2022

THE THREE-WAY INTERPLAY AMONG EARLY LIFE EXPOSURES, THE GUT MICROBIOME, AND OUTCOMES IN INFANCY

Yuka Moroishi
Yuka.Moroishi.GR@dartmouth.edu

THE THREE-WAY INTERPLAY AMONG EARLY LIFE EXPOSURES, THE GUT MICROBIOME, AND OUTCOMES IN INFANCY

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

in

Quantitative Biomedical Sciences

by Yuka Moroishi

Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire
June 2022

Examining Committee:

_____
(Co-chair) Jiang Gui, Ph.D.

_____
(Co-chair) Margaret R. Karagas, Ph.D.

_____
Zhigang Li, Ph.D.

_____
Juliette C. Madan, M.D.

_____
Hongzhe Li, Ph.D.

_____
F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

Abstract

The bidirectional relationship between the gut microbiome and immune system plays an important role in host immune status: the immune system provides the gut microbiome the optimal environment to thrive in, and the gut microbiome helps regulate the immune system. This relationship is especially important in infants, whose immune system is still premature and rely on innate immunity.

We investigated the three-way interplay among early-life exposures, the developing gut microbiome, and outcomes in infancy from the general population in New Hampshire, US. We used prospective cohort data from the New Hampshire Birth Cohort study to 1) determine whether timing of baby rice cereal introduction is related to respiratory infections, symptoms, and allergy in infancy; 2) identify gut microbiome composition and bacterial species that may influence respiratory infections and symptoms; 3) identify bacterial species and metabolic pathways that associate with antibody response to pneumococcal capsular polysaccharide and tetanus toxoid vaccination; and 4) develop a statistical approach to test the mediating effect of the microbiome on the "causal" path between exposure and outcome. Our studies highlight the potential to modulate the infant gut microbiome to improve health outcomes in infancy.

Preface

I would like to thank my advisors Dr. Jiang Gui and Dr. Margaret R. Karagas for their mentorship. Their support and guidance have been critical to the success of my dissertation projects and other academic endeavors. Next, I would like to thank the members of my dissertation committee, Dr. Zhigang Li, Dr. Juliette C. Madan, and Dr. Hongzhe Li, for their important feedback on my projects. I would also like to thank members of my qualifying exam committee, Dr. H. Robert Frost, Dr. Anne G. Hoen, and Dr. A. James O'Malley, for their feedback on the early stages of these projects. I am also grateful for my colleagues and collaborators both at Dartmouth and outside of Dartmouth.

Lastly, this thesis would not be possible without the commitment of everyone involved in the New Hampshire Birth Cohort Study. I thank all participants, especially caregivers and infants, of the study who provided the rich data to make these analyses possible. I also thank the staff for ensuring smooth progress of the study.

This work was supported by the US National Institutes of Health under award numbers UH3OD023275, NIGMS P20GM104416, NIGMS R01GM123014, NIEHS P42ES007373, NIEHS P01ES022832, NLM R01LM012723, NLM K01LM011985, and NIEHS R21ES020936; and the U.S. Environmental Protection Agency under award RD83544201.

Table of Contents

List of Tables

List of Illustrations

Abbreviations

alr	additive log-ratio
AR	autoregressive
ASV	amplicon sequence variant
BMI	body mass index
CDC	Centers for Disease Control and Prevention
CI	confidence interval
clr	centered log-ratio
DTaP	Diphtheria, Tetanus, and acellular Pertussis vaccine
EU	European Union
FDA	Food and Drug Administration
FDR	false discovery rate
g	gram
GEE	generalized estimating equation
HT-MMIOW	hypothesis test for microbiome mediation using inverse odds weighting
Ig	Immunoglobulin
ilr	isometric log-ratio
kg	kilogram
L	liter
m	meter
N	sample size
NGS	next-generation sequencing
NHBCS	New Hampshire Birth Cohort Study
No.	number
PCA	principal component analysis
PCP	pneumococcal capsular polysaccharide
PCV	Pneumococcal Conjugate Vaccine
RR	relative risk
RSV	respiratory syncytial virus
RTI	respiratory tract infection
SD	standard deviation
TT	tetanus toxoid
UMAP	uniform manifold approximation and projection
US	United States
USA	United States of America

Chapter 1: General Introduction

1.1 Early life exposures and immune-related outcomes in infancy

Infections and allergy remain a leading cause of morbidity and mortality in the US. In additional to the physical burden on infants as a result of these outcomes are the side effects from treatment and long-term effects, and the substantial economic impacts[1,2]. There are several factors associated with infants' susceptibility to disease. Prenatal exposures, breast feeding status, and diet can all contribute to an infant's risk of developing an infection[3–7] or allergy[8–10]. Antibiotic use, including prenatal, peripartum, and prenatal exposures, can also influence risks[11–16].

Accumulating evidence indicates that contaminants in our environment, which may come through our diet, also affect our immunity to disease[17,18]. Arsenic, a widespread toxicant, has been shown to have adverse effects on human health through the life course. Rice grown in the United States is an important source of exposure to arsenic[19]. In particular, infants, including the U.S., commonly transition to solid food with rice cereal[20]. As a result, infants who have been fed rice products have high concentrations of urinary arsenic[21]. Arsenic exposure has been associated with impaired immune response and increased risk of infections and chronic diseases[22]. Previous research on *in utero* arsenic exposure found increase in maternal urinary arsenic associated with increased risk of infection and respiratory symptoms in the first year of infant life[23]. Findings of elevated

arsenic levels among infants consuming rice products[24] have raised concerns that rice cereal consumption may have detrimental effects on infant health. Exploring the health impacts of early life rice cereal exposure could elucidate this issue, but to date, very little research has been done on this.

1.2 Gut microbiome and immunity

The bidirectional relationship between the gut microbiome and immunity evolves following birth; the gut microbiome aids in the maturation of the early immune system while the immune system regulates host-microbe symbiosis[25,26]. While neonates rely on innate immunity to protect against pathogens, the both the innate and the adaptive immune systems co-evolve over time with the maturation of intestinal epithelial cells [27]. The relationship between gut microbiome composition and immune response has been largely studied in mice models. Colonization of *Clostridia*-related Segmented Filamentous Bacterium induces innate and adaptive T cell[28,29] and IgA response in mice[30,31]. Clostridium-abundant mice also demonstrate higher T-regs and lower IgE levels compared to control mice[32]. Bacterial symbionts, such as *Bifidobacterium adolescentis*, found in the human gut, trigger Th17 cells induction and accumulation in mice[33,34]. Both *Bacteroides fragilis* and *Faecalibacterium prausnitzii* species increase CD4T cells that produce IL-10[35–37]. *Lactobacillus reuteri* has shown to induce double-positive intraepithelial T lymphocytes[38], which aid in the prevention of inflammatory bowel disease[39]. Studies also have shown the gut microbiome to regulate hematopoiesis that can control host bacterial infection[40]

A growing body of research has linked human gut microbiome to a variety of health outcomes including inflammatory bowel disease and metabolic diseases in humans [41]. Much of this work has focused on adults, whereas far less is known on the effect of early-life gut microbiome composition on infant outcomes. The work, thus far, provides evidence of possible associations between infant gut microbiome and diseases such as type I diabetes, allergy, and necrotizing enterocolitis, with associations lasting into childhood and adulthood[42]. The infant gut microbiome has also been linked to child allergy and atopy[43], and there are data suggesting that modulating the gut microbiome could reduce risk[44–46]. The link between the early gut microbiome and respiratory infections[47,48] and vaccine response[49–52] in infancy, while plausible, is not understood.

1.3 Gut microbiome as a mediator between exposures and outcomes

The driving factors associated with adverse health outcomes in early life also shape the developing gut microbiome in early life, including delivery mode, diet, and antibiotic use[42,53]. Experimentally, the microbiome can modify contaminants[54]. In both laboratory and epidemiologic studies, arsenic may impact the gut microbiome, and *in vitro* undergoes pre-systemic metabolism to compounds with greater toxic potential [55].

Mediation analyses is becoming popular in microbiome research to elucidate the contribution of the infant gut microbiome on the relationship between an exposures or treatments and downstream health outcomes. Studies now suggest that a healthy gut microbiome, possibly with probiotic supplementation, may prevent infection-related outcomes such as necrotizing enterocolitis and late-onset sepsis associated with preterm

birth [56]. The infant gut microbiome also may mediate the possible beneficial effect of farm exposure on school-age asthma[57]. Similarly, one study found that the infant gut microbiome mediated in the association between maternal pre-pregnancy overweight status and child overweight status between the ages of 1 and 3 years[58]. Mice and human studies also both suggest that the early gut microbiome may mediate the adverse effect of antibiotic exposure on later disease[59–61].

1.4 Statistical methods analyzing microbiome as a mediator

The primary approach to examining mediating mechanisms is through the classical three-step method for mediation analysis [62]. While several approaches exist to mediation analyses, the main ones in the literature are structured equation models [63,64] and the causal inference-based approach predicated on counterfactuals with a single mediator [65–67]. Methods for evaluating multiple mediators are beginning to emerge [68,69], including those to incorporate high dimensional mediators [70–81]. Modelling the microbiome is especially challenging due to the high-dimensional, compositional, and zero-inflated nature of microbiome data. Methods for microbiome mediation analysis include those examining the microbiome as a whole or those identifying specific mediating taxa[82–88]. However, few existing limited methods can account for exposure-mediator interactions and multiple types of outcomes (continuous and dichotomous).

1.5 Thesis Aims

This thesis aims to understand the three-way interplay among early-life exposures, the infant gut microbiome, and outcomes in infancy. We first investigated the association

between timing of rice cereal introduction and respiratory infections, respiratory symptoms, and allergy in a general population cohort of maternal-infant dyads. We then examined the association between the infant gut microbiome at 6 weeks of life and outcomes including respiratory infections, respiratory symptoms, wheeze, diarrhea, and vaccine response to PCP and TT in early childhood. Finally, we proposed a novel hypothesis test for microbiome mediation that accounts for the high dimensional and compositional nature of microbiome data and dichotomous outcomes.

Chapter 2: Infant infections, respiratory symptoms, and allergy in relation to timing of rice cereal introduction in a United States cohort

Published as Moroishi, Y., Signes-Pastor, A.J., Li, Z. et al. Infant infections, respiratory symptoms, and allergy in relation to timing of rice cereal introduction in a United States cohort. Sci Rep ***12,*** *4450 (2022).* https://doi.org/10.1038/s41598-022-08354-2.

Authors:

Yuka Moroishi, Antonio J. Signes-Pastor, Zhigang Li, Kathryn L. Cottingham, Brian P. Jackson, Tracy Punshon, Juliette Madan, Kari Nadeau, Jiang Gui, Margaret R. Karagas

Author Contributions:

YM conducted the statistical analysis and drafted the manuscript. AJSP helped conduct literature review. ZL helped design statistical approach and supervise the analysis. KLC contributed to the design of the rice intake questionnaire. BPJ And TP contributed to the exposure assessment. JM and KN assisted in the conduct and interpretation of the study. JG contributed to the statistical methods design and implementation. MRK designed the study, and oversaw the data collection.

2.1 Abstract

Rice products marketed in the USA, including baby rice cereal, contain inorganic arsenic, a putative immunotoxin. We sought to determine whether the timing of introduction of rice cereal in the first year of life influences occurrence of infections, respiratory symptoms, and allergy. Among 572 infants from the New Hampshire Birth Cohort Study, we used generalized estimating equation, adjusted for maternal smoking during pregnancy, marital status, education attainment, pre-pregnancy body mass index, maternal age at enrollment, infant birth weight, and breastfeeding history. Among 572 infants, each month earlier of introduction to rice cereal was associated with increased risks of subsequent upper respiratory tract infections (relative risk, RR = 1.04; 95% CI: 1.00–1.09); lower respiratory tract infections (RR = 1.19; 95% CI: 1.02–1.39); acute respiratory symptoms including wheeze, difficulty breathing, and cough (RR = 1.10; 95% CI: 1.00–1.22); fever requiring a prescription medicine (RR = 1.22; 95% CI: 1.02–1.45) and allergy diagnosed by a physician (RR = 1.20; 95% CI: 1.06–1.36). No clear associations were observed with gastrointestinal symptoms. Our findings suggest that introduction of rice cereal earlier may influence infants' susceptibility to respiratory infections and allergy.

2.2 Introduction

Early life is a critical period of immune system development and impacts health lifelong[89,90]. Infections remain the leading causes of mortality in children under five years

of age around the world[91]. Feeding practices, in particular breast feeding, are known to protect against infections and improve child health outcomes[7]; however, far less is known about the impact of diet during infants' transition to solid foods. Rice is an important dietary source of arsenic, including rice products commonly fed to infants as a first food and as snacks[92,19,93,94]. In flooded rice paddy fields, rice grains accumulate arsenic at rates about 10 times higher than that of other grains[19,95,96]. Additionally, arsenicals such as monosodium methanearsonate and disodium methanearsonate were used in pesticides and herbicides. Although these compounds are now mostly banned, residues remain in soil[95,97,98]. While rice cereal fortified with iron may be a good source of nutrients, concerns have been raised about this practice because of the arsenic content of rice-based products[99]. In previous studies from our cohort, infant urinary arsenic concentrations increased with consumption of rice products during infants transition to solid food[24], and at one year of age, infants fed rice products had elevated urinary concentrations of arsenic compared to those who were not fed these products[21].

Exposure to arsenic early in life has been specifically associated with an impaired immune response and increased risk of infection [22,100,101]. Infants are especially vulnerable to respiratory infections, in part due to their immature immune system [102]. Studies have reported associations between *in utero* arsenic exposure and a number of adverse outcomes including infant infections and respiratory outcomes among highly exposed populations in Bangladesh and among US infants [23,103,104]. These findings are supported by mechanistic evidence that *in utero* arsenic exposure may influence the epigenome of the placenta[105], immune cell profiles in newborn cord blood[106], and the

infant gut microbiome[107]. While epidemiologic data are lacking on allergy outcomes, maternal urinary arsenic concentrations during pregnancy were related to higher activated Th2 cells, which produce cytokines responsible for IgE production, a marker of allergic response[108–111]. In addition, there is evidence that early arsenic exposure influences childhood infections risk in highly exposed populations[112,113].

Despite health concerns[114,115], a regulatory limit for arsenic in infant rice cereal has not yet ratified in the USA. The European Union (EU) established a standard for inorganic arsenic in infant rice products to a maximum level of 100 μg/kg[116]. The US Food and Drug Administration (FDA) proposed the same guidance for infant rice cereal in 2016[117]. In 2018, the Governmental Accountability Office recommended that the FDA and US Department of Agriculture coordinate their efforts to identify contaminants in food including arsenic and establish a timeline for finalizing the guidance[118]. An action level has been set by the FDA for apple juice, but not other foods[119]. In 2006, the USA set the maximum contaminant level for inorganic arsenic in drinking water to be 10 μg/L[120], but evidence on the detrimental health impacts at even lower levels of exposure led to the reduction of the drinking water standard in certain states, including New Jersey[121] and New Hampshire[122], to 5 μg/L. In light of the vulnerability of infants to early life environmental exposures, we investigated the timing of introduction of rice cereal during their transition to solid food in first year of life and subsequent risk of infections, immune-related symptoms, and allergies as part of the New Hampshire Birth Cohort Study (NHBCS).

2.3 Methods

2.3.1 Study Design

Participants in this study include mother-infant dyads from the NHBCS. Pregnant women aged 18 to 45, receiving prenatal care at study clinics in New Hampshire, USA, were recruited starting in January 2009 as described previously[123,124]. The cohort includes only women who were living in the same household since their last menstrual period, not planning to move, living in a household served by a private water system, and with a singleton pregnancy. Participants completed surveys, including questions on sociodemographic factors, lifestyle such as smoking history, and pre-pregnancy body mass index (BMI), and infant birth characteristics were ascertained from a review of the delivery medical records. Home tap water samples were collected and analyzed by inductively coupled mass spectrometry to detect arsenic species[125]. The Committee for the Protection of Human Subjects at Dartmouth College approved all protocols, and participants provided written informed consent upon enrollment. All methods were performed in accordance with relevant guidelines and regulations.

2.3.2 Data Collection

Telephone interviews were conducted with caregivers when the infants turned 4, 8, and 12 months of age and at 6-month intervals thereafter. The survey asked whether or not their infant ever consumed rice cereal from birth and the day of the telephone interview (yes/no) and the month (or age in months) that rice cereal was introduced to their diet. Caregivers were asked whether their child had any infections, acute respiratory symptoms (e.g., wheeze, difficulty breathing, and cough), diarrhea or fever since the last time they

were interviewed. For positive responses, participants were asked to report if the condition lasted for more than 2 days, if the child saw a doctor, and if the child received prescription medicine for the condition. Participants were further asked whether their child had any known allergy (e.g., cats or dogs, antibiotics, dust, grass and plants, pollen, insect bites, peanuts, other nuts, eggs, and other foods), and for positive responses, if the allergy was diagnosed by a doctor.

2.3.3 Statistical Analysis

We examined rice cereal intake prior to the occurrence of infection. Specifically, we examined months since first introduction of rice cereal by 4 months of age on 8-month outcomes, months since first introduction of rice cereal by 8 months of age on 12-month outcomes, and months since first introduction of rice cereal by 12 months of age on 18-month outcomes. For example, we examined number of months since rice cereal consumption of a subject at 8 months, on occurrence of health outcome at 12months, which is the subsequent survey collection interval. For each interval (i.e., 4, 8, or 12 months), we computed the number of months since rice cereal was introduced and included an indicator variable of whether rice cereal was consumed in that interval (yes/no). We multiplied this indicator variable by number of months since rice cereal was introduced as our main predictor.

We then used Generalized Estimating Equation (GEE) with Poisson regression with AR(1) correlation structure and robust variance to assess the association between months since introduction of rice cereal exposures and repeated measures of longitudinal

outcomes[126]. Factors associated with both rice exposure, as determined from our data (Supplementary Table 2.7), and outcomes, considered a priori and used in previous studies[23], were considered potential confounders and included in our models. These included smoking during pregnancy (yes/no), maternal relationship status (married, single, separated/divorced), maternal education (≤ high school/GED, some college, college graduate, postgraduate schooling), maternal pre-pregnancy BMI, maternal age of enrollment (years), infant birth weight (grams), breastfeeding status as of four months (ever/never), and consumption of other solid food than rice cereal (ever/never) at each time point. Because water can be used to make rice cereal and is also a surrogate for *in utero* arsenic exposure, we also adjusted for total arsenic concentrations measured in household tap water samples (μg/L) in our analyses. For interpretability, we exponentiated the coefficient values to obtain relative risk (RR) and 95% confidence intervals (CI). Of the 572 participants included in the analyses, 482 participants (84.3%) had complete data for at least one time interval on rice cereal consumption and on subsequent health outcomes along with all potentially confounding variables considered. For the other 90 participants, we assumed for the values for potential confounders were missing at random using multiple imputation by chained equations and the predictive mean matching method to impute missing data[127]. All analyses were performed using R version 3.4.3 and functions *mice* and *geeglm* in packages 'mice' and 'geepack'.

2.4 Results

2.4.1 Baseline Characteristics

Of the 1760 pregnancies enrolled in the NHBCS as of October 2017, a total of 983 infants had complete follow-up data up to at least age of 8 months. After removing missing values on rice cereal consumption (54 missing) and subsequent health outcome information (357 missing), the final dataset contained 572 infants (Supplementary Figure 2.1). We found that the characteristics of the 411 subjects excluded from our analyses were generally similar to that of included subjects, with the exception of marital status (Supplementary Table 2.1). Our study group included a roughly equal distribution of male (54%) and female (46%) infants (Table 2.1). Among infants who were introduced to rice cereal in the first year of life, the average age at introduction was 5.2 months (SD: 1.3 months) (Supplementary Table 2.2). At the 4 month, 8 month, and 12 month time periods, rice cereal was consumed in 11.7%, 69.6%, and 68.6% of infants respectively (Supplementary Table 2.2). Overall, 96.5% of infants were reporting as having at least one infection or symptom of any duration reported up to age 18 months, 91.4% having at least one lasting 2 or more days, 65.2% having at least one involving a doctor's visit, and 52.3% having at least one resulting in a prescription medication (Supplementary Table 2.2). For allergies, 13.5% of infants were reporting as having at least one allergy, and 7.9% having at least one diagnosed by a doctor (Supplementary Table 2.2). Sample sizes and proportions of each outcome for each follow-up interval are reported in Supplementary Table 2.3. Household tap water arsenic concentrations were generally low, with a mean 2.2 μg/L (SD: 7.1; range: 0.0 to 92.3), but with 11.1% of the study population having levels above the New Hampshire drinking water standard of 5 μg/L (Table 2.1).

