7 (8.3)	0.83
Parental asthma	109 (13.0)	11 (13.1)	0.99
Maternal smoking during pregnancy	46 (5.5)	4 (4.8)	0.99
Birth by Caesarean section	111 (13.2)	13 (15.5)	0.68
Premature (<37 weeks)	34 (4.1)	4 (4.8)	0.77
Small for gestational age	16 (1.9)	2 (2.4)	0.68
Intrapartum antibiotics	105 (12.5)	12 (14.3)	0.78
Breastfeeding during the first 2 months of life[b]	571 (68.1)	56 (66.7)	0.80
Systemic antibiotic use before age 2-month nasal sample	141 (16.8)	13 (15.5)	0.51
Age 2-month scheduled visit			
Mild respiratory symptoms	135 (16.1)	15 (20.0)	0.48
Respiratory virus in nasal sample	134 (16.0)	7 (11.3)	0.43
Clinical outcomes during age 2-24 months[c]			
ARIs, incidence rate /100 children-years (95% CI)			
All ARIs	656 (635-679)	649 (574-735)	0.74
Days with ARI symptoms	4982 (4698-5288)	4364 (3588-5379)	0.57
LRTIs	24 (20-29)	19 (8-54)	0.22
Recurrent wheezing[d]	92 (11.6)	7 (11.9)	0.99

Data are no. (%) of infants unless otherwise indicated. Percentages may not equal 100 because of missingness.

Abbreviations: ARI, acute respiratory infection; IQR, interquartile range; LRTI, lower respiratory tract infection.

[a] Infants who did not undergo nasal sampling (n=23) and those with unqualified microbiota data (n=61).

[b] Data for breastfeeding available for 716 (78%) infants.

[c] Data on ARIs at age 2-24 months was available for 876 (95%) children. Incidence rates were calculated using negative binomial distribution and log link with natural logarithm of follow-up time as an offset.

[d] ≥3 wheezing episodes at age 2-24 months.

Table E2. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles in asymptomatic infants at age 2 months (n=704) with rate of acute respiratory infections (ARI) and lower respiratory infections (LRTI) during age 2-24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Outcome by microbiota profile	**Incidence rate ratio (95% CI)**	**P-value**	**Incidence rate ratio (95% CI)**	**P-value**
ARIs (n=7,810)				
Moraxella-dominant	1.27 (1.09-1.49)	0.003	1.19 (1.01-1.39)	0.03
Streptococcus-dominant	1.14 (0.97-1.33)	0.12	1.12 (0.96-1.30)	0.15
Dolosigranulum-dominant	1.09 (0.93-1.27)	0.31	1.05 (0.90-1.23)	0.50
Staphylococcus-dominant	1.13 (0.96-1.33)	0.15	1.11 (0.95-1.30)	0.20
Corynebacteriaceae-dominant	reference	-	reference	-
LRTIs (n=285)				
Moraxella-dominant	3.43 (1.16-10.90)	0.03	3.14 (1.01-10.46)	0.049
Streptococcus-dominant	1.86 (0.61-6.08)	0.29	1.61 (0.52-5.41)	0.41
Dolosigranulum-dominant	2.33 (0.77-7.52)	0.15	2.18 (0.71-7.15)	0.18
Staphylococcus-dominant	3.69 (1.23-11.96)	0.02	3.28 (1.08-10.69)	0.04
Corynebacteriaceae-dominant	reference	-	reference	-

Abbreviations: ARI, acute respiratory infection, CI, confidence interval, LRTI, lower respiratory tract infection

[a] For the sensitivity analysis, infants with respiratory symptoms at the time of age 2-month nasal sampling were excluded (n=135). Incidence rates of ARIs and LRTIs were compared by using unadjusted and multivariable-adjusted negative binomial regression models with follow-up time from age 2-month visit to age 24 months as an offset variable. Multivariable analysis adjusted for 8 patient-level covariates (age, sex, household siblings, parental asthma, birth by Caesarean section, breastfeeding, history of systemic antibiotic use, and respiratory virus in age 2-month sample). *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