Table 2. 1: Selected Characteristics of Mothers and Infants (N = 572) in the New Hampshire Birth Cohort Study Followed to Age 18 Months

Variable	Sample Size	Mean (SD) or No. (%)
Maternal Characteristics		
Smoking during any trimester of pregnancy, No. (%)	552	
Yes		61 (11.1)
No		491 (88.9)
Relationship status, No. (%)	538	
Married		486 (90.3)
Single		43 (8.0)
Separated/Divorced		9 (1.7)
Highest level of educational attainment, No. (%)	537	
≤high school/GED		52 (9.7)[a]
Some college		90 (16.8)[a]
College graduate		214 (39.9)[a]
Postgraduate schooling		181(33.7)[a]
BMI before pregnancy (kg/m^2), mean (SD)	560	26.1 (5.7)
Age at enrollment (years), mean (SD)	572	31.9 (4.8)
Arsenic in water (µg/L), mean (SD)	552	2.2 (7.1)
Water Arsenic > 5 µg/L, No. (%)	552	61 (11.1)
Infant Characteristics		
Sex, No. (%)	572	
Male		310 (54.2)
Female		262 (45.8)
Birth weight (g), mean (SD)	555	3416.7 (522.8)
Ever breast fed at 4 months, No. (%)	537	
Yes		513 (95.5)
No		24 (4.5)
Other solid food consumption at 4 months, No. (%) [b]	358	
Yes		14 (3.9)
No		344 (96.1)
Other solid food consumption at 8 months, No. (%) [b]	437	
Yes		120 (27.5)
No		317 (72.5)
Other solid food consumption at 12 months, No. (%) [b]	544	
Yes		137 (25.2)

No		407 (74.8)

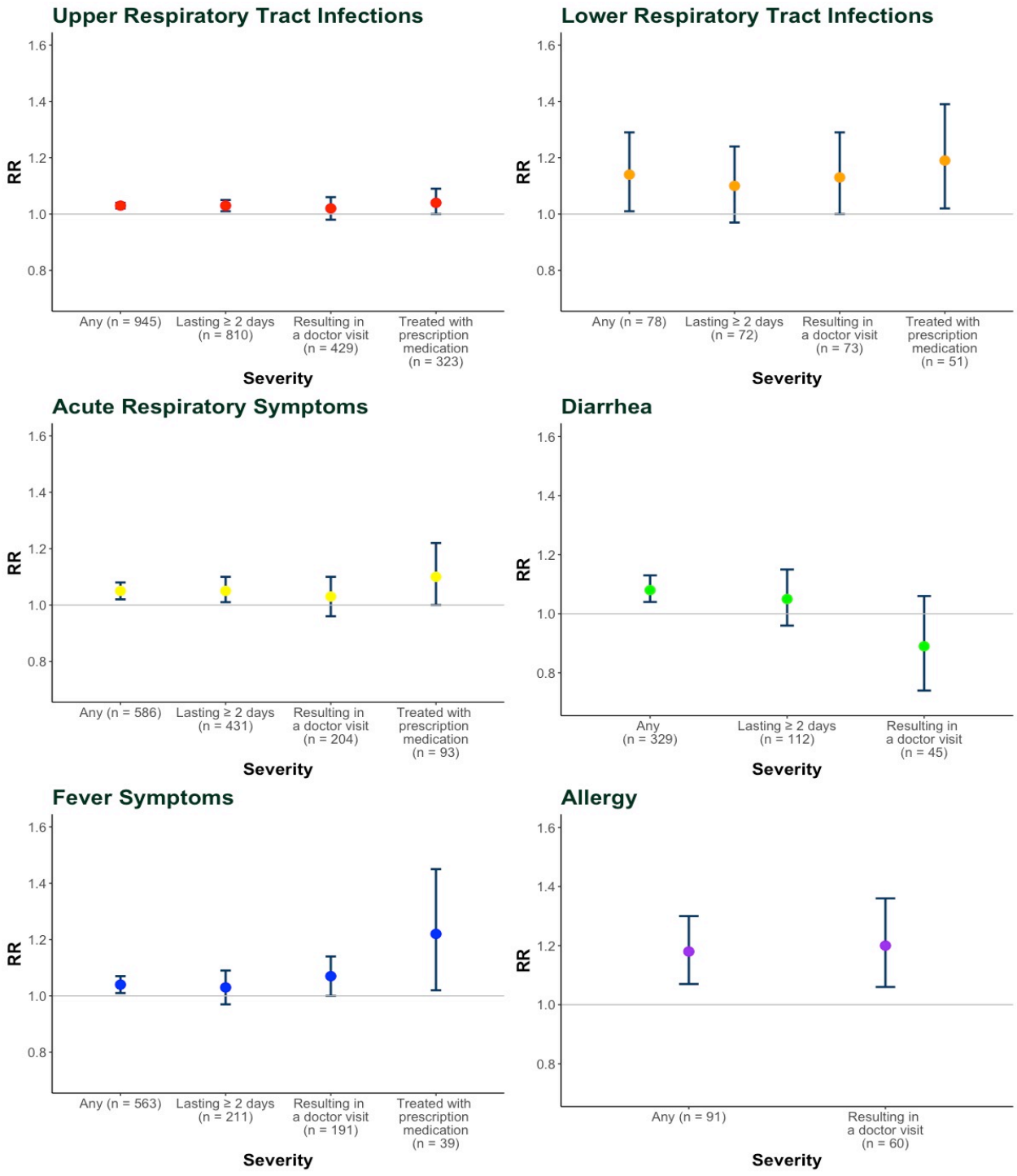
Abbreviations: *SD*, standard deviation; *No*., number; *BMI,* body mass index.
Of the 572 participants included in the analyses, 482 participants (84.3%) had data at least one time-period of rice cereal consumption, health outcome data for a subsequent time period, and complete data on confounders. Smoking during pregnancy was missing for 20 (3.50%) mothers, relationship status was missing for 34 (5.94%) mothers, education was missing for 35 (6.12%) mothers, and pre-pregnancy BMI was missing for 12 (2.10%) mothers. Birth weight was missing for 17 (2.97%) and breast-feeding status was missing for 59 (10.31%) infants. Other solid food consumption at 8 months was missing for 1 (0.23%) infant. A total of 20 (3.50%) participants had missing data for arsenic species in water.
[a] Percentages do not sum to 100 due to rounding
[b] Percentage calculated using different sample sizes due to missing values. Sample sizes were 321, 373, and 464 for 4 months, 8 months, and 12 months respectively.

2.4.2 Rice Cereal and Infections, Respiratory Symptoms, and Allergy

In our GEE analysis, earlier introduction of rice cereal was associated with increased risks of lower respiratory tract infections (i.e. bronchitis, pneumonia, bronchiolitis, whooping cough, and respiratory syncytial virus), respiratory symptoms, fever, and allergies, and to a lesser extent upper respiratory tract infections (i.e. runny stuffed nose, eye infection, ear infection, severe flu, sinus infection, strep throat, and laryngitis), but not gastrointestinal symptoms (Figure 2.1). While the magnitudes of the associations did not differ greatly across the variables for a given outcome and the confidence intervals overlapped, there was a tendency for the risk ratios to be higher for outcomes involving a health care provider visit or a medication prescribed (Figure 2.1, Supplementary Table 2.4). Relative risk estimates for upper and lower respiratory tract infections requiring a prescription medicine increased by 4% (RR = 1.04; 95% CI: 1.00-1.09) and by 19% (RR=1.19; 95% CI: 1.02-1.39) for each month earlier that rice cereal was introduced. A 10% increase in the relative risk of acute respiratory symptoms requiring a prescription medicine (RR = 1.10; 95% CI: 1.00-1.22) and 22% increase of fever symptoms requiring

prescription medicine (RR = 1.22; 95% CI: 1.02-1.45) were observed for each month earlier that rice cereal was introduced. For reported allergies diagnosed by a doctor, the relative risk estimate was 20% higher for each month earlier that rice cereal was introduced (RR = 1.20; 95% CI: 1.06-1.36), and for this outcome, the relative risk estimate was similar to any reported allergy. We did not observe any consistent associations with diarrhea, including symptoms requiring a doctor's visit (RR = 0.89; CI, 0.74-1.06). Only three cases of diarrhea had a medication prescribed, so this outcome was not included in the GEE analysis. Risk ratio estimates and confidence intervals for covariates included in our models are provided in Supplementary Table 2.5. Results for crude analyses are provided in Supplementary Table 2.6.

Figure 2. 1: Adjusted Risk Ratios (95% CIs) for Each Month Earlier of Introduction of Rice Cereal According to Outcome Severity, N = 572 infants.



Abbreviations: *RR* risk ratio.
RR indicates risk ratio for health outcomes according to reported severity. The circles indicate RR at each severity level for each outcome. The lines indicate the 95% confidence intervals. The grey line represents the null RR=1. Number of total outcome cases in three repeated measures is denoted by n. Number of observed outcomes can be higher due to repeated events among time points for each infant. Sample size N = 571 for fever analyses.

2.5 Discussion

In our prospective cohort study, we found that earlier introduction of rice cereal to an infant's diet was associated with higher risks of both upper and lower respiratory tract infections, respiratory symptoms, fever, and allergies. These associations were slightly stronger for what may have been more severe outcomes of lower respiratory tract infection, respiratory symptoms, and fever, i.e., those characterized by having a medication prescribed.

Infections remain the most important cause of morbidity in young children, and allergic and atopy diseases are becoming more widespread[91]. In the USA, an estimated 42.8% of infant hospitalizations in 2003 were due to infections[128]. Of these, 59.0% were due to lower respiratory tract infections and 6.5% to upper respiratory tract infections[128]. In 2013, the Centers for Disease Control and Prevention (CDC) noted an increasing trend in childhood food and skin allergies from 1997 to 2011[129]. In a 2017 CDC survey, 13% of children under the age of 18 years had been told they had asthma, 11% a respiratory allergy, 6.5% a food allergy, and 13.5% a skin allergy[130]. Thus, efforts to reduce infection and allergy prevalence in infants is of critical public health importance.

Early feeding practices play a critical role in the developing immune system of infants[7], but there are limited studies of the impact of infants' transition to solid foods. In a large prospective study from Dundee, Scotland (N = 545), early introduction of solid food was related to an increased risk of infant wheeze, but not other respiratory illnesses[131]. One prospective cohort study from the United Kingdom found infant introduction to solid foods before 4 months of age was associated with higher odds of any diarrhea compared to those introduced after 4 months of age (N = 615)[132]. Introduction of wheat after 6 months of age compared to before or equal to 6 months of age was associated with a reduced risk of wheat allergies in a longitudinal birth cohort from Denver (N = 1612)[133]. While research linking infant rice cereal exposure to later health outcomes is lacking, our results align with previous studies that observed increased risks of infection with arsenic exposure in early childhood. A Bangladeshi cohort study of children aged 7- 17 years who were exposed to high levels of arsenic *in utero* and early childhood from contaminated drinking water and sex- and age-matched controls without such exposures observed increased respiratory symptoms such as wheezing and shortness of breath (N = 650) in the high arsenic exposure group[112]. In a separate case control study from Bangladesh of children aged 28 days to 59 months who were hospitalized with severe and very severe pneumonia and age-matched controls, the odds of pneumonia were elevated among children with higher urinary arsenic concentrations measured both during hospitalization and at the convalescent period (30 days after) (N = 449)[113]. Other cohort studies from Bangladesh and from the USA have also associated childhood infections and diarrhea in relation to *in utero* arsenic exposure[23,103,104].

Rice products are a well-recognized route of exposure to arsenic. Arsenic exposure has been shown to increase risk of infections and other diseases, and emerging evidence points to the toxicological effects of arsenic on immune function[22,134]. The World Health Organization established a guideline for arsenic in drinking water of 10 μg/L in 1993 and had acknowledged arsenic contamination in rice and rice products as a public health concern[135,136]. In a 2014 EU report, concentrations of inorganic arsenic ranged from 56 μg/kg to 268 μg/kg for infant rice products[137]. A similar report from the USA covering 2012 to 2016 reported inorganic arsenic concentrations ranging from 21 μg/kg to 151 μg/kg for infant white rice cereal and from 30 μg/kg to 254 μg/kg for infant brown rice cereal [138]. In our cohort, 80% of infants were introduced to rice cereal in the first year of life, and more than half of our infants were eating rice products at one year of age[21]. Further, urinary arsenic concentrations increased with the number of rice and rice product servings[21]. The American Academy of Pediatrics has raised awareness about arsenic exposure from feeding infant rice products and recommends feeding infants a variety of foods with a variety of textures[94,139].

Our study has a number of strengths, but also has limitations that need to be noted. Our study benefitted from the availability of prospective cohort data of carefully collected repeated measurements of infection occurrences, respiratory symptoms, diarrhea, and allergies; timing of introduction of rice cereal; and a broad range of potential confounding factors. Among our main limitations was our inability to quantify the concentrations of arsenic to which infants were exposed through rice cereal. This is in part because the concentrations of arsenic in rice depends on a number of factors

including genotype, cultivation, and irrigation techniques, and concentration can be altered by cooking techniques[140,141]. Despite the heterogeneity of arsenic concentrations in rice, we previously found rice cereal to be a contributor to arsenic exposure among our infants [24]. In our analyses, we also adjusted for household water arsenic concentration along with indicator variables for breastfeeding and consumption of other solid foods. Participant recall is a potential source of bias. Efforts to minimize non-differential misclassification were made by including questions on duration of the illness and asking whether the infant saw a doctor or was prescribed medicine for their condition. Stronger associations were found with outcomes involving medical care, which would be expected to have higher validity and reflect greater severity of illness. Furthermore, self-reporting of allergies that were not medically confirmed was another potential source of misclassification. We have found that responses from caregiver responses for infections and symptoms that involved a doctor visit in our cohort tend to be at least 80% concordant with their pediatric medical records (unpublished data). Our findings of an increased risk of any reported diarrhea associated with earlier introduction rice consumption should thus be interpreted with caution as the relative risk estimates were not consistently elevated for diarrhea lasting 2 or more days or associated with a doctor's visit. The possibility of unmeasured confounding also cannot be excluded; however, we assessed the potential confounding of factors previously found to be related to our outcomes of interest and found that certain factors such as family history of allergy, day care attendance, and parity were not associated with timing of rice cereal introduction. The price of rice cereal in our cohort may be indicative of socio-economic status. Although we did control for maternal marital status and highest level of education

attainment, residual confounding remains a possibility. We inspected the costs of infant cereals from five major grocery stores in a region of the state where recruitment took place. We did not find that the costs of infant rice-based cereal products differed appreciably from that of other grain cereals. Lastly, we computed 95% confidence intervals for our analyses, which do not account for multiple comparisons.

In conclusion, our findings suggest that earlier introduction of rice cereal may increase an infant's risk of infections, respiratory symptoms, and allergies. The widespread occurrence of these outcomes in young children and use of rice cereal as a first food underscores the importance of considering the types and timing of foods introduced when providing dietary recommendations for infants.

Chapter 3: The infant gut microbiome in relation to subsequent risk of respiratory infections and symptoms among vaginally and cesarean delivered infants

Manuscript currently under review

Authors:

Yuka Moroishi, Jiang Gui, Anne G. Hoen, Hilary G. Morrison, Emily R. Baker, Kari Nadeau, Hongzhe Li, Zhigang Li, Juliette Madan, Margaret R. Karagas

Author Contributions:

YM drafted the manuscript and conducted all statistical analyses. JG, AGH, HL and ZL helped design the statistical methodology. HGM conducted raw data processing. AGH, ERB, KN, JM and MRK contributed to designing the study, overseeing data collection and processing, and interpreting results.

3.1 Abstract

Emerging evidence points to a critical role of the developing gut microbiome in immune maturation and infant health; however, prospective studies are lacking. We examined the occurrence of infections and associated symptoms during the first year of life in relation to the infant gut microbiome at six weeks of age using bacterial 16S rRNA V4-V5 gene sequencing (N = 465) and shotgun metagenomics (N = 185). Higher infant gut microbiota alpha diversity was associated with an increased risk of infections or respiratory symptoms treated with a prescription medicine, and specifically upper respiratory tract infections. Among vaginally delivered infants, a higher alpha diversity was associated with an increased risk of all-cause wheezing treated with a prescription medicine and diarrhea involving a visit to a health care provider. Positive associations were specifically observed with *Veillonella* species among all deliveries and *Haemophilus influenzae* among cesarean-delivered infants. Our findings suggest that intestinal microbial diversity and the relative abundance of key taxa in early infancy may influence susceptibility to respiratory infection, wheezing, and diarrhea.

3.2 Introduction

Infections remain the leading causes of mortality in infants globally[142]. The human gut microbiome is becoming increasingly recognized for its critical role in immune function and the inflammatory response[47,143]. A bidirectional relationship emerges following birth whereby the gut microbiome aids the maturation of the immune system and the immune system regulates host-microbe symbiosis[25,26]. The impacts of perturbing these intricate relationships are evident in high-risk infants. For example, among infants with cystic

fibrosis, the composition of the gut microbiome is a determinant of colonization with opportunistic pathobionts[144,145]. Likewise, in preterm infants, the gut microbiome is associated with fatal occurrences of necrotizing enterocolitis and infection[143,146,147]. Factors driving the establishment of the gut microbiome, including delivery mode and breast feeding[47,124,148–150], have also been related to the risk of infections[3,4,47,151–153]. Furthermore, the use of antibiotics during pregnancy, which has been found to influence the gut microbiome of offspring[154–156], increased the risk of infant infection-related hospitalizations[15]. Encouraging results from probiotic trials suggest health benefits from altering the gut microbiome, including an enhanced immune response to pathogens[157,158]. While studies have found possible links between early gut microbiome composition and infant infection[47,159], few prospective studies have been conducted, particularly in the general population.

We report on gut microbiome diversity and composition among infants during the critical period of early immune training and the subsequent occurrence of respiratory infections and symptoms, such as wheezing and diarrhea, in the first year of life as part of a prospective study of a cohort of pregnant women and their offspring from the general population in New Hampshire. Here, we measured the fecal microbiome to measure the gut microbiome. Wheeze and diarrhea outcomes for this study included those of any cause. Based on amplicon sequence variant (ASV) data generated from 16S rRNA sequencing, higher alpha diversity at six weeks of age was associated with having an additional respiratory infection or symptom of respiratory infection requiring a prescription medicine, with associations varying by delivery mode. Using next-

generation sequencing (NGS), shotgun metagenomics, *Veillonella* in all deliveries and *Haemophilus* in cesarean deliveries were among the species in six-week stool identified as being related to an additional subsequent respiratory infection or symptom of respiratory infection requiring a prescription medicine during an infant's first year of life.

3.3 Methods

3.3.1 Study Population

Participants included mother-infant dyads from the NHBCS from whom we obtained infant stool samples at approximately six weeks of age. Pregnant women aged 18 to 45 were recruited from prenatal clinics in New Hampshire, USA, starting in January 2009, as described previously[123]. Women who were living in the same household served by a private water system since their last menstrual period, had no plans to move, and had a singleton pregnancy were included in the cohort. Participants completed surveys on infant lifestyle questions such as feeding mode, solid food introduction, and daycare. Infant birth characteristics were ascertained from newborn medical records, and maternal characteristics were abstracted from prenatal and delivery records, including age at enrollment, prenatal use of antibiotics, and prepregnancy body mass index in kilograms per meter squared. The Committee for the Protection of Human Subjects at Dartmouth College approved all protocols, and participants provided written informed consent upon enrollment.

3.3.2 Ascertainment of Infant Health Outcomes

Telephone interviews were conducted with infants' caregivers in the first year of life, i.e., when infants turned approximately 4 months, 8 months, and 12 months of age. Caregivers were asked whether their child had any upper RTIs or associated symptoms (e.g., runny nose, stuffy nose, eye infection, ear infection, influenza, sinus infection, pharyngitis, or laryngitis), lower RTIs (e.g., bronchitis, pneumonia, bronchiolitis (including respiratory syncytial virus (RSV)), or whooping cough), acute respiratory symptoms (e.g., difficulty breathing, wheezing, fever, or cough), or diarrhea since the previous interview. For each positive response, participants were then asked whether the condition lasted more than 2 days, whether the child saw a physician, and whether the child received any prescription medications for the condition.

3.3.3 Stool Sample Collection, DNA Extraction, Sequencing, and Profiling

We measured the fecal microbiome of infant stool as a measure of the infant gut microbiome. Infant stool samples were collected at regularly scheduled ~6-week postpartum follow-up appointments as described previously[124,160]. Samples were aliquoted and frozen at -80 ^{o}C within 24 hours of receipt. A Zymo DNA extraction kit (Zymo Research) was used for DNA extraction from thawed samples, and an OD260/280 nanodrop was used to measure sample quality and purity. The V4-V5 hypervariable region of the bacterial 16S rRNA gene was sequenced using Illumina MiSeq at the Marine Biological Laboratory in Woods Hole, Massachusetts. ASVs were then inferred using DADA2[161], and taxonomies were assigned using the SILVA database[162]. Quality control measures were conducted as described previously[124]. A subset of stool samples was also sequenced with NGS and shotgun metagenomics sequencing as previously

described[156]. Extracted DNA samples were sheared to a mean insert size of 400 bp using a Covaris S220 focused ultrasonicator. The sequencing libraries were constructed using Nugen's Ovation Ultralow V2 protocol, and samples were sequenced using Illumina NextSeq. DNA reads were merged and trimmed using KneadData[163] for quality control before species-level taxonomic profiles were generated using Metaphlan2[164].

3.3.4 Statistical Analysis

We examined the association between gut microbiome composition and health outcomes ascertained during interval interviews over the subjects' first year of life. For our analyses, we examined the total number of reported outcomes, specifically upper RTIs and lower RTIs, as well as symptoms such as wheezing with a reported visit to a physician and treatment with a prescription medication. Diarrhea is not typically treated with prescription medications in infants; therefore, we focused the analyses on reports of diarrhea that involved a doctor visit. We imputed missing outcomes if the caregiver completed the interview but a specific question was unanswered using multiple imputation by chained equations and the predictive mean matching method.