Table E3. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of non-LRTIs (i.e., ARIs not including physician-diagnosed LRTIs) during age 2-24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Outcome by nasal microbiota profile	**Incidence rate ratio (95% CI)**	**P-value**	**Incidence rate ratio (95% CI)**	**P-value**
Non-LRTIs (i.e., ARIs not including physician-diagnosed LRTIs) (n=7,525)				
Moraxella-dominant	**1.30 (1.13-1.50)**	**<0.001**	**1.16 (1.01-1.34)**	**0.03**
Streptococcus-dominant	1.14 (0.99-1.32)	0.08	1.12 (0.97-1.29)	0.11
Dolosigranulum-dominant	1.09 (0.94-1.27)	0.23	1.06 (0.92-1.22)	0.43
Staphylococcus-dominant	1.10 (0.95-1.28)	0.21	1.11 (0.96-1.28)	0.17
Corynebacteriaceae-dominant	reference	-	reference	-

Abbreviations: ARI, acute respiratory infection; CI, confidence interval; LRTI, lower respiratory tract infection

[a] Incidence rates of non-LRTIs (i.e., ARIs not including physician-diagnosed LRTIs) were compared by using unadjusted and multivariable-adjusted negative binomial regression models with follow-up time from age 2-month visit to age 24 months as an offset variable. Multivariable analysis adjusted for 9 patient-level covariates (age, sex, household siblings, parental asthma, birth by Caesarean section, breastfeeding, history of systemic antibiotic use, respiratory symptoms at the time of nasal sampling, and respiratory virus in age 2-month sample). *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

Table E4. Sensitivity analysis: Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of days with symptoms of acute respiratory infections (ARI) during age 2-24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Variable	**Incidence rate ratio (95% CI)**	**P-value**	**Incidence rate ratio (95% CI)**	**P-value**
Nasal microbiota profile				
Moraxella-dominant	1.88 (1.47-2.38)	<0.001	1.61 (1.25-2.06)	<0.001
Streptococcus-dominant	1.56 (1.21-1.99)	<0.001	1.56 (1.22-1.99)	<0.001
Dolosigranulum-dominant	1.43 (1.11-1.82)	0.005	1.42 (1.11-1.81)	0.005
Staphylococcus-dominant	1.44 (1.11-1.86)	0.005	1.52 (1.18-1.96)	0.001
Corynebacteriaceae-dominant	reference	-	reference	-
Age at nasal sample collection (per each incremental week)	-		0.99 (0.95-1.03)	0.59
Male sex	-		1.09 (0.97-1.22)	0.13
Household sibling	-		1.20 (1.05-1.37)	0.006
Parental asthma	-		1.18 (1.00-1.42)	0.06
Birth by Caesarean section	-		0.78 (0.66-0.93)	0.005
Breastfeeding during the first 2 months of life	-		1.21 (1.01-1.44)	0.04
Systemic antibiotic use before age 2-month nasal sampling	-		0.96 (0.83-1.13)	0.62
Respiratory symptoms at the time of age 2-month sampling	-		1.45 (1.22-1.73)	<0.001
Respiratory virus detected in age 2-month sample	-		0.92 (0.78-1.10)	0.33

Abbreviations: ARI, acute respiratory infection; CI, confidence interval

[a] Rates of days with ARI symptoms were compared by using unadjusted and multivariable-adjusted negative binomial regression models with completed follow-up time of symptom diary from age 2-month visit to age 24 months as an offset variable. *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

Table E5. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of acute respiratory infections (ARI) during age 2-24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Variable	**Incidence rate ratio (95% CI)**	**P-value**	**Incidence rate ratio (95% CI)**	**P-value**
Nasal microbiota profile				
Moraxella-dominant	1.34 (1.16-1.54)	<0.001	1.19 (1.04-1.37)	0.01
Streptococcus-dominant	1.16 (1.00-1.35)	0.04	1.14 (0.99-1.31)	0.08
Dolosigranulum-dominant	1.11 (0.96-1.29)	0.15	1.08 (0.94-1.24)	0.31
Staphylococcus-dominant	1.14 (0.98-1.32)	0.10	1.14 (0.98-1.31)	0.09
Corynebacteriaceae-dominant	reference	-	reference	-
Age at nasal sample collection (per each incremental week)	-		0.99 (0.97-1.01)	0.46
Male sex	-		1.07 (1.00-1.14)	0.04
Household sibling	-		1.20 (1.12-1.29)	<0.001
Parental asthma	-		1.04 (0.94-1.14)	0.46
Birth by Caesarean section	-		0.92 (0.84-1.02)	0.10
Breastfeeding during the first 2 months of life	-		1.09 (0.98-1.21)	0.09
Systemic antibiotic use before age 2-month nasal sampling	-		1.01 (0.93-1.10)	0.76
Respiratory symptoms at the time of age 2-month sampling	-		1.13 (1.04-1.24)	0.005
Respiratory virus detected in age 2-month sample	-		0.98 (0.90-1.08)	0.75