For models using 16S data, we aggregated ASVs to the genus level and calculated alpha diversity on read counts per genus using the inverse Simpson index. We then used a GEE for repeated measures with Poisson regression and AR(1) correlation structure to assess the association between $\log_2$-transformed 16S-based genus-level alpha diversity and each of the outcomes of interest. For models using metagenomics species data, we calculated the $\log_2$-transformed relative abundance using a pseudocount of 5×10^{-20} for zero values. We also used a GEE for repeated measures with Poisson regression and

AR(1) correlation structure to estimate relationships with species present in at least 10% of subjects. We applied an FDR threshold of 0.1 to adjust for multiple testing[165]. Factors associated with both the gut microbiome and health outcomes were considered potential confounders and included in all GEE analyses. These confounders included maternal prepregnancy BMI (kg/m^2), delivery mode (vaginal/cesarean), infant sex (male/female), breast feeding at six weeks (exclusively breastfed/mixed fed or exclusively formula fed), antibiotic use during pregnancy (yes/no), and gestational age (complete weeks). We also conducted stratified analyses by delivery type for both alpha diversity using 16S data and microbial species based on metagenomics data. Due to sample size limitations, we conducted stratified analyses only for species-specific analyses on all outcomes combined and for upper RTIs.

For interpretability, we exponentiated the coefficient values to obtain relative risk (RR) and 95% confidence intervals (CIs). Of the 465 participants included in the 16S analyses, 391 participants (84.1%) had complete data for all potentially confounding variables. Of the 185 participants included in the metagenomics analyses, 160 participants (86.5%) had complete data for all potentially confounding variables. We assumed missing confounder entries were missing at random and used multiple imputation by chained equations and the predictive mean matching method to impute missing data. All analyses were performed using R version 3.4.3 using the functions *diversity*, *mice*, and *geeglm* in the 'vegan', 'mice', and 'geepack' packages.

3.4 Results

3.4.1 Baseline Characteristics

As of July 2019, we had completed 16S rRNA V4-V5 hypervariable region gene sequencing on 513 infant stool samples and whole-genome metagenomics sequencing on 202 infant stool samples collected at approximately six weeks of age. After removing infants for whom health information was unavailable in telephone surveys, our analysis included 465 infants with 16S data and 185 infants with metagenomics data. Our study population had an approximately equal distribution of male (53.4%) and female infants (46.6%) (Table 3.1). Nearly half of the infants (56.2%) had been exclusively breastfed at approximately six weeks of age, and approximately one-fifth of mothers (18.5%) had reported antibiotic use during pregnancy (Table 3.1). Cesarean section deliveries accounted for one-third of deliveries (30.3%) (Table 3.1). The five most common genera in our 16S data were *Escherichia/Shigella*, *Bacteroides*, *Bifidobacterium*, *Klebsiella*, and *Enterococcus* (Supplementary Figure 3.1). The five most common species in our metagenomics data were *Bifidobacterium longum*, unclassified *Escherichia* species, *Escherichia coli*, *Bifidobacterium breve*, and *Gemella haemolysans* (Supplementary Figure 3.1). The numbers of each respiratory infection and symptom at each age are provided in Supplementary Table 3.1 and Supplementary Table 3.2.

Table 3. 1: Selected Baseline Characteristics of Mothers and Infants in the New Hampshire Birth Cohort Study

Abbreviations: SD, standard deviation; No., frequency; kg, kilograms; m, meters.

Variable	16S V4-V5 rRNA		Metagenomics	
	Sample Size	Mean (SD) or No. (%)	Sample Size	Mean (SD) or No. (%)
Maternal Characteristics				
Age at enrollment, mean (SD), years	465	31.9 (4.6)	185	31.9 (4.3)
Body Mass Index before pregnancy, mean (SD), kg/m^2	462	25.8 (5.9)	185	25.7 (5.7)
Parity, No. (%)	461		184	
0		217 (47.1)		92 (50.0)
1		166 (36.0)		60 (32.6)
2+		78 (16.9)		32 (17.4)
Antibiotic use during pregnancy, No. (%)	427		173	
Yes		79 (18.5)		36 (20.8)
No		348 (81.5)		137 (79.2)
Infant Characteristics				
Delivery mode, No. (%)	465		185	
Vaginal		339 (72.9)		129 (69.7)
Cesarean		126 (27.1)		56 (30.3)
Infant sex, No. (%)	464		185	
Male		248 (53.4)		107 (57.8)
Female		216 (46.6)		78 (42.2)
Breast feeding at six weeks, No. (%)	427		171	
Exclusively breast fed		240 (56.2)		92 (53.8)
Mixed fed or exclusive formula fed		187(43.7)		79 (46.2)
Gestational age, Mean (SD), weeks	465	39.1 (1.6)	185	39.0 (1.7)

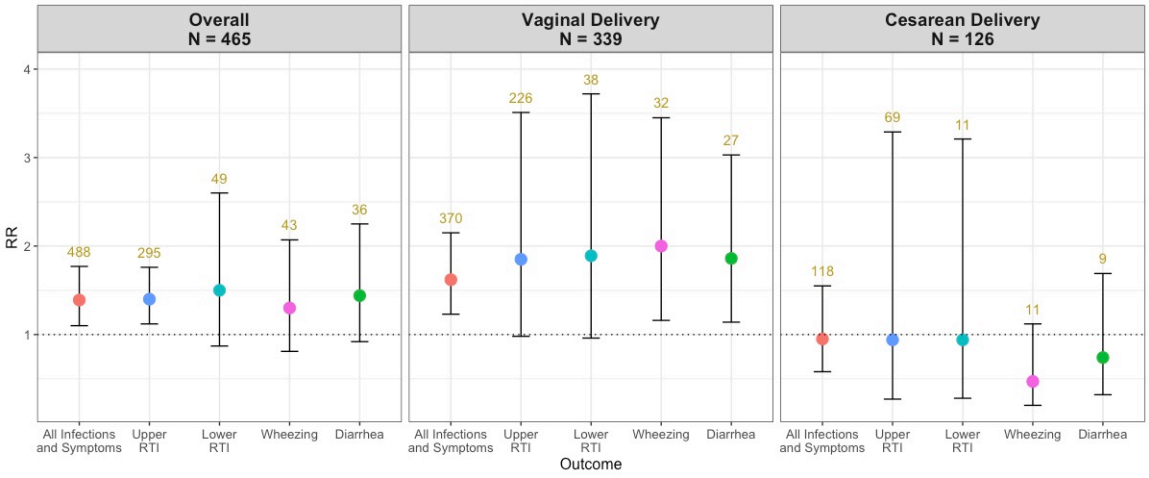
Of the 465 mothers included in the 16S analyses, maternal BMI was missing for 3 mothers. Parity was missing for 4 mothers, and antibiotic use during pregnancy was missing for 38 mothers.

Of the 465 infants included in the 16S analyses, infant sex was missing for 1 infant, and feeding type was missing for 38 infants. Of the 185 mothers included in the metagenomics analyses, parity was missing for 1 mother, and antibiotic use during pregnancy was missing for 12 mothers. Of the 185 infants included in the metagenomics analyses, feeding type was missing for 14 infants.

3.4.2 16S V4-V5 rRNA Gene: Alpha Diversity

Associations were determined via the Wald test in generalized estimating equation (GEE) analyses with a p-value threshold of 0.05. Alpha diversity was positively associated with the occurrence of any respiratory infection or symptom of infection, which included upper respiratory tract infections (RTIs), lower RTIs, and acute respiratory symptoms. Upper RTI outcomes were specifically associated. Each doubling in alpha diversity was associated with a 39% increase in having an additional respiratory infection or symptom of respiratory infection (RR = 1.39, 95% CI: 1.1-1.77) and a 40% increase in an additional upper RTI (RR = 1.40, 95% CI: 1.12-1.76) (Figure 3.1, Supplementary Table 3.3). Among vaginally delivered infants, a doubling of alpha diversity was associated with a 62% increase in in having an additional respiratory infection or symptom of respiratory infection (RR = 1.62, 95% CI: 1.23-2.15) (Figure 3.1, Supplementary Table 3.3). A doubling of alpha diversity was associated with a doubling of the risk of wheezing for which a medication was prescribed (RR = 2.00, 95% CI: 1.16-3.45) and an 86% increase in diarrhea requiring a doctor visit (RR = 1.86, 95% CI: 1.14-3.03) among vaginally delivered infants (Figure 3.1, Supplementary Table 3.3). We did not observe consistent associations among cesarean-delivered infants.

Figure 3. 1**:** Dot and Whisker Plots of Adjusted Relative Risk Estimates and 95% Confidence Intervals from GEE Analysis of Infant 6-Week-old Stool 16S V4-V5 rRNA Sequencing Alpha Diversity and Infections and Symptoms of Infection over the First Year of Life



Abbreviations: GEE, generalized estimating equation; N, sample size; RR, relative risk; RTI, respiratory tract infection. Overall GEE adjusted for maternal BMI, delivery type, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. GEE stratified by delivery mode (vaginal and cesarean) adjusted for maternal BMI, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. Points represent relative risk, and vertical lines above and below points represent upper and lower confidence bands. Relative risk estimates represent an increased risk of having an additional infection or symptom of infection or an increased risk of experiencing wheezing or diarrhea with each doubling of the inverse Simpson index. Upper RTI, lower RTI, and wheezing outcomes are those diagnosed by a physician for which a medication was prescribed. Diarrhea outcomes are those diagnosed by a physician for whom no medication was prescribed. Numbers above upper confidence bands indicate the total number of outcomes, which may be greater than N due to repeated measures. Sample sizes were N = 464 for overall and N = 125 for cesarean delivery for diarrhea analyses due to missing data.

3.4.3 Metagenomics: Species-Level

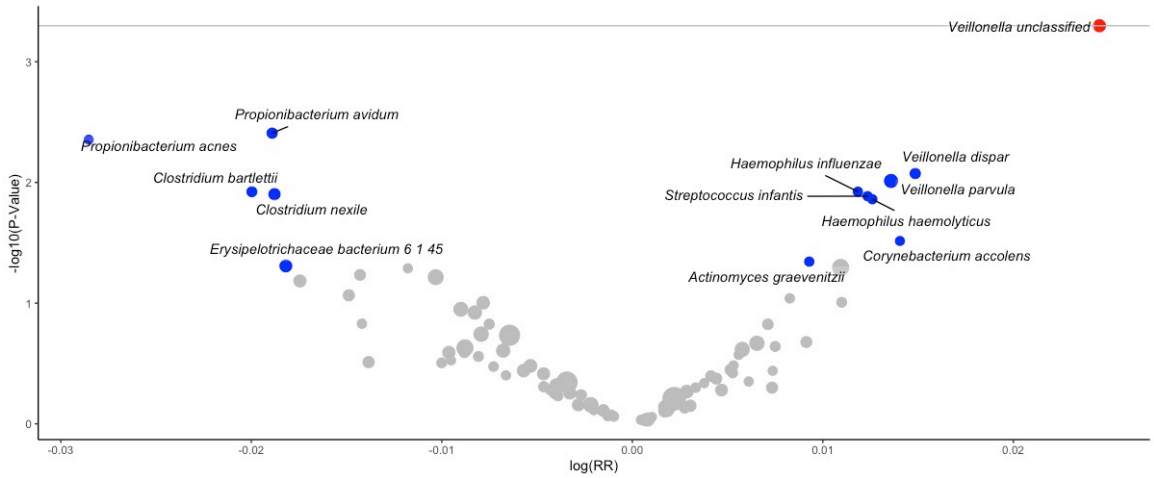
In the GEE of metagenomics species data, the doubling of the relative abundance of *Veillonella* unclassified was positively associated with having an additional respiratory infection or symptom of respiratory infection in the first year of life (RR = 1.02; 95% CI: 1.01-1.04) (Figure 3.2.1). In examining specific outcomes, we found that diarrhea was

positively associated with the relative abundance of *Streptococcus peroris* and negatively associated with the relative abundance of *Streptococcus salivarius* (Figure 3.3).

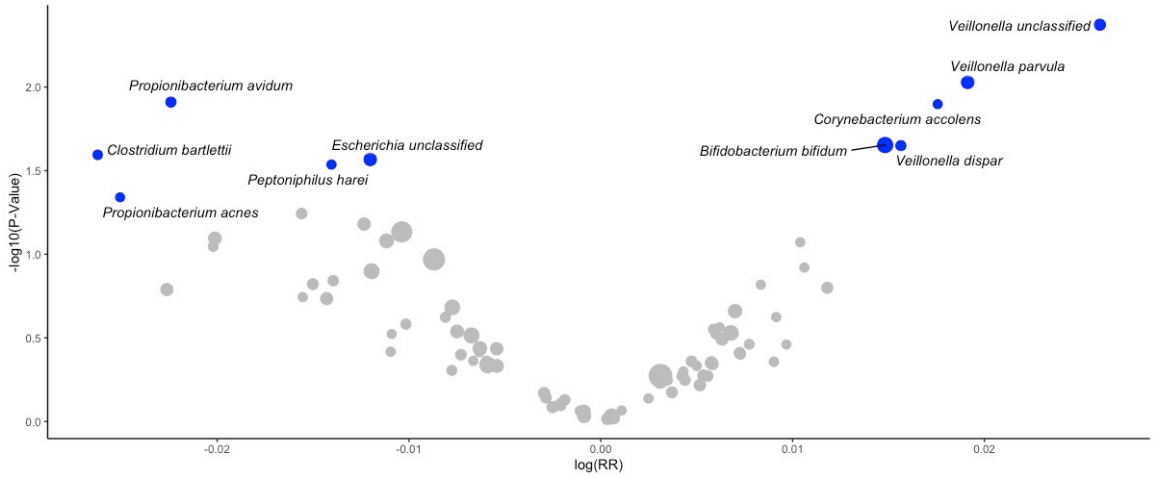
Stratified by delivery mode, we found that having an additional respiratory infection or symptom of respiratory infection was positively associated with *Haemophilus influenzae* among cesarean-delivered infants (RR = 1.02; 95% CI: 1.01-1.04) (Figure 3.2.3). *Veillonella parvula*, *Corynebacterium pseudodiphtheriticum*, and *Corynebacterium pseudodiphtheriticum* were positively associated, while *Clostridium butyricum* and *Coprobacillus unclassified* were negatively associated, with a risk of an additional upper RTI among infants delivered by cesarean section (Figure 3.3).

Figure 3. 2**:** Volcano Plots of GEE Adjusted Relative Risk Estimates of the Number of Infections and Symptoms over the First Year of Life in Relation to 6-Week Metagenomics Species Relative Abundance

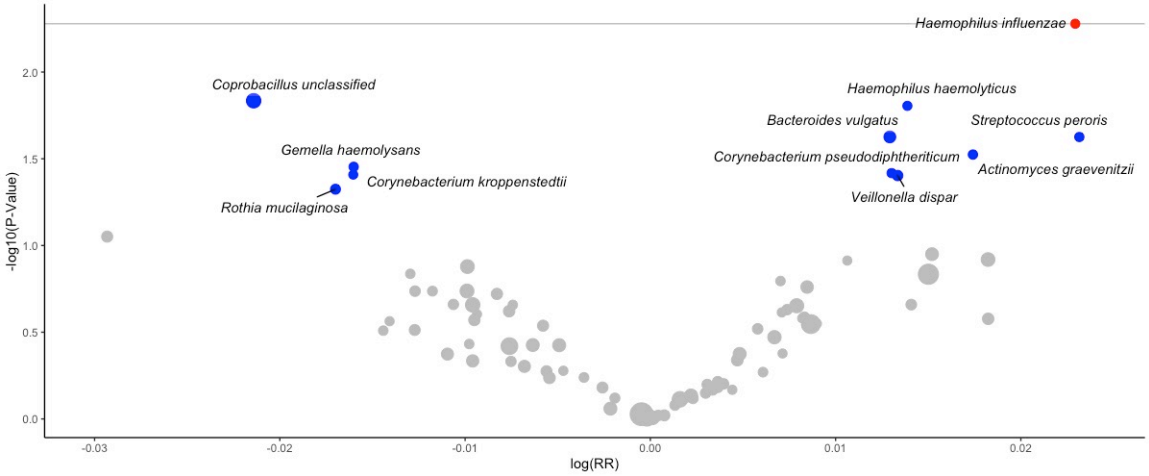
1. Overall (N = 185)



2. Vaginal Delivery (N = 129)

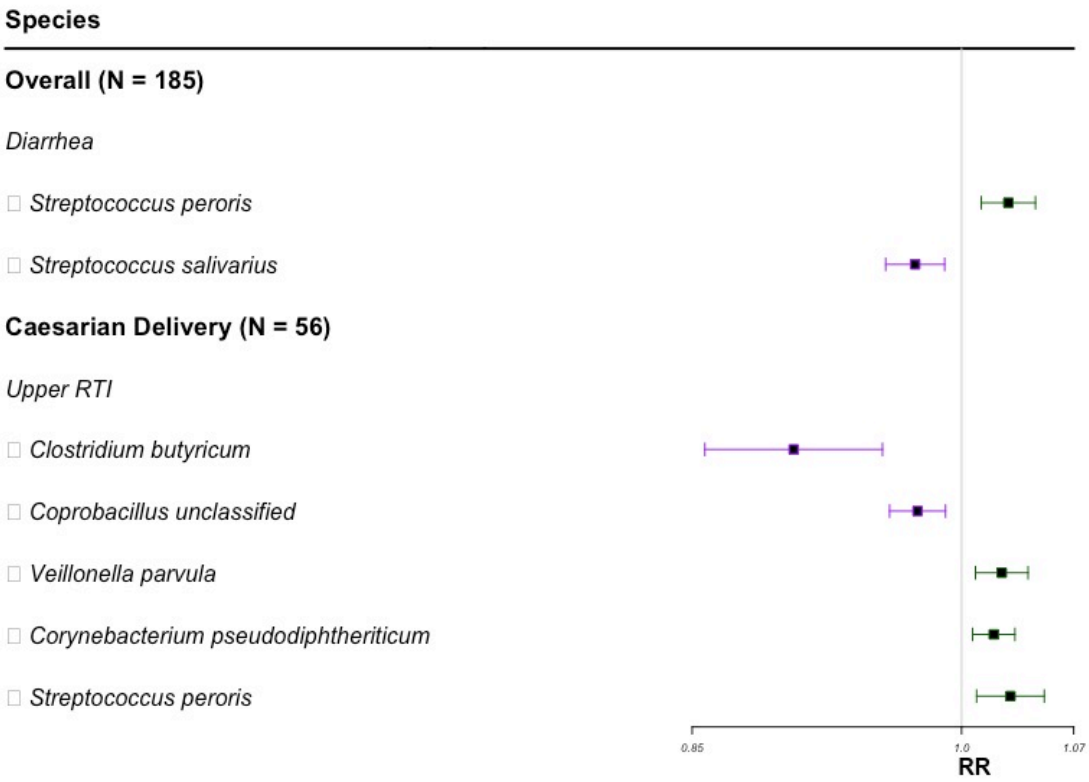


3. Cesarean Delivery (N = 56)



Abbreviations: GEE, generalized estimating equation; N, sample size; RR, relative risk.
Estimates shown for taxa prevalent in over 10% of subjects. The gray line represents a log10-transformed FDR threshold of 0.1. Blue points indicate statistically significant taxa at $\alpha = 0.05$. Red points indicate taxa selected by FDR correction. The size of the points is scaled by relative abundance. Relative risk estimates represent an increased risk of having an additional infections or symptoms of infection with each doubling of relative abundance. **a.** Volcano plot of unstratified GEE adjusted for maternal BMI, delivery type, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. **b.** Volcano plot of vaginal deliveries adjusted for maternal BMI, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. **c.** Volcano plot of cesarean deliveries adjusted for maternal BMI, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. Three taxa were removed due to high RRs and low p-values.

Figure 3. 3: Forest Plot of Metagenomics Species Associated with the Number of Infections and Symptoms of Infections in the First Year of Life



Abbreviations: N, sample size; RTI, respiratory tract infection; RR, relative risk.
Species selected by FDR correction presented in the forest plot. Species for vaginal deliveries did not meet the FDR threshold of 0.1. Squares represent RR, and horizontal lines represent 95% confidence intervals. Green represents a positive association, and purple represents a negative association. Relative risk estimates represent an increased risk of having an additional upper respiratory infection or an increased risk of experiencing diarrhea with each doubling of the relative abundance.

3.5 Discussion

In our prospective study of infants from the general population in New Hampshire, USA, we observed patterns of the early microbiome that were related to the occurrence of infant respiratory infections, wheezing, and diarrhea. Higher diversity of the early infant gut microbiome was associated with a greater number of respiratory infections and

symptoms over the first year of life. Relationships between early microbial patterns and infant outcomes differed by delivery mode, a known contributor to the developing microbiome[13], with stronger associations with alpha diversity among infants born by cesarean section. Using metagenomic sequencing, we found that *Veillonella* in any delivery mode and *Haemophilus* in cesarean deliveries were among the species associated with an increased risk of infant respiratory infections and symptoms.

Our analyses found associations with many bacterial species that are commonly found in oral flora, though these bacteria have also been detected in the gut. An early driver of the gut microbiome is diet. One prospective study found that exclusive breastfeeding was inversely related to lower respiratory tract infections among infants and asthma and allergic rhinitis among children four years of age[166]. The same study highlighted the potential mediating effect of the gut microbiome on the relationship between exclusive breastfeeding and outcomes. Additionally, infants born operatively may be relatively more likely to acquire the genera of such species through breast milk[160].

Respiratory Infections

In our study, *Veillonella*, specifically *Veillonella parvula,* was positively associated with upper respiratory infections, especially in cesarean-delivered infants. *Veillonella parvula* is commonly found in oral flora, although it is observed in both oral and gut ecosystems[167]. We are not aware of studies that have examined *Veillonella parvula*. However, consistent with our findings, a prospective study of 120 Dutch infants found an abundance of three *Veillonella* operational taxonomic units using 16S V4 rRNA

sequencing among one-week-old infants that was associated with a higher number of respiratory infections in the first year of life[47]. In mechanistic studies, *Veillonella parvula* produces propionate in the human gut, which may stimulate IL10-producing Treg differentiation[168,169], and in the small intestine, it induces IL-8, IL-1 β , IL-10, and TNF-α [170].