Abbreviations: ARI, acute respiratory infection; CI, confidence interval

[a] Rates of ARIs were compared by using unadjusted and multivariable-adjusted negative binomial regression models with follow-up time from age 2-month visit to age 24 months as an offset variable. *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

Table E6. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of lower respiratory tract infections (LRTI) during age 2-24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Variable	**Incidence rate ratio (95% CI)**	**P-value**	**Incidence rate ratio (95% CI)**	**P-value**
Nasal microbiota profile				
Moraxella-dominant	3.89 (1.47-11.06)	0.008	2.79 (1.04-8.09)	0.04
Streptococcus-dominant	2.62 (0.95-7.67)	0.07	1.89 (0.69-5.55)	0.22
Dolosigranulum-dominant	2.38 (0.87-6.95)	0.10	1.95 (0.72-5.69)	0.19
Staphylococcus-dominant	3.42 (1.24-10.13)	0.02	2.90 (1.07-8.49)	0.04
Corynebacteriaceae-dominant	reference	-	reference	-
Age at nasal sample collection (per each incremental week)	-		0.98 (0.84-1.13)	0.75
Male sex	-		1.37 (0.92-2.03)	0.11
Household sibling			1.26 (0.82-1.96)	0.28
Parental asthma	-		1.24 (0.69-2.24)	0.46
Birth by Caesarean section	-		1.30 (0.74-2.30)	0.36
Breastfeeding during the first 2 months of life	-		1.09 (0.56-2.14)	0.79
Systemic antibiotic use before age 2-month nasal sampling	-		1.14 (0.68-1.91)	0.61
Respiratory symptoms at the time of age 2-month sampling	-		1.79 (1.06-3.07)	0.03
Respiratory virus detected in age 2-month sample	-		0.52 (0.28-0.96)	0.04

Abbreviations: ARI, acute respiratory infection; CI, confidence interval

[a] Rates of lower respiratory tract infections were compared by using unadjusted and multivariable-adjusted negative binomial regression models with follow-up time from age 2-month visit to age 24 months as an offset variable. *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

Table E7. Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with risk of recurrent wheezing by age 24 months[a]

	Unadjusted analysis		Multivariable-adjusted analysis	
Variable	**Hazard ratio (95% CI)**	**P-value**	**Hazard ratio (95% CI)**	**P-value**
Nasal microbiota profile				
Moraxella-dominant	1.75 (0.95-3.21)	0.07	1.82 (0.94-3.52)	0.08
Streptococcus-dominant	reference	-	reference	-
Dolosigranulum-dominant	1.58 (0.82-3.04)	0.17	1.75 (0.90-3.41)	0.10
Staphylococcus-dominant	1.45 (0.73-2.91)	0.29	1.59 (0.78-3.21)	0.20
Corynebacteriaceae-dominant	1.13 (0.41-3.11)	0.81	1.21 (0.43-3.36)	0.72
Age at nasal sample collection (per each incremental week)	-		0.96 (0.83-1.11)	0.56
Male sex	-		0.60 (0.39-0.91)	0.02
Household sibling			0.97 (0.60-1.55)	0.89
Parental asthma	-		2.42 (1.45-4.03)	<0.001
Birth by Caesarean section	-		1.11 (0.60-2.03)	0.74
Breastfeeding during the first 2 months of life	-		0.99 (0.50-1.94)	0.97
Systemic antibiotic use before age 2-months nasal sampling	-		1.29 (0.76-2.18)	0.35
Respiratory symptoms at the time of age 2-month sampling	-		2.49 (1.50-4.14)	<0.001
Respiratory virus detected in age 2-month sample	-		0.41 (0.19-0.85)	0.02

Abbreviations: CI, confidence interval

[a] Risk of recurrent wheezing (≥3 wheezing episodes during age 2-24 months) was compared by using unadjusted and multivariable-adjusted Cox proportional hazards models. *Streptococcus*-dominant microbiota profile was used as the reference.

Figure E1. The optimal number of nasal microbiota profiles
The optimal number of nasal microbiota profiles was identified by using gap statistics based on Bray-Curtis distance and principal coordinates analysis

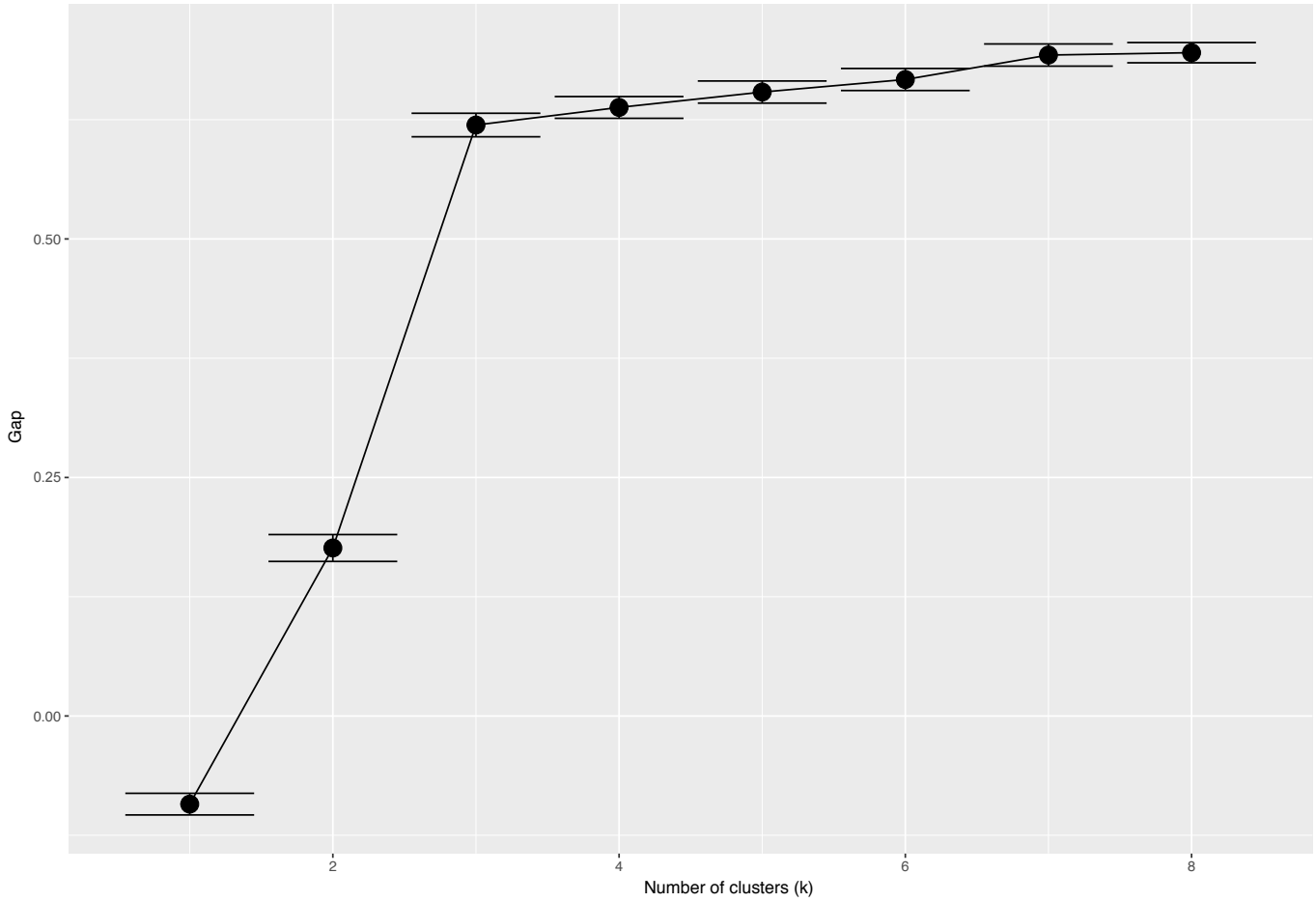