Among cesarean-delivered infants in our study, a higher relative abundance of *Corynebacterium* species was associated with a greater risk of upper RTIs. *Corynebacterium* species are generally characterized as pathobionts in the respiratory tract[171]. Case series have suggested that *Corynebacterium pseudodiphtheriticum* in the sputum is a driver of pulmonary infection[172,173], and a case-control study of the nasopharyngeal microbiome from France found enriched *Corynebacterium pseudodiphtheriticum* in subjects with viral respiratory tract infections compared to healthy controls (N = 178)[174].

Upper respiratory tract infections in our cohort were associated with a decreased abundance of *Clostridium butyricum* in infant stool samples. One *Clostridium butyricum* strain upregulated inhibitory cytokines such as IL-10 in a mouse model[175]. Clostridium species can promote Treg production and inhibit inflammatory cytokines[32,176], with some associated with systemic infections in humans[177]. Thus, the potential inhibitory impact of *Clostridium butyricum* on an infant's immune response to infection requires further investigation.

Cesarean-delivered infants in our study had a positive association between *Haemophilus influenzae* and the number of any respiratory infections and symptoms. *Haemophilus influenzae* is a bacterial species known to cause several types of infectious diseases, including those of the respiratory tract. Although previous studies have not found associations between *Haemophilus influenza* in the gut microbiome and respiratory infections, the species do reside in the intestinal tract[178]. Further explorations of the gut-lung axis, as well as the origin of such bacteria in the gut, are warranted[179].

Other studies have also found associations between the gut microbiome and respiratory infections. Observations from epidemiologic studies that were not found in our study included a higher abundance of *Bifidobacterium* and *Enterococcus* and a lower abundance of *Escherichia-Shigella, Prevotella, Faecalibacterium* and *Enterobacter* in subjects aged 0 to 3 years with pneumonia compared to healthy controls (N = 33) in a cross-sectional study of Mongolian children[46]. A case-control study of US infants found that infants with a higher gut alpha diversity of *Bacteroides* had a higher likelihood of developing bronchiolitis (N = 155)[48]. Findings from the aforementioned Dutch prospective study found several associations between the bacterial taxa of the infant gut microbiome and the number of respiratory infections, including associations with *Bifidobacterium, Bacteroides,* and *Enterococcus* (N = 120). These findings, as well as ours, require further confirmation.

We found a positive prospective association between the overall alpha diversity at six weeks of life and the risk of upper RTIs. Although our findings for alpha diversity

may seem contradictory to some prior studies that reported negative associations between alpha diversity and health outcomes[180], our study focuses on the early microbiome when diversity is low in healthy babies. We further prospectively examined associations with respiratory infections. Similar to our study, other studies found no associations between alpha diversity and adverse health outcomes but observed differences in the abundance of specific microbes[158]. A cross-sectional study from the US found that infants hospitalized for severe RSV infection had slightly lower alpha diversity of the gut than infants with moderate RSV infection and controls (N = 95)[159]. However, as the infants' gut microbiomes were assessed after the onset of disease, reverse causality was possible in this study; thus, further prospective studies are needed to understand how the overall microbial diversity and colonization of the neonatal and early infant gut reflect immune response to infections.

Wheezing

Wheezing is a respiratory symptom associated with infection, atopy, allergy or a later diagnosis of asthma. We found an association between a higher alpha diversity and an increased risk of wheezing identified by a physician, and this was largely among vaginally delivered infants. Epidemiologic studies have reported associations between the infant gut microbiome and atopic wheezing and asthma in childhood[181–184]. Whether our findings translate to a later risk of asthma will require longer-term follow-up of our cohort.

Diarrhea

Among infants in our study, higher alpha diversity was associated with an increased risk of diarrhea requiring a physician visit in the first year of life. Other prospective studies are lacking, and to our knowledge, prior work includes only case-control studies that measured the infants' stool microbiomes at the time of symptoms. For example, in a small case-control study from South Africa, lower alpha diversity was observed in the stool of infants with gastrointestinal disease compared to infants with respiratory disease and infants with other diseases (N = 34)[185]. In addition to the issue of reverse causality, diarrheal diseases in the US vary in etiology and consequences compared to those in other geographic regions.

We observed that an increased relative abundance of *Streptococcus peroris* and a reduced relative abundance of *Streptococcus salivarius* were associated with a higher risk of diarrhea seen by a physician. Limited data also exist on microbiome composition in relation to diarrheal disease, again with most studies being cross-sectional and with relatively small sample sizes. A study from China of 20 infants with diarrheal illness and 13 controls found differences in gut microbiome composition, with two patients having a higher abundance of *Streptococcus peroris* than controls[186]. Other studies designed to detect pathogens among ill infants and young children compared to controls using 16S rRNA sequencing have also noted a higher abundance of *Streptococcus* species associated with diarrheal illness[187,188]. In mice, *Streptococcus salivarius* strains inhibited inflammation with severe and moderate colitis[189]. In the same experiment, *Streptococcus salivarius* inhibited the activation of the NF- κ B pathway, which induces

proinflammatory gene expression in intestinal epithelial cells[189]. Together, these findings raise the possibility of a role of *Streptococcus* species in susceptibility to diarrhea in early childhood.

Our study had a number of strengths as well as limitations. We carefully collected infant stool samples at approximately six weeks of age from a cohort of pregnant women and their offspring from the general population, and we examined repeated measurements of infection occurrences, respiratory symptoms, and diarrhea and a broad range of potential confounding variables, such as maternal prepregnancy BMI, delivery mode, infant sex, breast feeding at six weeks, antibiotic use during pregnancy, and gestational age, for the analyses. A major challenge to the analysis of microbiome sequencing data is the ability to fully capture their correlated, compositional, and high-dimensional nature when assessing longitudinal outcomes. Therefore, we performed our analyses on the relative abundance of individual species and corrected for multiple hypotheses using the false discovery rate (FDR). Our outcomes were not ascertained by viral or bacterial culture or PCR to confirm the type of infection; as a result, we relied on responses of telephone surveys from caregivers. Additionally, we could not differentiate between the various causes of wheezing and diarrhea in our dataset. Participant recall is a potential source of bias; however, efforts were made to reduce misclassification by including questions on the duration and severity of illness and limiting our analyses to outcomes that involved either a physician visit or a prescription medication. In a review of infants' pediatric medical records, we found caregiver responses to be at least 80% concordant with physician assessments documented in the medical record (unpublished data).

Furthermore, while our study is one of the largest overall, we had limited statistical power in our analyses stratified by delivery mode.

In conclusion, our findings from a prospective birth cohort of US infants suggest that the composition of the microbiome in early life influences the most common health outcomes of infancy, which in turn may have consequences for lifelong disease risk. While higher alpha diversity was associated with respiratory infections and symptoms overall and among vaginal deliveries, the doubling of the relative abundance of unclassified *Veillonella* species and *Haemophilus influenza* species increased the risk of an additional respiratory infection or symptom overall and among cesarean-born infants respectively. Our findings may help to inform interventions aimed at altering the microbiome during this critical window of immune training[27].

Chapter 4: A Prospective Study of the Infant Gut Microbiome in Relation to Vaccine Response

Manuscript currently in press at Pediatric Research

Authors

Yuka Moroishi, Jiang Gui, Kari C. Nadeau, Hilary G. Morrison, Juliette C. Madan, Margaret R. Karagas

Author Contributions

YM, MRK, and JCM contributed to conception of the study. MRK, JCM, KCN, and HGM contributed to data acquisition and processing. YM and JG contributed to statistical analyses. YM, MRK, JCM, and KCN interpreted results.

4.1 Abstract

The establishment of the gut microbiome plays a key symbiotic role in the developing immune system; however, its influence on vaccine response is yet uncertain. We prospectively investigated the composition and diversity of the early-life gut microbiome in relation to infant antibody response to two routinely administered vaccines. Eighty-three infants enrolled in the New Hampshire Birth Cohort Study were included in the analysis. We collected blood samples at 12 months of age and assayed the isolated serum to quantify total IgG and measured antibody to pneumococcal capsular polysaccharide and tetanus toxoid. Stool samples were collected from infants at six weeks of age and sequenced using 16S rRNA, and a subset of 61 samples were sequenced using shotgun metagenomics sequencing. We observed differences in beta diversity for 16S six-week stool microbiome and pneumococcal and tetanus IgG antibody responses. Metagenomics analyses identified species and metabolic pathways in six-week stool associated with tetanus antibody response, in particular, negative associations with the relative abundance of *Aeriscardovia aeriphila* species and positive associations with the relative abundance of species associated with CDP-diacylglycerol biosynthesis pathways. The early gut microbiome composition may influence an infant's vaccine response.

4.2 Introduction

Vaccines reduce infant mortality and morbidity from infections worldwide[190]. The World Health Organization and the US American Association of Pediatrics guidelines outline a standard schedule of immunizations for infants to provide protection against potentially fatal infectious diseases. Rates of vaccine administration vary globally[191]. Among

children born in the US in 2016 and 2017, approximately 99% of infants received vaccines, with only 1.2% not receiving any vaccines by 24 months of age[192]. However, immune responses to vaccinations vary by host with reported factors being sociodemographic characteristics, perinatal exposures, breast or formula feeding, antibiotic use, and variation in timing and other characteristics of the vaccination itself[193].

The early establishment of infant gut microbiome is now known to play an essential role in the development of the immune system[25,26]. Growing evidence points to the impact of the gut microbiome on immune response to vaccination, including among infants treated with antibiotics [59,194–197]. For example, recent studies examined the relation between infant gut microbiota and response to oral poliovirus, bacille Calmette-Guérin, tetanus toxoid, hepatitis B virus, and rotavirus vaccines among infants living in Bangladesh, Ghana, and Pakistan[49–51]. These studies identified specific bacterial taxa in the intestinal microbiome associated with differential vaccine response and microbial compositions in high vaccine responders to be similar to those of healthy infants from high income countries[50,51]. Prospective studies are lacking in common vaccine response in relationship to the very earliest development of the intestinal microbiome. To gain a better understanding of the role of the infant gut microbiome on vaccine response, we investigated the association between the early infant gut microbiome composition and antibody response to two common vaccines administered in the first year of life -- pneumococcal capsid protein and tetanus toxoid -- at one year of age in the New Hampshire Birth Cohort Study (NHBCS).

4.3 Methods

4.3.1 Study Population

The current study included participants in the NHBCS who had stool samples collected at approximately 6 weeks of age for microbiome analyses and infant blood samples at approximately 12 months of age for vaccine response assays. The NHBCS comprised of pregnant women aged 18-45 with a singleton pregnancy who received care at prenatal clinics in New Hampshire, USA as described previously[123]. We collected longitudinal survey data for maternal and infant lifestyles, and we ascertained infant birth characteristics and vaccine information from delivery and pediatric medical records. The Committee for the Protection of Human Subjects at Dartmouth College approved all protocols, and informed consent was obtained from all participants upon enrollment.

4.3.2 Blood sample collection, serum isolation, and assaying

Infant blood samples were collected from well child visits scheduled at approximately 12-months of age post-partum. Serum was isolated from the blood samples, frozen at -80^{o}C, and shipped to Stanford University in Palo Alto, California for serological analysis. Antibody responses to pneumococcus capsule-based vaccines were measured using VaccZyme Anti-PCP IgG Enzyme Immunoassay kit (Binding Site, Birmingham, UK). Antibody responses to tetanus toxoid-based vaccines were measured using Tetanus Toxoid IgG ELISA (Genway Biotech, San Diego, CA).

4.3.3 Stool sample collection, DNA extraction, sequencing, and profiling

Infant stools samples were collected at approximately 6 weeks of age as described previously[124,160]. These samples were aliquoted and frozen at -80^{o}C, and DNA was extracted from thawed samples using Zymo DNA extraction kit (Zymo Research, Irvine, CA). OD260/280 nanodrop was used to measure sample quality and purity. Samples were sent to Marine Biological Laboratory in Woods Hole, Massachusetts for bacterial 16S rRNA gene sequencing of the V4-V5 hypervariable region using Illumina MiSeq (Illumina, San Diego, CA). We conducted quality control measures internally by amplifying in triplicate with one negative control as described previously[160]. We inferred amplicon sequence variants (ASVs) using DADA2[161] and assigned taxonomies using the SILVA database[162]. A subset of stool samples also underwent shotgun metagenomics sequencing using Illumina NextSeq (Illumina, San Diego, CA) as previously described[156]. DNA samples were extracted and sheared to a mean insert size of 400 bp using a Covaris S220 focused ultrasonicator, and sequencing libraries were constructed with Nugen's Ovation Ultralow V2 protocol. We merged and trimmed DNA reads using KneadData[163], inferred taxonomic profiles at the species-level using Metaphlan3[198], and profiled metabolic pathways using HUMAnN3.0[198].

4.3.4 Statistical Analysis

We examined the association between the stool microbiome and antibody response to pneumococcal capsular polysaccharide (PCP) and tetanus toxoid (TT). Using the 16S data, we aggregated ASVs to genera and calculated beta diversity using Bray-Curtis dissimilarity. We used PERMANOVA to test the differences between groups. Using the metagenomics data, we $\log_2$-transformed the relative abundance of bacterial species and

metabolic pathways. We removed species that were present in less than 10% of subjects and performed linear regression on all remaining taxa. We further investigated metabolic pathways by removing pathways that were present in less than 10% of subjects and standardizing the relative abundance of pathways. We used elastic-net to select pathways with possible associations with vaccine response. We then conducted linear regression on each elastic-net selected pathway with FDR correction.

We applied a False Discovery Rate (FDR) threshold of 0.1 to adjust for multiple testing. Factors associated with both the gut microbiome and vaccine response were considered potential confounders and included in all analyses, including infant birth weight (g), breast feeding at 6 weeks (exclusively breast fed/ever formula fed), and maternal pre-pregnancy BMI (kg/m^2). Since delivery mode is not associated with vaccine response, it was not considered a potential confounder. For missing confounding data, we assumed entries were missing at random and used multiple imputation by chained equations and the predictive mean matching method to impute missing data in a randomly chosen iteration. All analyses were performed in R version 3.4.3, using functions *diversity*, *vegdist, ade4, adonis*, *mice*, and *cv.glmnet* in packages 'vegan', 'mice', and 'glmnet'.

We also performed a sensitivity analysis of 65 infants in whom we verified that they had received at least one dose of Pneumococcal Conjugate Vaccine (PCV) 13 or Diphtheria, Tetanus, and acellular Pertussis vaccine (DTaP) in their pediatric medical records.

4.4 Results

4.4.1 Baseline Characteristics

We measured vaccine response in 155 study infants for PCP and 133 subjects for TT. Of those with PCP vaccine response data, 80 stool samples were analyzed by 16S and 59 by metagenomics sequencing. For TT response, 68 infants had 16S data and 53 had metagenomics sequencing data. Our study group had a roughly equal distribution of male (54.2%) and female (45.8%) infants (Table 4.1). Nearly 80% of subjects were delivered vaginally, and roughly half (53.4%) had been exclusively breast fed at 6 weeks of life (Table 4.1). Mean infant birth weight was 3342g and mean maternal pre-pregnancy BMI was 26.1 (Table 4.1). The mean PCP antibody response was 30.4mg/L (SD = 22.5), and the mean tetanus toxoid antibody response was 0.932 IU/mL (SD = 0.759). The 80 infants in our study who had both pneumococcal vaccine response measurements and 16S sequence data all had pneumococcal vaccine response measurements above the preferred PCP protection threshold of 0.2 mg/L[199]. Of the 68 infants who had tetanus vaccine response measurements and 16S data, 6 subjects (8.8%) had measurements below the preferred TT protection threshold of 0.1 IU/mL (Supplementary Figure 4.1)[200].

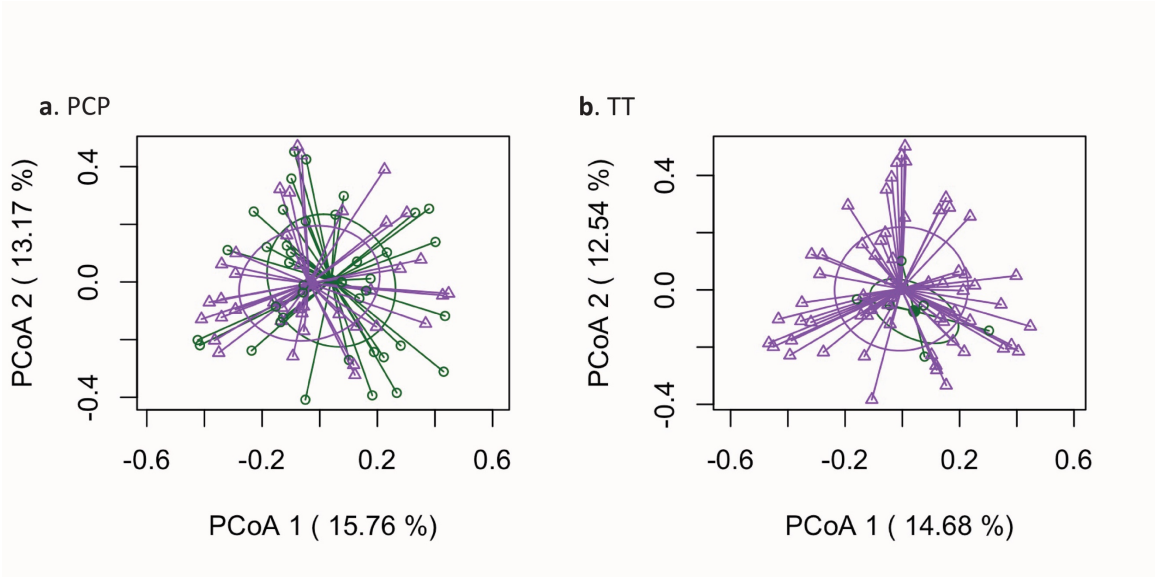
Table 4. 1: Selected Baseline Characteristics of Infants and Mothers in the New Hampshire Birth Cohort Study (N = 83)

Variable	Sample Size	No. (%) or Mean (SD)
Infant Characteristics		
Sex, No. (%)	83	
Female		38 (45.8%)
Male		48 (54.2%)
Delivery Mode, No. (%)	83	
Vaginal		65 (78.3)
Caesarian		18 (21.7)
Breast Feeding at 6 Weeks, No. (%)	73	
Exclusively breast fed		39 (53.4)
Fed any formula		34 (46.6)
Breast Feeding at 1 Year, No. (%)	73	
Exclusively breast fed		19 (26.0)
Fed any formula		54 (74.0)
Daycare (ever in first year of life), No. (%)	63	
Yes		35 (55.6)
No		28 (44.4)
Birthweight (g), mean (SD)	82	3342g (558)
Received PCV13 vaccine, No. (%)	66	
At least one dose recorded in medical record		65 (98.5)
None		1 (1.5)
Received DTaP vaccine, No. (%)	66	
At least one dose recorded in medical record		65 (98.5)
None		1 (1.5)
PCP Vaccine Response (mg/L), mean (SD)	80	30.4 (22.5)
TT Vaccine Response (IU/mL), mean (SD)	68	0.932 (0.759)
Maternal Characteristics		
Maternal pre-pregnancy BMI (kg/m^2), mean (SD)	83	26.1 (5.7)
Parity, No. (%)	83	
0		44 (53.0)
1+		39 (47.0)

4.4.2 16S V4-V5 rRNA Gene: Beta Diversity

In PERMANOVA analyses of pair-wise community composition, we observed a weak association between PCP antibody response at or below versus above the median (PERMANOVA P = 0.112) (Figure 4.1.1). We further observed a borderline statistically significant difference in beta diversity for TT antibody response at or below the TT protection threshold and above the threshold (PERMANOVA P = 0.065) (Figure 4.1.2).

Figure 4. 1: PCoA plots of bacterial 16S V4-V5 rRNA sequencing Bray-Curtis Dissimilarity for PCP and TT. PCP groups assigned by median PCP IgG concentration threshold. TT groups assigned by preferred protection threshold of 0.1IU/mL. Percentages on the X and Y axis of plots represent percentage of variance explained by first two eigenvectors.