Figure E2. Enrolment and follow-up of the study children

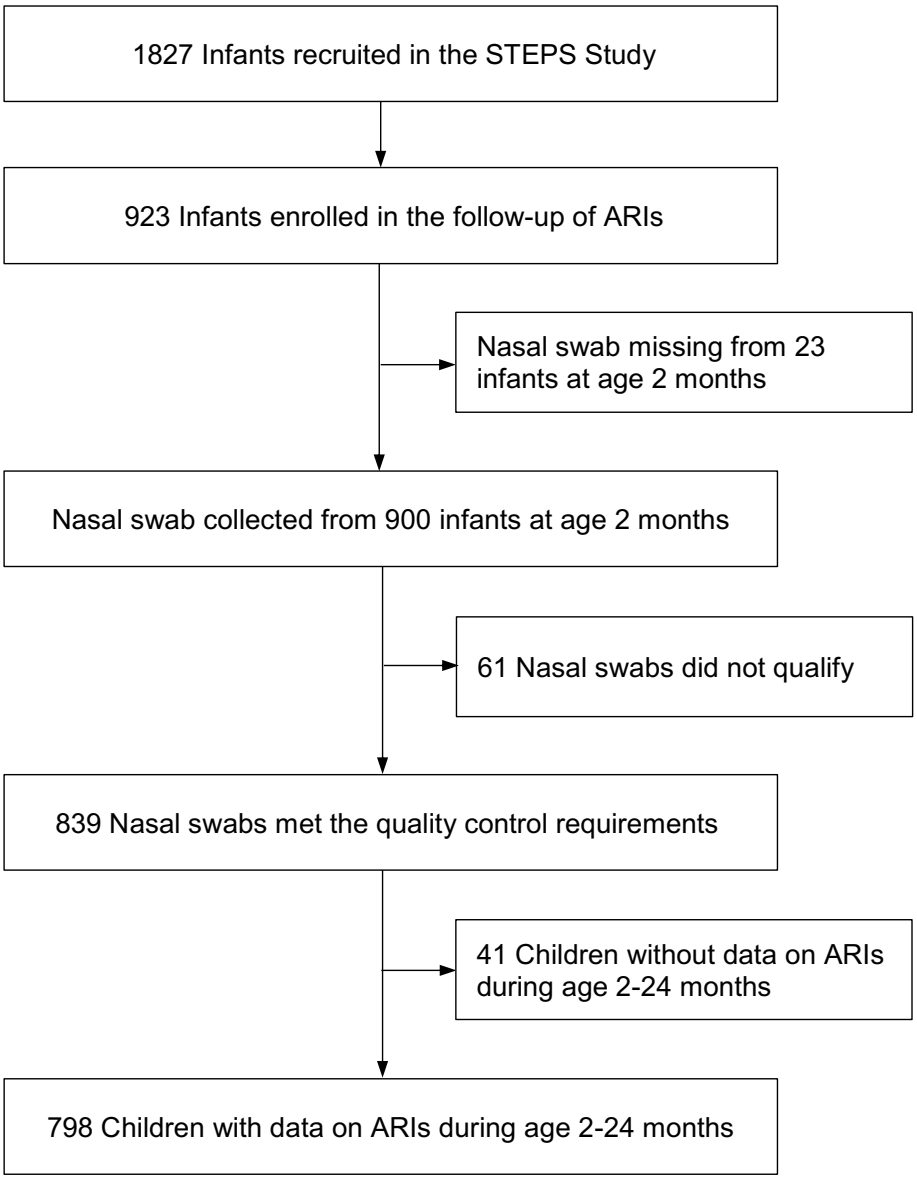

Figure E3. Kaplan-Meier curve of nasal microbiota profiles at age 2 months and risk of recurrent wheezing by age 24 months. P-value of the log-rank test=0.40. DP=dominant profile.