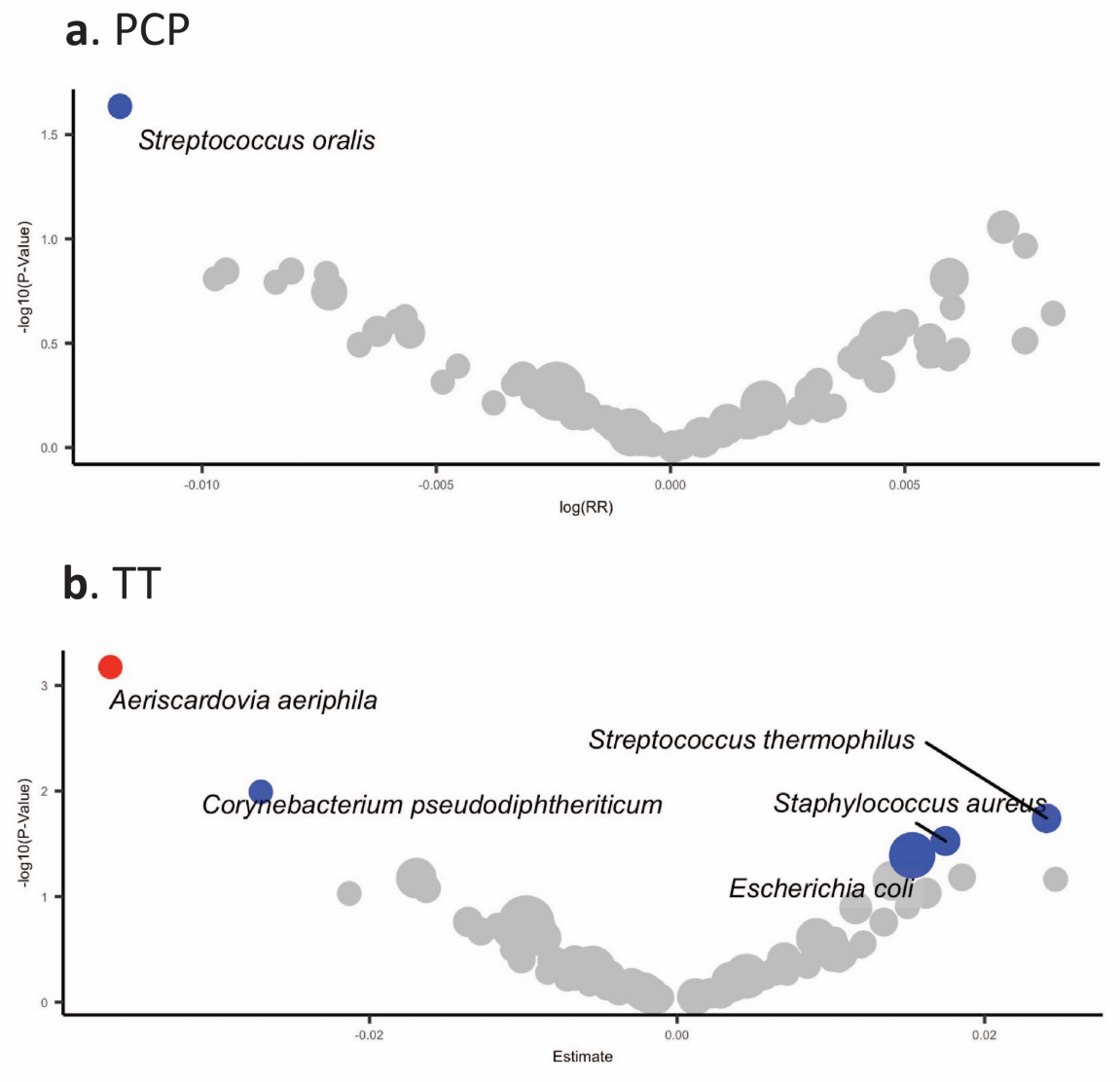


4.4.3 Metagenomics: Species and Pathways

Streptococcus oralis species was inversely associated with PCP antibody response although the association did not meet FDR threshold (Figure 4.2.1). *Aeriscardovia aeriphila* was inversely associated with TT antibody response, whereas *Staphylococcus aureus*, *Escherichia coli*, *Streptococcus thermophilus*, and *Anaerococcus vaginalis* were

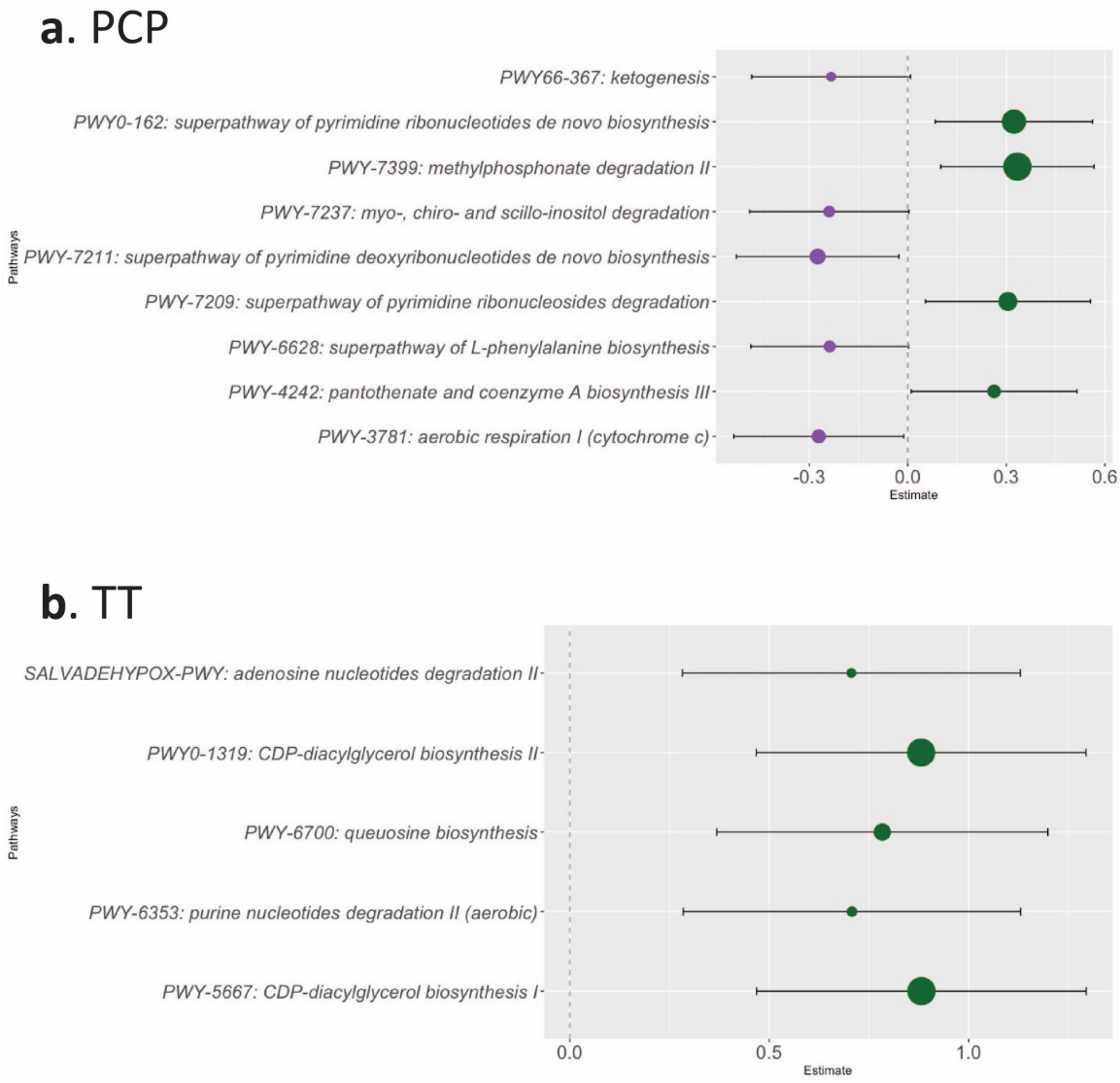
positively associated with TT antibody response (Figure 4.2.2). Of these, only *Aeriscardovia aeriphila* remained statistically significant after FDR correction (Figure 4.2.2) (Supplementary Table 4.1).

Figure 4. 2: Associations between metagenomics bacterial species and vaccine response. Dots indicate bacterial species, and size of dots vary by mean abundance. Blue indicates species with p-value < 0.05. Red indicates species with p-values <0.05 and meet FDR correction.



Several metabolic pathways were found to be associated with vaccine response. The 9 pathways associated with lower PCP vaccine response included higher abundance of taxa related to aerobic respiration I (cytochrome c), superpathway of L-phenylalanine biosynthesis, superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis, myo-, chiro-, and scillo-inositol degradation, and ketogenesis (Fig. 4.3.1) (Supplementary Table 4.2). In contrast, higher abundance of taxa related to pantothenate and coenzyme A biosynthesis III, superpathway of pyrimidine ribonucleosides degradation, methylphosphonate degradation II, and superpathway of pyrimidine ribonucleotides de novo biosynthesis were associated with higher PCP antibody response (Fig. 4.3.1) (Supplementary Table 4.2). Five pathways were positively associated with TT antibody response, including CDP-diacylglycerol biosynthesis I, purine nucleotides degradation II (aerobic), queuosine biosynthesis, CDP-diacylglycerol biosynthesis II, and adenosine nucleotides degradation II (Fig. 4.3.2) (Supplementary Table 4.2). No inverse associations were observed for TT.

Figure 4. 3: Associations between elastic-net and metabolic pathways and vaccine response. Dots indicates effect size, and horizontal bands indicate 95% CI. Green dots represent positive association, while purple dots represent negative association. Size of dots vary by p-value: larger dot indicates smaller p-value. Only pathways selected by elastic net and FDR correction shown here.



Our sensitivity analyses of infants with medical record confirmation of vaccination revealed similar results. Results are provided in Supplementary Figure 4.2, Supplementary Figure 4.3, and Supplementary Figure 4.4 for beta diversity, bacterial species, and metabolic pathways respectively.

4.5 Discussion

In a prospective study of the infant gut microbiome using both 16S and metagenomic sequencing, we observed differences in infant gut microbiome composition at six weeks of age in relation to pneumococcal conjugate and tetanus toxoid vaccines at one year of age. Microbial community structure (beta diversity) was associated with both PCP and TT antibody response although with limited statistical power. In analyses of individual bacterial species, associations were observed with decreased antibody response to TT including *Aeriscardovia aeriphila* after correction for multiple comparisons. Pathway analyses indicated several potential mechanisms by which microbial metabolites might influence vaccine response especially those related to CDP-diacylglycerol biosynthesis.

Guidelines by the AAP in the US state that children under 2 years of age should receive three doses of PCV13 that provide protection against 13 strains of *Streptococcus pneumoniae*[201]. These microbes are responsible for the majority of invasive bacterial infections in infants and young children, including otitis media infections, resulting in significant morbidity, antibiotic exposure, and at times difficulty with hearing[202]. Streptococcus also results in pneumonia, “pink eye”, and other infectious illnesses in children 2 months and older, hindering daycare attendance, and with consequent financial loss to working parents. Following PCV, antibodies to PCP antigens conjugated to a carrier protein are induced by both T cell and B cell responses[203]. Tetanus is a life-threatening infection that is caused by *Clostridium tetani* and is preventable by a toxoid vaccine. It is part of several combination vaccines including DTaP[201]. DTaP primarily

induces T cell response[204]. By age one, most infants in the US would have received three doses of PCV13 and DTaP[201].

Using 16S V4-5 rRNA sequencing data, we found marginally statistically significant between-group (beta diversity) differences in PCP and TT antibody response. One prior study of Chinese infants observed differences in beta diversity among those with and without a positive oral poliovirus IgA vaccine response based on 16S rRNA analyses of stool collected the day of vaccination. However, many prior studies have not detected such differences with either oral poliovirus[205] or rotavirus vaccination[206,207].

The association between specific microbial taxa and infant vaccine response has been examined in only a few prospective studies of infants. In a study of 48 Bangladeshi infants, positive associations were found for *Corynebacterium* and *Bifidobacterium* and negative associations for *Escherichia/Shigella* and *Acinetobacter* and T cell proliferation response to tetanus toxoid vaccine among young infants (measured at 15 weeks of life)[49]. In this study, a positive association for *Actinomyces* and a negative association for *Staphylococcus* with IgG response to TT vaccine also were observed. Further, a positive association was identified between *Bifidobacteriaceae* and tetanus toxoid T cell proliferation response. This contrasts with our finding of an inverse association with *Aeriscardovia aeriphila*, a bacterial species of the family *Bifidobacteriaceae*. A later publication from the Bangladesh cohort of infants found a positive association between *Bifidobacterium* and CD4 and IgG response to vaccines, including to the TT vaccine at age 2 years[52]. Other studies have focused on vaccine response to oral live attenuated

rotavirus. For instance, a case-control study of 78 Ghanaian infants found negative correlations between rotavirus vaccine response in serum collected four weeks after the last vaccine dose and *Bacteroides* and *Prevotella* species, and positive correlations with bacteria in the *Bacilli* phylum in 6 week stool samples collected before vaccination[50]. Similarly, in a study performed in Pakistan, a case-control study found positive correlation between rotavirus vaccine response 28 days after the last vaccine dose and relative abundance of *Clostridium cluster XI* and Proteobacteria, including *Serratia* and *Escherichia coli* in the pre-vaccination stool of 20 infants[51]. A randomized control trial found correlations between *Bifidobacteria* and antipoliovirus IgA response in 30 French infants[208]. Another study of polio vaccine in 107 Chinese infants found increased *Firmicutes* and decreased *Actinobacteria* in stool collected the day of the last oral poliovirus vaccine dose among infants with negative IgA response[205]. Many of these studies investigated response to oral vaccines, and gut microbiota may interact differently with oral vaccines compared to intramuscular vaccines due to direct contact between oral vaccines and the gut. Thus, further studies are needed in diverse study populations and with response to multiple types of vaccines.

External factors that influence both the gut microbiome and vaccine response highlight opportunities for recommendations and interventions to improve vaccine response. Germ-free mice treated with antibiotics and mice deficient in toll-like receptor 5 expression have demonstrated lower responses to influenza vaccine, but not other vaccines[197]. Diminished vaccine response was restored after reconstituting their gut microbiome with flagellated *E. coli*. In another mouse experiment, lower antibody response after

ovalbumin and Freund's adjuvant immunization was observed in infant mice whose mothers were treated with antibiotics during pregnancy[196]. The study also observed that germ-free mice with deficient antibody response can increase their response after introduction to normal gut flora, raising the possibility that interventions may be able to enhance immune response. A similar study in mice found associations between maternal and early-life antibiotic exposure and gut microbiome dysbiosis and lower IgG responses for several vaccines, including PCV13 and Hexa, which produces antibodies to tetanus[59]. Further, impaired PCV13 response in mice treated with antibiotics was not observed if they received fecal transplant from age-matched untreated mice, while impaired response remained in mice treated with antibiotics and fecal transplant from antibiotic treated mice. Further experimental studies and mediation analyses will help establish the potential for microbiome-directed interventions to bolster immune response to vaccines.

Further evidence of the importance of the microbiome on vaccine response comes from studies on beneficial effects of probiotics. A study of 20 French infants found those given *Bifidobacterium breve* strain C50 and *Streptococcus thermophilus* supplementation in the first 4 months of life had higher antipoliovirus IgA response at 4 months[208]. Interestingly, in our analyses, we observed a positive association between *Streptococcus thermophilus* and TT antibody response, although this was no longer statistically significant after FDR correction. Prenatal and early life supplementation with *Lactobacillus rhamnosus GG*, *L. rhamnosus LC705*, *Bifidobacterium breve Bbi99*, and *Propionibacterium freudenreichii* ssp. *shermanii* in the first 6 months of life was associated with higher IgG antibody response to *Haemophilus influenzae* type b at 6 months of age in a randomized controlled

study of 61 allergy-prone Finnish infants[209]. A randomized control study of 61 mothers and their infants found maternal *Lactobacillus rhamnosus GG* supplementation starting at 36 weeks gestation was associated with lower IgG response of tetanus toxoid, *Haemophilus influenzae* type b, and several serotypes of PCV7 vaccines in infants at 12 months of age[210]. Thus, further research on antibiotic use and prenatal and early life probiotic supplementation is warranted to elucidate immunomodulation by gut microbiota.

Based on metagenomic analysis, we found several metabolic pathways associated with PCP and TT vaccine response. Nucleotides including purine and pyrimidine have an important role in the immune system. Several pathways involving pyrimidine were associated with differential response to PCP. A link between pyrimidine biosynthesis and the immune system has been observed[211,212]. Furthermore, the purine nucleotide degradation II pathway was observed to be positively associated with TT response in our study. CDP-diacylglycerol biosynthesis I and II were positively associated with TT response in our study. CDP-diacylglycerol has an important function in lipid metabolism, which could affect the immune system[213].

The main strength of our study is the prospective analyses of early-life infant gut microbiome and vaccine response. However, there also were limitations. Due to the high dimensional nature of microbiome data and available sample size, our analyses had limited statistical power. Therefore, we applied a filtering process as well as regularization techniques for dimension reduction and variable selection. Delivery mode

was not related to vaccine response, and therefore, not included as a covariate in our analyses. Several potential confounders, such as breast feeding, were adjusted for in our analyses; however, residual confounding remains a possibility. Further work is warranted to investigate the gut microbiome as a possible mediator or effect modifier on the relationships between exposures and vaccine response. Delivery mode and breast feeding are two such exposures associated with relative abundance of gut bacteria[214], and associations with these exposures have been observed in our cohort[124,215]. Breast feeding may influence a child's vaccine response[193]. Analyses with larger sample sizes may be able to identify mediating or modifying effects as well as to corroborate or refute associations lacking statistical significance after correction for multiple hypotheses found in our species and pathway analyses.

In conclusion, we found patterns of the developing gut microbiome captured at a very early time point in development associated with differential response to vaccines administered in the first year of life. The microbes and pathways associated with antibody response to PCP and TT may offer clues to the critical role the developing microbiome plays in shaping the immune system as measured by vaccine response. Our findings provide insight into possible interventions to optimize antibody response and improve vaccine efficacy during a critical time in early immune maturation and when susceptibility to infection is at its highest.

Chapter 5: HT-MMIOW: A Hypothesis Test approach for Microbiome Mediation using Inverse Odds Weighting

Authors

Yuka Moroishi, Zhigang Li, Juliette C. Madan, Hongzhe Li, Margaret R. Karagas, Jiang Gui

Author Contribution

YM, MRK, and JG conceived the study. YM developed the statistical method, ran the simulations, and implemented the R package. ZL, HL, and JG contributed to the development of the method. MRK and JCM contributed to data acquisition and processing.

5.1 Abstract

The human microbiome has an important role in determining health. Mediation analyses quantify the contribution of the microbiome in the causal path between exposure and disease; however, current mediation models cannot fully capture the high dimensional, correlated, and compositional nature of microbiome data and do not typically accommodate dichotomous outcomes. We propose a novel approach that uses inverse odds weighting to test for the mediating effect of the microbiome. We use simulation to demonstrate that our approach gains power for high dimensional mediators, and it is agnostic to the effect of interactions between the exposure and mediators. Our application to infant gut microbiome data from the New Hampshire Birth Cohort Study revealed a mediating effect of 6-week infant gut microbiome on the relationship between maternal prenatal antibiotic use during pregnancy and incidence of childhood allergy by 5 years of age.

5.2 Introduction

The human gut microbiome and immune system interact to form a bidirectional relationship [26], and the developing gut microbiome plays a crucial role in immune system maturation in infancy [25]. As a result, perturbations in the gut microbiome is linked to clinical outcomes through infancy and childhood, and into adulthood. External factors shape the gut microbiome including delivery mode, diet, and antibiotic use [53]. Understanding the three-way interplay among external factors, the gut microbiome, and

clinical outcomes can help generate opportunities for intervention such as therapeutics and probiotics for modulating the gut microbiome to improve health outcomes.

Mediation analysis quantifies the contribution of a variable in the causal pathway between an exposure or treatment and an outcome. These analyses are especially useful in health research to determine possible interventions to prevent disease or produce better health outcomes. Upon development of approaches to mediation analysis including traditional approaches [63,64] and causal inference approaches [65–68,216], many extensions of mediation analyses have also been published in recent years. These include models for multiple mediators [68,69] and those that account for interactions between exposures and mediators [217,218]. Other approaches include inverse odds ratio weighting (IORW) [219] or inverse odds weighting (IOW) [220] to quantify the total, direct, and indirect effects. Testing the mediation effect can be conducted using common approaches such as the Sobel test [221] and the joint significance test [64].

Mediation analysis for high dimensional mediators has become increasingly popular for modeling biomarkers. These include methods using PCA and regularization for dimension reduction [70–73], multiple testing procedures [74–79], and others [80,81]. Methods for mediation analysis have also extended to microbiome data: many test for mediation with continuous outcome [82–87], some account for interactions between exposure and mediators [84,87], and one conducts mediation analysis for dichotomous outcomes [88]. Most of these methods, however, require specification of the microbiome model, which is difficult due to the sparse, high dimensional, and compositional nature of microbiome data.

Furthermore, to our knowledge, none of them can accommodate both dichotomous outcomes and high dimensional microbiome mediators.

We propose a novel approach, HT-MMIOW, a Hypothesis Test approach for Microbiome Mediation using IOW, that tests the mediation effect of high dimensional microbiome data on both continuous and dichotomous outcomes using IOW. This approach uses the isometric log-ratio transformation (ilr) to account for the compositional nature of microbiome data and reduces dimensions using the Uniform Manifold Approximation and Projection (UMAP). The components generated form UMAP serve as mediator variables. A permutation test is performed to determine the statistical significance of the overall microbial mediation effect. Our simulation results demonstrate that this new approach is well powered when the number of true mediators is large or the indirect effect is large. We present an application of our approach to infant gut microbiome data to detect a mediating effect between antibiotic exposure and a diagnosis of allergy before the age of 5.

5.3 Methods

5.3.1 Identification

Suppose the exposure, mediator, and outcome are denoted by an $n \times 1$ vector E, an $n \times p$ matrix $\boldsymbol{M}$, and an $n \times 1$ vector Y respectively. Y is known to succeed $\boldsymbol{M}$, and $\boldsymbol{M}$ is known to succeed E. The microbiome matrix $\boldsymbol{M}$ is composed of abundance of p taxa for each subject in n, and there are t true mediators in $\boldsymbol{M}$. For simplicity, let both E and Y be dichotomous. Let X be an $n \times q$ matrix of covariates. Under the Baron and Kenny's

classical mediation framework for a single mediator M, mediation analysis is represented by the following equations:

$$logit(\Pr(Y = 1 \mid e)) = a_0 + a_1 E$$

$$E[M|e] = b_0 + b_1 E$$

$$logit(\Pr(Y = 1 \mid e, m)) = c_0 + c_1 E + c_2 M$$

The direct effect of E on Y is c_1, the indirect effect of E on Y through M is $b_1 c_2$, and the total effect is $a_1 = c_1 + b_1 c_2$. The Sobel test determines the existence of an indirect effect with the hypothesis $H_0: b_1 c_2 = 0$ vs $H_A: b_1 c_2 \neq 0$.

5.3.2 Dimension reduction of microbiome data

The standard mediation model requires the knowledge of true mediators. Mediation models for multiple mediators exist, but many assume that mediators are independent or that the causal relationship between mediators are known. However, microbiome data is high dimensional, compositional, and correlated. We do not know which microbial taxa are true mediators, and we do not know the causal relationships among taxa. We propose mapping the microbiome data from the Aitchison-simplex to Euclidean space and reducing the dimensions to obtain a smaller set of mediation features.

The ilr is an extension of log-ratio transformations for compositional data that extends on traditional additive (alr) and centered (clr) log-ratio transformations. Briefly, alr is the log-ratio of a value in the composition and a reference value, and clr is the log-ratio of a value and the geometric mean of values in the composition. There are, however,

limitations to these approaches; alr produces transformations do not preserve distances, and clr produces transformations with a singular covariance matrix. In Ilr, distance is preserved when data is transformed from the p-dimensional Aitchison space to the $p-1$ dimensional Euclidean space, and the resulting vectors are orthogonal and interpretable in analyses [222]. Each value of ilr transformation is a "balance" between two subsets of $\boldsymbol{M}$, denoted by R for those on the left of the balance and S for those on the right of the balance:

$$ilr(R,S) = \sqrt{\frac{rs}{r+s}}\log\left(\frac{g(m_R)}{g(m_S)}\right)$$

where r is the cardinality of R, s is the cardinality of S, m_R are the values in R, m_S are the values in S, and $g(\cdot)$ is the geometric mean function.

We impute a pseudocount of 0.5 to zero values and apply ilr to the microbiome data to account for compositionality. After ilr, we reduce dimensionality using UMAP. Briefly, UMAP is a fast dimension reduction procedure that models the manifold with a fuzzy topological structure by searching for a low-dimensional projection with the closest possible equivalent fuzzy topological structure [223]. UMAP allows for non-linear dimension reduction and meaningful separation between clusters. Applying UMAP on $ilr(\boldsymbol{M})$ transforms our microbiome data to an $n \times c$ matrix $\boldsymbol{U}$, where c is a user-specified number of components.

5.3.3 Compute the indirect effect and perform permutation test for mediation

The IORW and IOW approach for causal mediation analysis allows for the decomposition of the total effect to direct effect and indirect effect. This approach can accommodate multiple mediators due to a weighting procedure that removes the need to specify the model to regress the multiple mediators on exposure. This approach is advantageous for microbiome data because of the difficulty in modeling the joint conditional density of high dimensional, compositional, and correlated microbiome features. Using the IOW approach, the total effect of exposure E on outcome Y, given covariates $\boldsymbol{X}$, can be estimated using the model:

$$\hbar(Y|E,\boldsymbol{X}) = \beta_0 + E\beta_1 + \boldsymbol{X}\beta_2 \tag{1}$$

where β_1 represents the total effect of E on Y, given $\boldsymbol{X}$. $\hbar(\cdot)$ is the user-specified link function and ε represents the error. We implement IOW to condense the association between E and mediators $\boldsymbol{U}$, conditional on $\boldsymbol{X}$. The weights can be estimated using the model with a user-specified link function $j(\cdot)$:

$$j(E|\boldsymbol{U},\boldsymbol{X}) = \alpha_0 + \boldsymbol{U}\alpha_1 + \boldsymbol{X}\alpha_2 \tag{2}$$

For each observation, the weight is the inverse of the predicted odds in the exposed group and 1 for the unexposed group. Using IOW as opposed to IORW stabilizes the weights, though it may introduce small bias to estimates [220]. We then use a weighted regression model to estimate the direct effect of E on Y, conditional on $\boldsymbol{X}$.

$$\hbar(WY|WE,W\boldsymbol{X}) = \gamma_0 W + WE\gamma_1 + W\boldsymbol{X}\gamma_2 \tag{3}$$

where γ_1 represents the direct effect of E on Y, given $\boldsymbol{X}$, and W is an $n \times 1$ vector of weights. Due to this weighting procedure, we do not need to specify interactions between E and $\boldsymbol{U}$.

We can now estimate the indirect effect using parameters from (1) and (3), which becomes our observed test statistic for our permutation test: $T_{obs} = \beta_1 - \gamma_1$. The null hypothesis of no mediation effect of the microbiome is expressed by:

$$H_0 : \beta_1 - \gamma_1 = 0,$$

and the alternative hypothesis that a mediation effect of the microbiome exists is expressed by:

$$H_A : \beta_1 - \gamma_1 \neq 0.$$

Our permutation test can calculate the P-value using the formula $\left(\sum_{j=1}^{B} I(\left|T^{(i)}\right| > |T_{obs}|)\right)/B$, where B is the total number of permutations, $I(\cdot)$ is an indicator function, $T^{(j)}$ is the test statistic under the null hypothesis for permutation $j = 1, \ldots, B$.

5.4 Simulation

3.1 Simulation overview

We conducted simulations to evaluate the power of our proposed approach on continuous and dichotomous outcomes. We simulated E and $\boldsymbol{M}$ using SparseDOSSA [224]. Briefly, this tool uses zero-inflated log-normal distribution to simulate realistic microbial community structure of human stool and covariates that correlate with the simulated microbiome

data. We applied UMAP on $\boldsymbol{M}$ to reduce dimensionality to 2 components. We set the effect size of the exposure on outcome at 5, the percentage of microbial features associated with the exposure at 50%, and effect size of the exposure on mediator at 3 for simulated data generation using SparseDOSSA. We simulated outcomes Y using a standard logistic regression with the exposure and scaled relative abundance of true microbial taxa. mediators selected at random from those associated with the exposure, with the effect size of the exposure on outcome set at 5 and a random noise of of $N(0,1)$. Exposure variables were dichotomous, and outcome variables were continuous and dichotomous.

To fine-tune our approach, we evaluated the performance of HT-MMIOW with various dimension reduction techniques. These include 1) UMAP as described previously; 2) PCA to n-components then UMAP; 3) PCA with components explaining 100% of the variance; and 4) PCA with components explaining 80% of the variance. We investigated varying effect sizes of true mediators on outcome (effect size = 0.5, 1, 5), which serve as proxy for smaller to larger indirect effects, and varying number of true mediators in the microbiome sample (t = 1, 5, 10). We also varied sample sizes (n = 50, 70, 100, 150, 300, 500) and the number of taxa in the microbiome data (p = n, 2n). We compared the performance of HT-MMIOW with a distance-based omnibus test using Bray-Curtis and Jaccard distances [83]. Though the omnibus test is designed for continuous outcomes, we sought to examine its performance on dichotomous outcomes due to the lack of methods for high dimensional microbiome mediators and dichotomous outcomes. All analyses

were conducted using R version 4.0.2 and packages “SparseDOSSA2”, “magrittr”, “dplyr”, “ggplot2”, “compositions”, “umap”, “foreach”, “doParallel”, and “bda”.

3.2 Simulation results

Figures 5.1 and 5.2 display the simulation results of HT-MMIOW for continuous outcomes under varying conditions compared with the omnibus test. HT-MMIOW with UMAP dimension reduction performs better than HT-MMIOW with PCA and UMAP when the number of taxa are twice the of sample size. HT-MMIOW also performs better than the omnibus test for all conditions, especially when the number of true mediators is small. Using HT-MMIOW with only PCA as the dimension reduction technique did not perform well. HT-MMIOW with PCA components explaining 80% of the variance generally performed better than HT-MMIOW with PCA components explaining 100% of the variance, and this may be because fewer components in the inverse odds ratio mediation approach provides more power. These power calculations were based on an empirical threshold of 0.05. Though type I error for continuous outcomes were around 0.05 based on the 0.05 empirical threshold, type I error based on an empirical threshold of 0.01 were smaller (Supplementary Figure 5.1). HT-MMIOW using a 0.01 empirical threshold continued to yield higher power than the omnibus distance test (Supplementary Figure 5.2) .

Figure 5. 1: Power calculations for varying sample size, effect size, and number of true mediators for continuous outcomes and $p = n$. Results are based on 200 simulations. Each line represents the hypothesis test procedures evaluated.

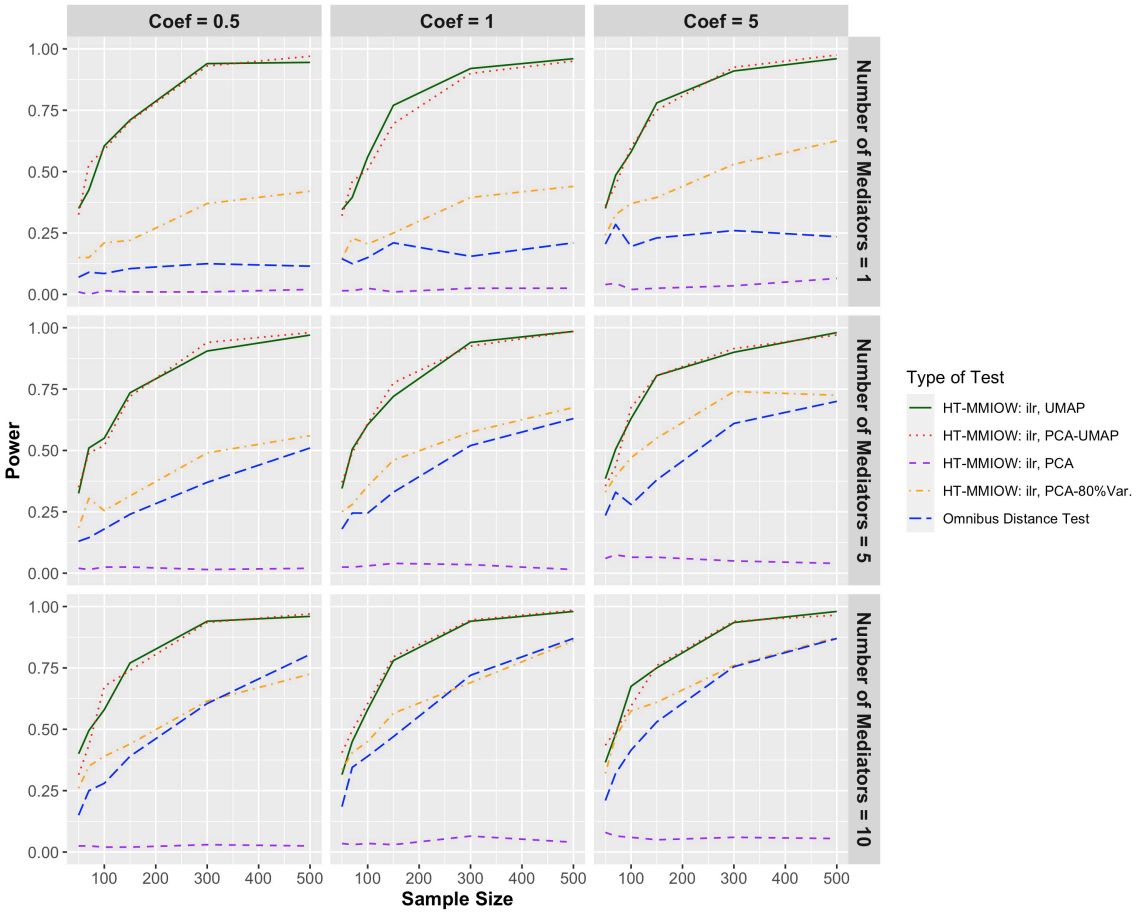
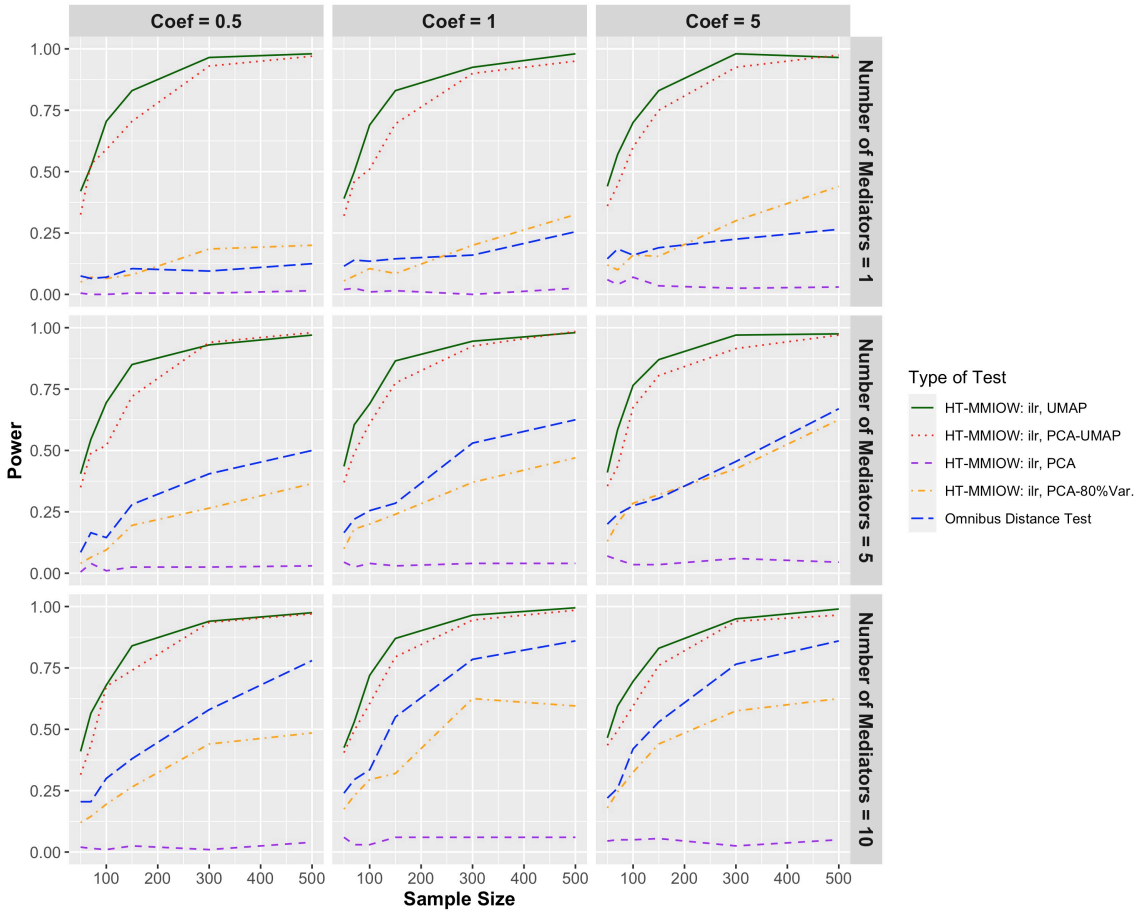


Figure 5. 2: Power calculations for varying sample size, effect size, and number of true mediators for continuous outcomes and $p = 2n$. Results are based on 200 simulations. Each line represents the hypothesis test procedures evaluated.



For dichotomous outcomes, HT-MMIOW performed well with stronger effect of mediators on Y and with larger numbers of true mediators in M (Figure 5.3, Figure 5.4). Using UMAP performed just as well as using PCA and UMAP. HT-MMIOW also performed better than the omnibus test in all conditions. Again, HT-MMIOW's performance was lower when using only PCA as the dimension reduction technique. As expected, performances for both continuous and dichotomous outcomes increased with increasing sample size. Again, the above power calculations were based on an empirical threshold of 0.05. The type I error for dichotomous outcomes were inflated when an

empirical threshold of 0.05 was used; however, type I error based on an empirical threshold of 0.01 were close to zero (Supplementary Figure 5.1). HT-MMIOW using a 0.01 empirical threshold for dichotomous outcomes yielded higher power than the omnibus distance test (Supplementary Figure 5.3).

Figure 5. 3: Power calculations for varying sample size, effect size, and number of true mediators for dichotomous outcomes and $p = n$. Results are based on 200 simulations. Each line represents the hypothesis test procedures evaluated.

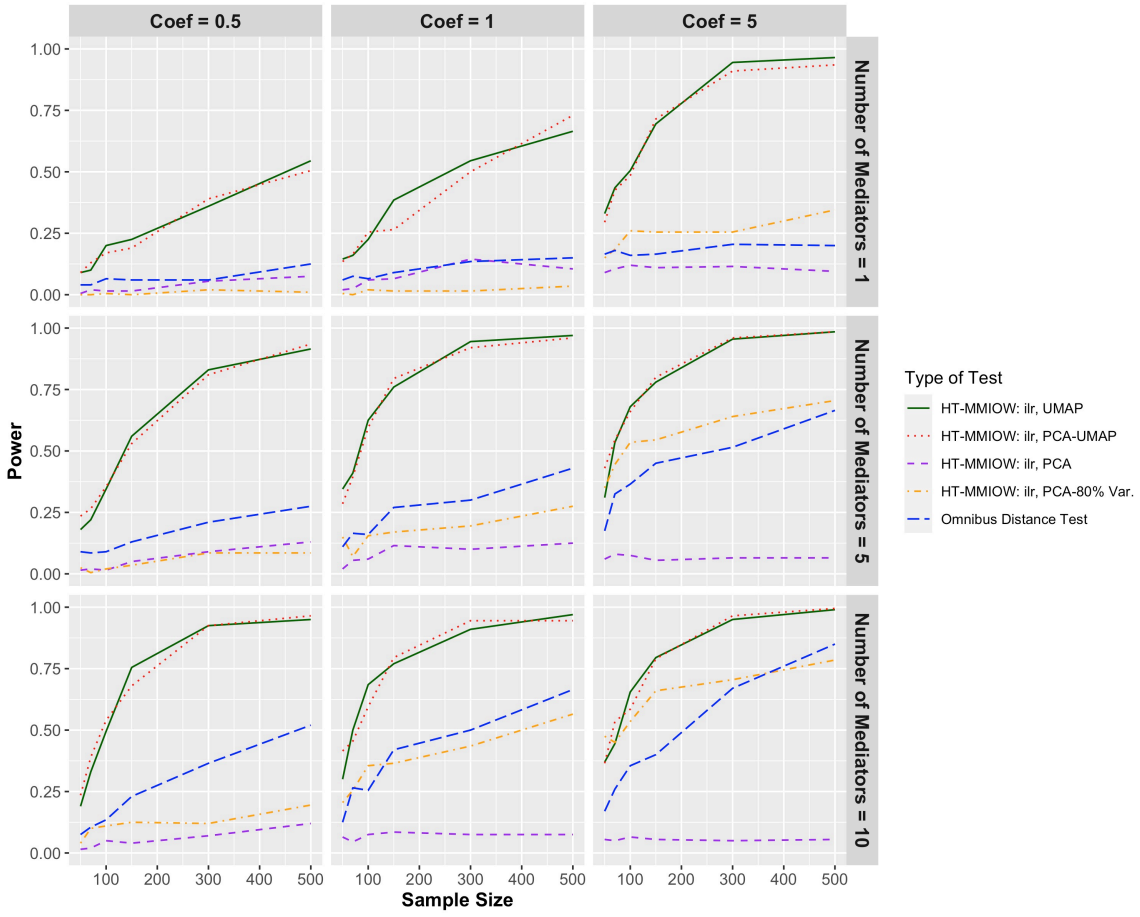
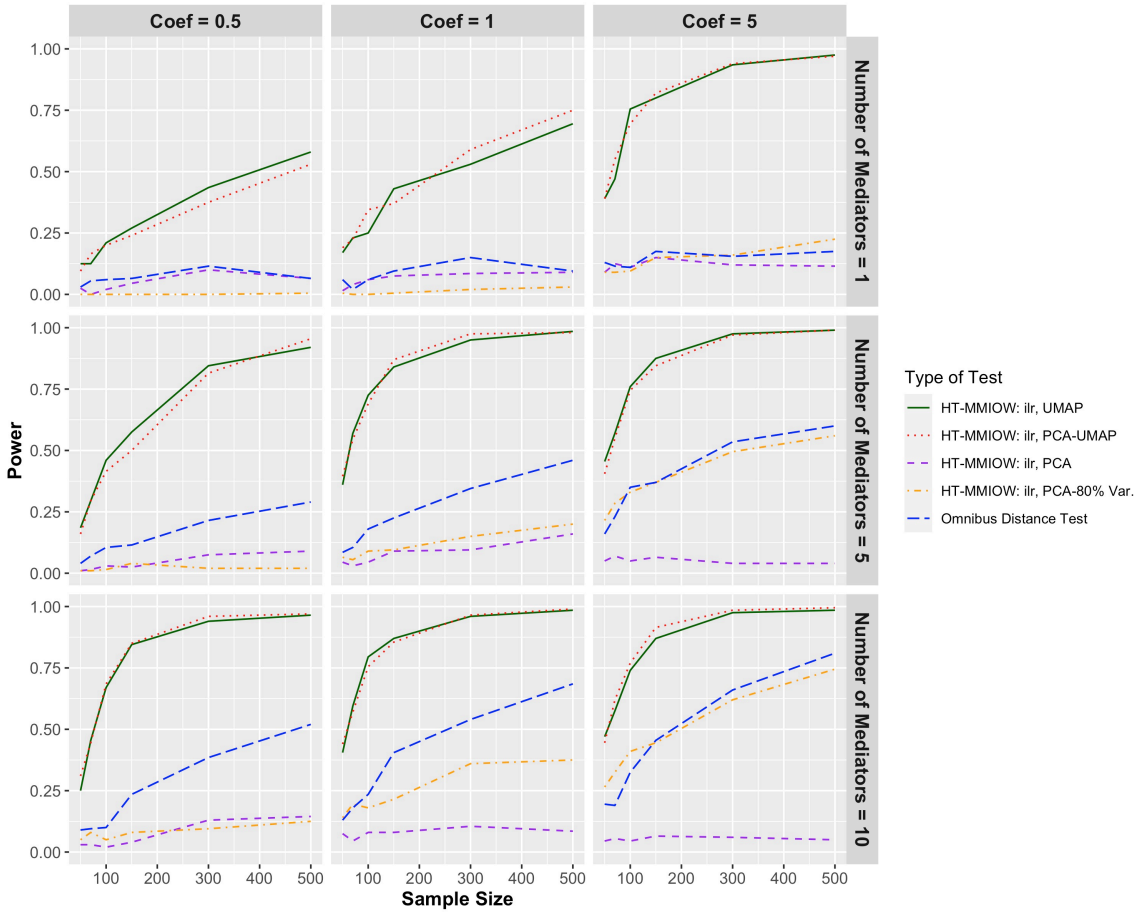


Figure 5. 4: Power calculations for varying sample size, effect size, and number of true mediators for dichotomous outcomes and $p = 2n$. Results are based on 200 simulations. Each line represents the hypothesis test procedures evaluated.



5.5 Application to NHBCS data

Prenatal antibiotic use has been linked to development of infant and childhood allergy [16,156], and it is also associated with compositional differences in the infant gut microbiome [225]. We applied HT-MMIOW to test the effect of the infant gut microbiome as a potential mediator in the causal path between prenatal antibiotic use and allergy on data from the New Hampshire Birth Cohort Study (NHBCS). The NHBCS is a prospective birth cohort of mother-infant dyads who received prenatal care in New Hampshire clinics. The study recruited pregnant women with ages ranging between 18

and 45, who had a singleton pregnancy and were served by a private water system as described previously [123]. Prenatal use of antibiotics was reported in prenatal records. The child's caregivers reported whether the child had any allergies diagnosed by a physician during telephone interviews conducted when infants turned approximately 4, 8, 12, and 18 months of age, and then at 1 year intervals thereafter. Infant stools samples were collected at approximately 6 weeks of age and sequenced using Illumina MiSeq (Illumina, San Diego, CA) for bacterial 16S rRNA gene sequencing of the V4-V5 hypervariable region at Marine Biological Laboratory in Woods Hole, Massachusetts. We inferred amplicon sequence variants using DADA2 [161] and assigned taxonomies using the SILVA database [162].

A total of 412 mother-infant dyads were included in this analysis. Of these, 72 mothers (17.5%) self-reported to antibiotic use during pregnancy, and 39 children (9.5%) had been diagnosed with allergy by 5 years of age. Based on the 16S sequencing reads of our stool samples, we identified 705 different genera of bacteria. In our mediation test, HT-MMIOW produced a P-value of 0.02; this gives us evidence to reject the null hypothesis and suggest that the infant gut microbiome mediates the relationship between antibiotic use during pregnancy and incidence of allergy. Due to variable duration of follow-up, we adjusted for age at which allergy was first diagnosed or age at last follow-up in our model.

5.6 Discussion

We proposed a novel hypothesis test for microbiome mediation effect that utilizes a dimension reduction procedure for microbiome data and a mediation analysis procedure utilizing IOW to test if an indirect effect exists. Our simulation scenarios evaluated the power of our approach under varying conditions. HT-MMIOW was highly powered when the effect of the mediator on the microbiome was large or when the number of true mediators in the microbiome data was large. Compared to a conventional hypothesis test, HT-MMIOW generally performed better for both continuous and dichotomous outcomes. Our approach is one of the only hypothesis tests that can test for the indirect effect in high dimensional microbiome mediators and dichotomous outcomes.

Our proposed hypothesis test is flexible in its application. HT-MMIOW can be used for multiple types of exposure and outcome data; the user may specify their own link functions in the regression models. Our approach can also be adjusted to account for other types of high dimensional mediation analysis, including genomics, by replacing the centered log-ratio transformation step to a transformation of the user's choice. The total effect, direct effect, and indirect effect can be calculated using the IOW approach, and confidence intervals can be estimated using bootstrap methods.

The main strength of our approach is that HT-MMIOW reduces high-dimensional microbiome data to a few independent components that are representative of microbial community structure. The IOW procedure accommodates multiple mediators in the mediation model. Furthermore, the IOW frameworks eliminates the need to specify a

model to regress the exposure on the joint conditional density of multiple compositional mediators. Another strength is that IOW does not require us to specify the interaction between exposure and mediators, even if they exist.

Despite its strengths, HT-MMIOW is not without limitations. One limitation of our approach is that the mediation effect of true microbial mediators must be large for the approach to detect an indirect effect. We must also assume that there is no unmeasured confounding. Additionally, the number of true mediators must be large for smaller mediation effects and smaller samples sizes. Our method also cannot detect the mediation effect of specific microbes or clusters of microbes. Nevertheless, this approach serves as an important first step in determining whether a mediation effect exists. Further work is warranted to identify key microbial features and interactions responsible for driving this effect.

Chapter 6: Concluding Remarks

There is increasing recognition of the important role exposures and the early gut microbiome play in the health of infants and young children. In this thesis, I investigated the three-way interplay among exposures, the infant gut microbiome, and outcomes in infants enrolled in the NHBCS. Our longitudinal analysis of 572 infants found that each month earlier of introduction to rice cereal, which contains inorganic arsenic, was associated with increased risk of upper RTI, lower RTI, acute respiratory symptoms, and allergy. We also found positive associations between 16S alpha diversity of 6-week gut microbiome and respiratory infections in infancy. Microbial species profiled from metagenomics sequencing also were found to be associated with respiratory infections in infancy. Our analysis of six-week gut microbiome and vaccine response detected several bacterial species and metabolic pathways associated with PCP and TT antibody response. Lastly, we introduced HT-MMIOW, a hypothesis test procedure that detects microbiome mediation using UMAP to reduce the number of dimensions in the mediator and IOW to quantify the overall indirect effect. HT-MMIOW detected a statistically significant indirect effect of the six-week gut microbiome on the "causal" path between maternal antibiotic use during pregnancy and incidence of child allergy up to five years of age.

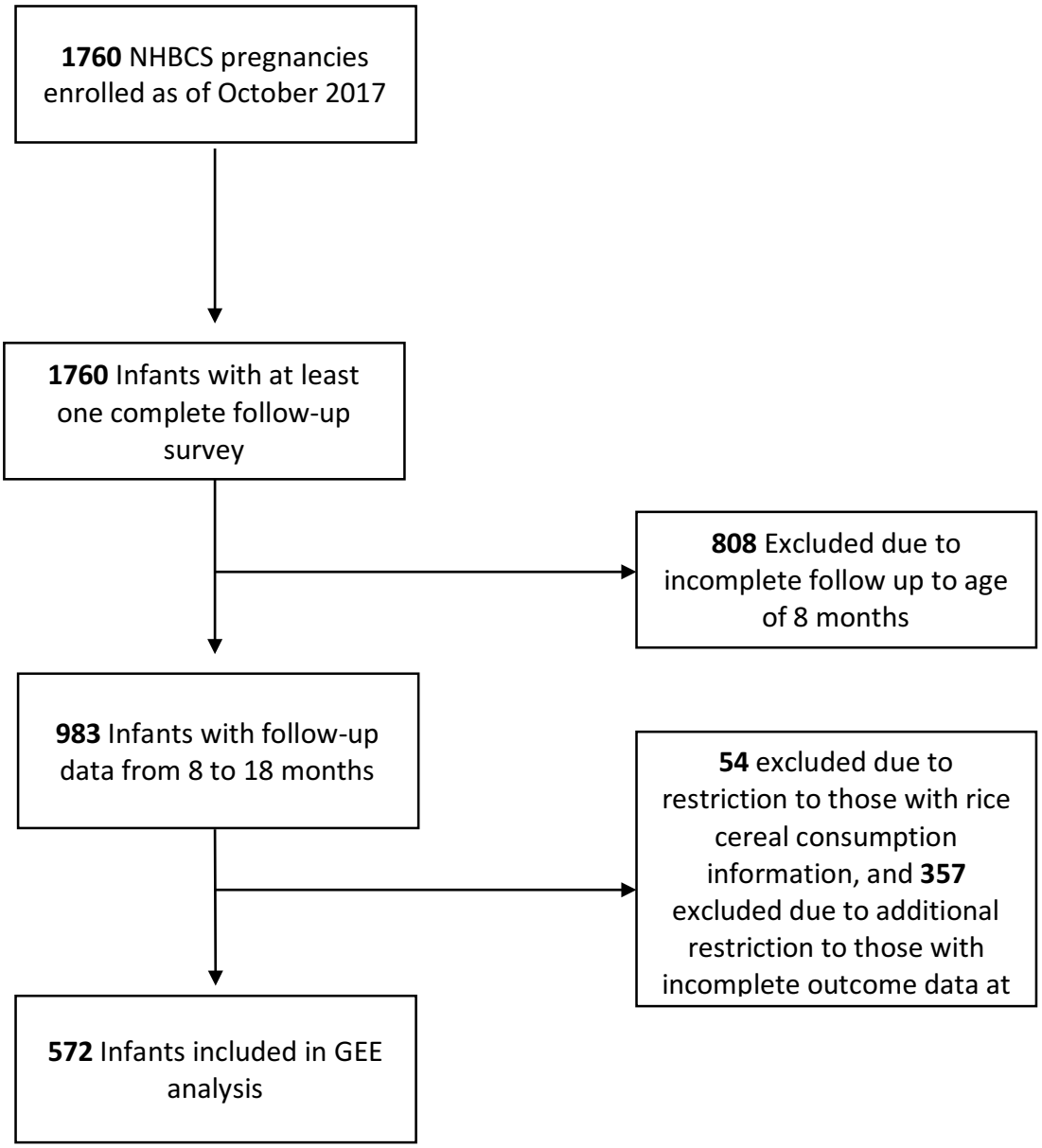
This thesis found associations between exposures, the early gut microbiome, and outcomes in infancy and early childhood; however, we cannot conclude causal relationships. We attempted to rule out associations that may be driven by other covariates with confounder adjustment in our analyses. In our mediation models assumed that unmeasured confounding did not have a large contribution to our risk ratios as this

assumption is required for performing mediation analysis with observational data. Further research is warranted to identify the mechanisms that explain the causal link among exposures, gut microbiota, and outcomes.

Infant infections are still a leading cause of morbidity and mortality in the US and other countries. Further, the widespread prevalence of allergy is of great public health concern. In addition to the health-related burden for the individual, allergies can pose a significant economic burden in terms of healthcare costs and loss of caregiver income.
Understanding the three-way interplay among exposures, the gut microbiome, and infant and child outcomes may aid the development of novel interventions or therapies aimed at modulating the gut microbiome to reduce the rates of infant morbidity and mortality with lifelong health impacts, such as infections and allergies.

Appendices

Supplementary Figure 2. 1: Flow Diagram of Study Participants with Inclusion and Exclusion Criteria



Abbreviations: *NHBCS*, New Hampshire Birth Cohort Study; *GEE*, generalized estimating equation.

Supplementary Table 2. 1: Selected Characteristics of Mothers and Infants (N = 411) in the New Hampshire Birth Cohort Study Not Included in Main Analyses Followed to Age 18 Months

Variable	Sample Size	Mean (SD) or No. (%)
Maternal Characteristics		
Smoking during any trimester of pregnancy, No. (%)	372	
Yes		55 (14.8%)
No		317 (85.2 %)
Relationship status, No. (%)	366	
Married		293 (80.1)
Single		56 (15.3)
Separated/Divorced		17 (4.6)
Highest level of educational attainment, No. (%)	367	
≤high school/GED		49 (13.4)[a]
Some college		70 (19.1)[a]
College graduate		135 (36.8) [a]
Postgraduate schooling		113 (30.8)[a]
BMI before pregnancy (kg/m^2), mean (SD)	398	26 (6.0)
Age at enrollment (years), mean (SD)	411	31.1 (4.9)
Arsenic in water (μg/L), mean (SD)	390	4.2 (13.5)
Water Arsenic > 5 μg/L, No. (%)	390	66 (16.9)
Infant Characteristics		
Sex, No. (%)	410	
Male		188 (45.9)
Female		222 (54.1)
Birth weight (g), mean (SD)	398	3447 (496)
Ever breast fed at 4 months, No. (%)	350	
Yes		333 (95.1)
No		17 (4.9)
Other solid food consumption at 4 months, No. (%) [c]	88	
Yes		78 (88.6)
No		10 (11.4)
Other solid food consumption at 8 months, No. (%) [c]	47	
Yes		26 (55.3)
No		21 (44.7)
Other solid food consumption at 12 months, No. (%) [c]	315	
Yes		53 (16.8)
No		262 (83.2)

[a] Percentages do not sum to 100 due to rounding

Supplementary Table 2. 2: Infant Rice Cereal Consumption and Immune-Related Outcome Prevalence

Variable	N	Mean (SD) or N (%)
Ever consumed rice cereal before age of 12 months	572	
Yes		435 (76.0)
No		137 (24.0)
Age of rice cereal introduction of those consumed before age of 12 months (months)	435	5.2 (1.3)
Rice cereal introduced before 4 months of age		27 (6.2)
Rice cereal introduced between 4-8 months of age		381 (87.6)
Rice cereal introduced between 8-12 months of age		27 (6.2)
Rice cereal consumption at 4 months[a]	358	
Yes		42 (11.7)
No		316 (88.3)
Time since rice cereal introduction at 4 months (months)	435	0.02 (0.1)
Rice cereal consumption at 8 months[a]	438	
Yes		305 (69.6)
No		133 (30.4)
Time since rice cereal introduction at 8 months (months)	572	1.4 (1.3)
Rice cereal consumption at 12 months[a]	544	
Yes		373 (68.6)
No		171 (31.4)
Time since rice cereal introduction at 12 months (months)	572	4.4 (2.7)
Infections or symptoms within 5-18 months of life	572	
At least one outcome		552 (96.5)
At least one outcome lasting ≥ 2 days		523 (91.4)
At least one outcome resulting in a doctor visit		373 (65.2)
At least one outcome treated with prescription medication		299 (52.3)
Allergy within 5-18 months of life	572	
At least one outcome		77 (13.5)
At least one outcome resulting in a doctor visit		45 (7.9)
Reported an allergy to peanuts		5 (0.9)

Reported an allergy to other nuts		3 (0.5)
Reported an allergy to eggs		7 (1.2)
Reported an allergy to other foods		39 (6.8)
Reported an allergy to antibiotics		21 (3.7)
Reported an allergy to cats or dogs		4 (0.7)
Reported an allergy to pollen		9 (1.6)
Reported an allergy to latex		2 (0.3)
Reported an allergy to dust		2 (0.3)
Reported an allergy to insect bites		2 (0.3)
Reported an allergy to grass		5 (0.9)

Abbreviations: *SD*, standard deviation.

[a]Percentage calculated using different sample sizes due to missing values. Sample sizes were 358, 438, and 544 for 4 months, 8 months, and 12 months respectively.

Supplementary Table 2. 3: Number of Immune-related Outcomes over 5-18 Months at Each Time Period, N = 572

Outcome	Time Period	Any report of the outcome No.(%)	Lasting ≥ 2 days No.(%)	Involving a doctor visit No.(%)	Requiring a prescription medication No.(%)
Upper Respiratory Infections	8 months	206 (74.9)	177 (64.4)	103 (37.5)	58 (21.1)
	12 months	340 (83.1)	292 (71.4)	155 (37.9)	119 (29.1)
	18 months	399 (90.9)	341 (77.7)	171 (38.9)	146 (33.3)
Lower Respiratory Infections	8 months	17 (6.2)	17 (6.2)	16 (5.8)	9 (3.3)
	12 months	28 (6.8)	27 (6.6)	26 (6.4)	18 (4.4)
	18 months	33 (7.5)	28 (6.4)	31 (7.1)	24 (5.5)
Acute Respiratory Symptoms	8 months	133 (48.4)	99 (36.0)	53 (19.3)	18 (6.5)
	12 months	202 (49.4)	143 (35.0)	70 (17.1)	31 (7.6)
	18 months	251 (57.2)	189 (43.1)	81 (18.5)	44 (10.0)
Diarrhea	8 months	46 (16.7)	15 (5.5)	11 (4.0)	0 (0.1)
	12 months	124 (30.3)	40 (9.8)	16 (3.9)	2 (0.5)
	18 months	159 (36.2)	57 (13.0)	18 (4.1)	1 (0.2)
Fever Symptoms [a]	8 months	101 (36.7)	32 (11.6)	36 (13.1)	4 (1.5)
	12 months	210 (51.3)	84 (20.5)	70 (17.1)	12 (2.9)
	18 months	252 (57.8)	95 (21.6)	85 (19.4)	23 (5.2)
Allergy	8 months	11 (4.0)	N/A[b]	6 (2.2)	N/A[b]
	12 months	27 (6.6)		17 (4.2)	
	18 months	53 (12.1)		37 (8.4)	

[a] Sample size N = 571 for fever analyses

[b] Participants only asked about any allergies and whether these allergies had been doctor diagnosed

Supplementary Table 2. 4: Adjusted Risk Ratio Estimates and 95% Confidence Intervals From GEE in Repeated Measures over 5-18 Months for One Month Earlier Introduction of Rice Cereal on Risk of Immune-related Outcomes, N = 572[a].

Outcome	Any report of the outcome RR (95% CI) N of total events	Lasting ≥ 2 days RR (95% CI) N of total events	Involving a doctor visit RR (95% CI) N of total events	Requiring a prescription medication RR (95% CI) N of total events
Upper Respiratory Infections	1.03 (1.02-1.04) 945	1.03 (1.01-1.05) 810	1.02 (0.98-1.06) 429	1.04 (1.00-1.09) 323
Lower Respiratory Infections	1.14 (1.01-1.29) 78	1.10 (0.97-1.24) 72	1.13 (1.00-1.29) 73	1.19 (1.02-1.39) 51
Acute Respiratory Symptoms	1.05 (1.02-1.08) 586	1.05 (1.01-1.10) 431	1.03 (0.96-1.10) 204	1.10 (1.00-1.22) 93
Diarrhea	1.08 (1.04-1.13) 329	1.05 (0.96-1.15) 112	0.89 (0.74-1.06) 45	[c] 3
Fever Symptoms [b]	1.04 (1.01-1.07) 563	1.03 (0.97-1.09) 211	1.07 (1.00-1.14) 191	1.22 (1.02-1.45) 39
Allergy	1.18 (1.07-1.30) 91	N/A [d]	1.20 (1.06-1.36) 60	N/A[d]

Abbreviations: RR, risk ratio; CI, confidence interval.

[a] GEE adjusted for smoking during pregnancy, maternal relationship status, maternal education, maternal pre-pregnancy BMI, maternal age of enrollment, arsenic concentrations in household tap water samples, infant birth weight, breastfeeding, and other solid food consumption. Risk ratios represent increased risk of health outcome with every month earlier of introduction to rice cereal.

[b] Sample size N = 571 for fever analyses

[c] Too few observations to perform analysis

[d] Participants only asked about any allergies and whether these allergies had been doctor diagnosed

Supplementary Table 2. 5: Crude Risk Ratio Estimates and 95% Confidence Intervals From GEE for all Other Covariates on Risk of Immune-related Outcomes, N = 572[a].

Variable	Upper Respiratory Infections	Lower Respiratory Infections	Acute Respiratory Symptoms	Diarrhea	Fever Symptoms[b]	Allergy
Intercept	0.21 (0.06, 0.66)	0.00 (0.00, 0.07)	0.02 (0.00, 0.20)	0.14 (0.00, 5.82)	0.01 (0.00, 0.26)	5.98 (0.17, 212.93)
Indicator of Rice Cereal Consumption	1.32 (0.98, 1.78)	0.70 (0.27, 1.83)	0.77 (0.41, 1.43)	2.35 (0.82, 6.72)	0.94 (0.24, 3.62)	1.42 (0.53, 3.84)
Other Solid Food Consumption	1.19 (0.89, 1.63)	1.07 (0.44, 2.57)	1.43 (0.83, 2.48)	2.22 ()0.86, 5.78	2.04 (0.66, 6.29)	2.68 (1.02, 7.05)
Smoking during any trimester of pregnancy	1.26 (0.91, 1.74)	0.80 (0.29, 2.18)	1.35 (0.67, 2.71)	1.36 (0.63, 2.94)	0.89 (0.35, 2.30)	0.56 (0.20, 1.59)
Relationship status – Married (baseline) VS Single	1.01 (0.65, 1.58)	0.95 (0.32, 2.80)	2.09 (1.01, 4.32)	1.91 (0.70, 5.19)	0.73 (0.19, 2.91)	0.89 (0.30, 2.63)
Relationship status – Married (baseline) VS Separated/Divorced	1.28 (0.76, 2.16)	0.86 (0.11, 6.58)	2.26 (0.83, 6.18)	1.41 (0.27, 7.39)	3.05 (0.87, 10.71)	1.90 (0.20, 17.77)
Highest level of educational attainment – ≤high school/GED (baseline) VS Some college	0.78 (0.50, 1.22)	0.82 (0.28, 2.39)	1.04 (0.46, 2.32)	0.98 (0.35, 2.73)	1.95 (0.43, 8.90)	0.51 (0.18, 1.43)
Highest level of educational attainment – ≤high school/GED (baseline) VS College graduate	1.03 (0.70, 1.52)	0.64 (0.24, 1.66)	1.04 (0.46, 2.38)	0.71 (0.24, 2.09)	1.47 (0.33, 6.55)	0.45 (0.17, 1.23)
Highest level of educational attainment – ≤high school/GED (baseline) VS Postgraduate schooling	0.89 (0.58, 1.36)	0.89 (0.34, 2.36)	1.48 (0.62, 3.52)	0.86 (0.23, 3.22)	1.31 (0.27, 6.32)	0.74 (0.26, 2.13)
BMI before pregnancy	1.00 (0.99, 1.02)	1.06 (1.02, 1.10)	1.02 (0.99, 1.06)	0.98 (0.92, 1.05)	0.99 (0.95, 1.04)	0.92 (0.87, 0.98)
Age at enrollment	1.00 (0.98, 1.02)	1.04 (0.98, 1.10)	1.02 (0.97, 1.06)	0.97 (0.90, 1.05)	1.05 (0.98, 1.11)	0.95 (0.89, 1.01)
Birth weight	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Ever breast fed at 4 months	1.25 (0.75, 2.08)	1.53 (0.39, 5.95)	1.71 (0.50, 5.87)	0.73 (0.15, 3.49)	0.58 (0.20, 1.65)	0.28 (0.11, 0.74)
Arsenic in water	0.99 (0.97, 1.01)	1.01 (0.98, 1.04)	1.01 (0.98, 1.04)	1.00 (0.97, 1.06)	0.88 (0.77, 1.01)	1.00 (0.97, 1.03)

[a] Outcomes include those treated with prescription medication for upper RTI, lower RTI, acute respiratory symptoms, and fever symptoms and those diagnosed by a doctor for diarrhea and allergy.

[b] Sample size N = 571 for fever analyses

Supplementary Table 2. 6: Crude Risk Ratio Estimates and 95% Confidence Intervals From GEE in Repeated Measures over 5-18 Months for One Month Earlier Introduction of Rice Cereal on Risk of Immune-related Outcomes, N = 572[a].

Outcome	Any report of the outcome RR (95% CI) N of total events	Lasting ≥ 2 days RR (95% CI) N of total events	Involving a doctor visit RR (95% CI) N of total events	Requiring a prescription medication RR (95% CI) N of total events
Upper Respiratory Infections	1.03 (1.01-1.04) 945	1.03 (1.01-1.04) 810	1.02 (0.98-1.05) 429	1.04 (0.99-1.08) 323
Lower Respiratory Infections	1.14 (1.01-1.28) 78	1.10 (0.97-1.24) 72	1.14 (1.00-1.29) 73	1.18 (1.01-1.38) 51
Acute Respiratory Symptoms	1.05 (1.02-1.08) 586	1.05 (1.01-1.09) 431	1.03 (0.96-1.09) 204	1.11 (1.01-1.23) 93
Diarrhea	1.08 (1.03-1.13) 329	1.06 (0.97-1.15) 112	0.89 (0.75-1.07) 45	[c] 3
Fever Symptoms [b]	1.03 (1.00-1.06) 563	1.02 (0.96-1.08) 211	1.06 (0.99-1.13) 191	1.25 (1.03-1.51) 39
Allergy	1.17 (1.07-1.29) 91	N/A [d]	1.19 (1.07-1.33) 60	N/A[d]

Abbreviations: RR, risk ratio; CI, confidence interval.

[a] Crude GEE. Risk ratios represent increased risk of health outcome with every month earlier of introduction to rice cereal.

[b] Sample size N = 571 for fever analyses

[c] Too few observations to perform analysis

[d] Participants only asked about any allergies and whether these allergies had been doctor diagnosed

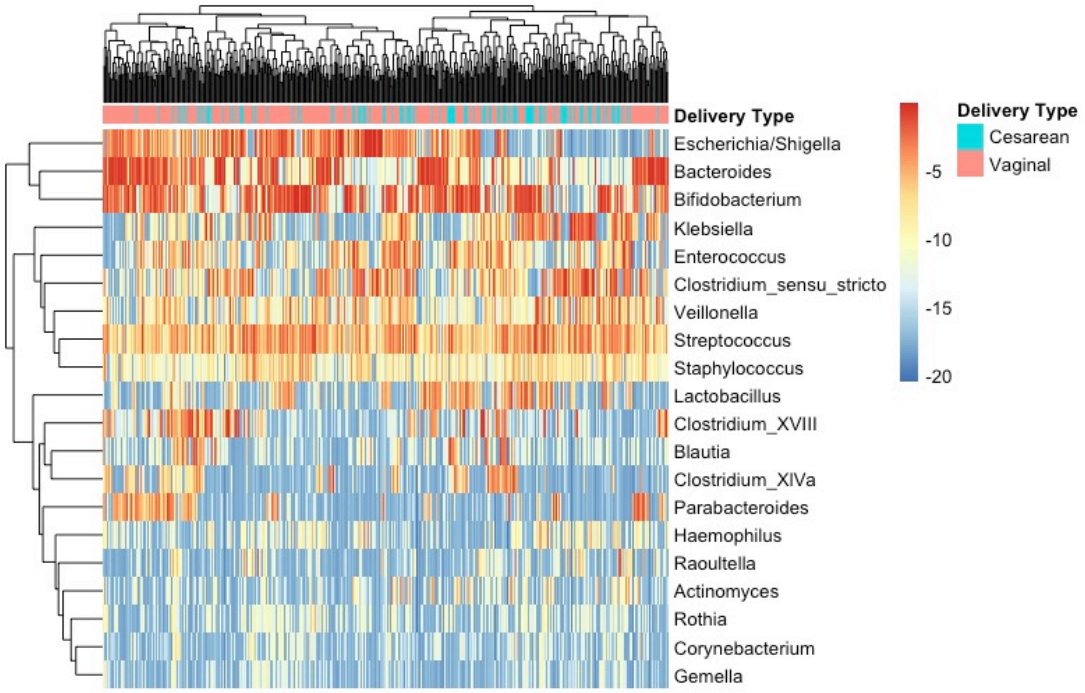
Supplementary Table 2. 7: Associations between Rice Cereal Introduction Age in Months and Confounders, N = 572[a]

Confounder	Estimate	Standard Error	P-Value
Smoking during any trimester of pregnancy			
Yes	0.1970	0.1970	0.0194
No (reference)			
Relationship status			
Married (reference)			
Single	-0.7455	0.2224	0.0009
Separated/Divorced	-1.3096	0.4715	0.0057
Highest level of educational attainment			
≤high school/GED (reference)			
Some college	0.2373	0.2563	0.3550
College graduate	0.4876	0.2317	0.0359
Postgraduate schooling	0.7134	0.2367	0.0027
BMI before pregnancy (kg/m^2)	-0.0255	0.0111	0.0223
Age at enrollment (years)	0.03234	0.0132	0.0150
Birth weight (g)	-0.0001	0.0001	0.2810
Ever breast fed at 4 months			
Yes	1.1638	0.3009	0.0001
No (reference)			

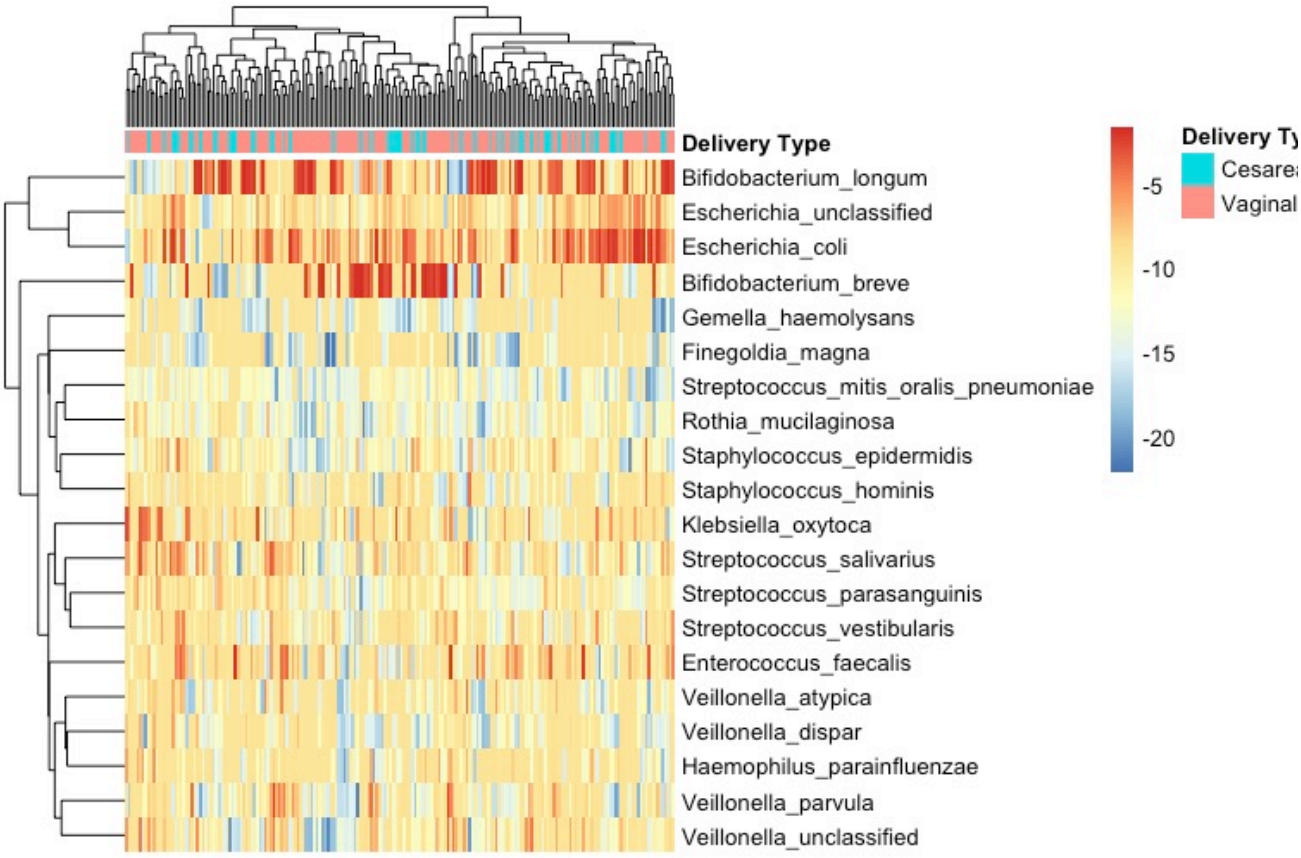
[a] Associations calculated using univariate linear models.

Supplementary Figure 3. 1: Composition of 6-week Gut Microbiome of Infants in the NHBCS

A. Heat map of 20 most common genera from 16S sequencing data (N = 465)



B. Heat map of 20 most common species from metagenomics sequencing data (N = 185)



Supplementary Table 3. 1: Number of Immune-related Outcomes over 12 Months at Each Time Period for 16S Analyses[a]

Outcome	Time Period	Overall N = 465	Vaginal Delivery N = 339	Cesarean Delivery N = 126
Any Infections or Symptoms	4 months	55	37	18
	8 months	189	133	56
	12 months	244	200	44
Upper Respiratory Infections	4 months	30	21	9
	8 months	109	79	30
	12 months	156	126	30
Lower Respiratory Infections	4 months	5	4	1
	8 months	26	18	8
	12 months	18	16	2
Wheezing	4 months	4	2	2
	8 months	19	14	5
	12 months	20	16	4
Diarrhea	4 months	7	5	2
	8 months	11	9	2
	12 months	18	13	5

[a] For any infections or symptoms, 18 entries were imputed. For upper RTI, 3 entries were imputed. For lower RTI, 3 entries were imputed.

Supplementary Table 3. 2**:** Number of Immune-related Outcomes over 12 Months at Each Time Period for Metagenomics Analyses[a]

Outcome	Time Period	Overall N = 185	Vaginal Delivery N = 129	Cesarean Delivery N = 26
Any Infections or Symptoms	4 months	16	11	5
	8 months	72	42	30
	12 months	119	96	23
Upper Respiratory Infections	4 months	12	10	2
	8 months	48	31	17
	12 months	72	58	14
Lower Respiratory Infections	4 months	0	0	0
	8 months	6	3	3
	12 months	8	7	1
Wheezing	4 months	1	0	1
	8 months	6	3	3
	12 months	11	8	3
Diarrhea	4 months	3	2	1
	8 months	4	3	1
	12 months	7	5	2

[a] For any infections or symptoms, 7 entries were imputed.

Supplementary Table 3. 3: Adjusted Relative Risk Estimates and 95% Confidence Intervals from GEE Analysis of Infant 6-Week Stool 16S V4-V5 rRNA Sequencing Alpha Diversity and Infections and Symptoms of Infection over the First Year of Life [a]

Outcome	Overall N = 465			Vaginal Delivery N = 339			Cesarean Delivery N = 126		
	No. of Outcomes[b]	RR	95% CI	No. of Outcomes[b]	RR	95% CI	No. of Outcomes[b]	RR	95% CI
Any Infection or Symptom[c]	488	1.39*	(1.1, 1.77)	370	1.62*	(1.23, 2.15)	118	0.95	(0.58, 1.55)
Upper RTI	295	1.40*	(1.12, 1.76)	226	1.85	(0.98, 3.51)	69	0.94	(0.27, 3.29)
Lower RTI	49	1.50	(0.87, 2.6)	38	1.89	(0.96, 3.72)	11	0.94	(0.28, 3.21)
Wheezing	43	1.30	(0.81, 2.07)	32	2.00*	(1.16, 3.45)	11	0.47	(0.20, 1.12)
Diarrhea[d]	36	1.44	(0.92, 2.25)	27	1.86*	(1.14, 3.03)	9	0.74	(0.32, 1.69)

Abbreviations: GEE, generalized estimating equation; N, sample size; No., number; RR, relative risk; CI, confidence interval; RTI, respiratory tract infection.

*indicates statistical significance at $\alpha = 0.05$.

[a] Overall GEE adjusted for maternal BMI, delivery type, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. GEE stratified by delivery mode (vaginal and cesarean) adjusted for maternal BMI, sex, breast feeding at six weeks, perinatal antibiotic use, and gestational age. Relative risk estimates represent an increased risk of having an additional infection or symptom of infection or an increased risk of experiencing wheezing or diarrhea with each doubling of the inverse Simpson index. Upper RTI, lower RTI, and wheezing outcomes are those diagnosed by a physician for which a medication was prescribed. Diarrhea outcomes are those diagnosed by a physician for which no medication was prescribed.

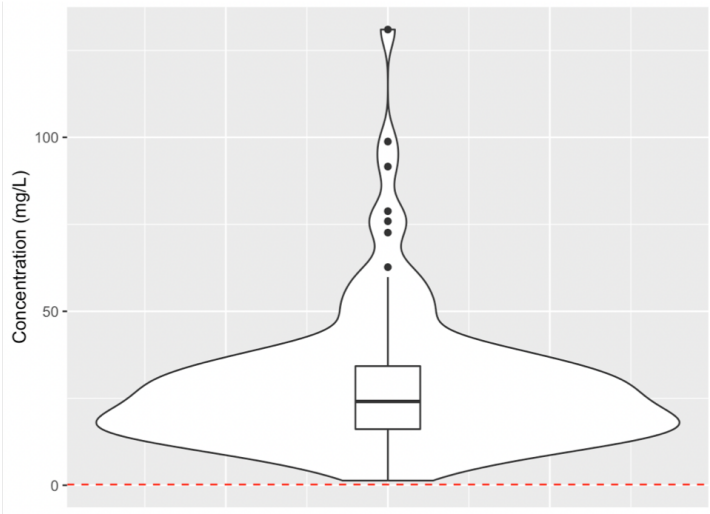
[b] Total number of outcomes may be greater than N due to repeated measures.

[c] Any infection or symptom is the sum of upper respiratory tract infections (RTI), lower RTI, and acute respiratory symptoms.

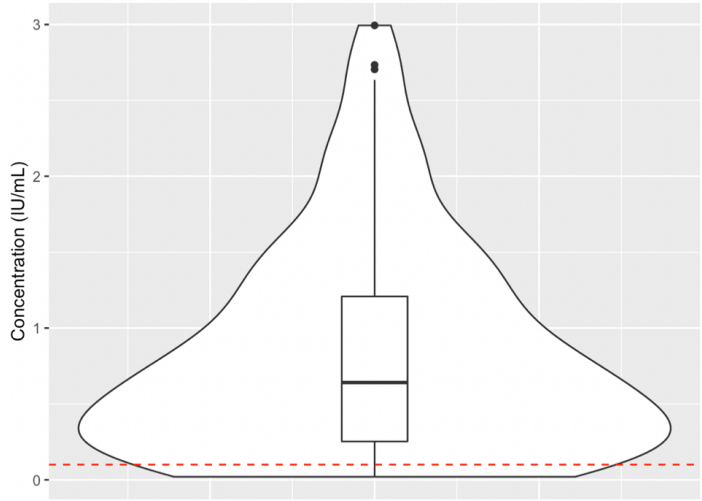
[d] Sample sizes N = 464 for overall and N = 125 for cesarean delivery for diarrhea analyses due to missing data.

Supplementary Figure 4. 1: PCP and TT IgG antibody concentrations.

a. PCP



b. TT



Violin plots of PCP and TT IgG concentration indicating concentration quartiles. Red dotted line indicates preferred protection threshold of 0.2mg/L for PCP and 0.1IU/mL for TT.

Supplementary Table 4. 1: Bacterial species associated with vaccine response after correction for multiple comparisons

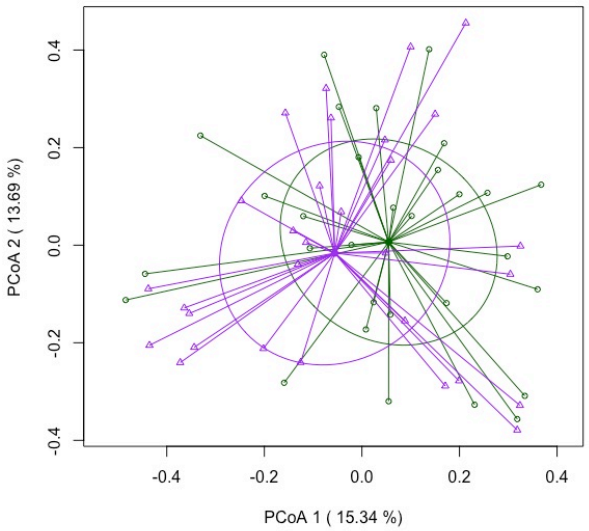
Outcome	Species	Estimate	P-value	95% CI
TT	*Aeriscardovia aeriphila*	-0.037	< 0.001	(-0.057, -0.017)

Supplementary Table 4. 2: FDR selected metabolic pathways associated with vaccine response

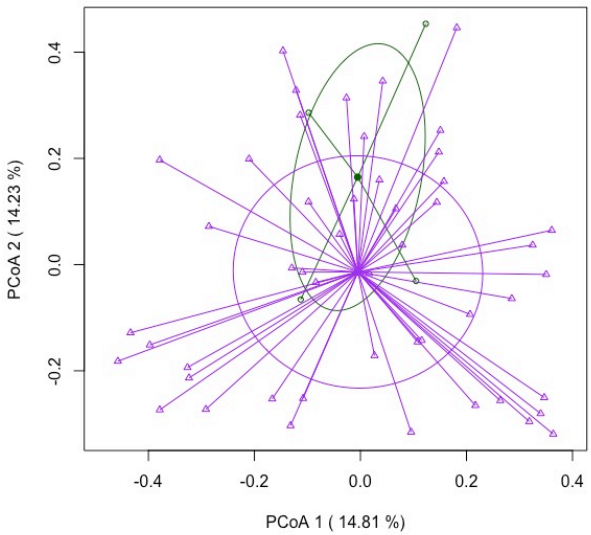
Outcome	Pathway	Estimate	P-value	95% CI
PCP	PWY-3781: aerobic respiration I (cytochrome c)	-0.271	0.045	(-0.530, -0.013)
PCP	PWY-4242: pantothenate and coenzyme A biosynthesis III	0.263	0.046	(0.010, 0.515)
PCP	PWY-6628: superpathway of L-phenylalanine biosynthesis	-0.238	0.057	(-0.478, 0.002)
PCP	PWY-7209: superpathway of pyrimidine ribonucleosides degradation	0.305	0.021	(0.053, 0.556)
PCP	PWY-7211: superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis	-0.275	0.034	(-0.522, -0.027)
PCP	PWY-7237: myo-, chiro- and scillo-inositol degradation	-0.239	0.058	(-0.482, 0.003)
PCP	PWY-7399: methylphosphonate degradation II	0.334	0.007	(0.100, 0.567)
PCP	PWY0-162: superpathway of pyrimidine ribonucleotides de novo biosynthesis	0.323	0.011	(0.084, 0.562)
PCP	PWY66-367: ketogenesis	-0.234	0.063	(-0.476, 0.008)
TT	PWY-5667: CDP-diacylglycerol biosynthesis I	0.882	< 0.001	(0.468, 1.295)
TT	PWY-6353: purine nucleotides degradation II (aerobic)	0.707	< 0.001	(0.284, 1.13)
TT	PWY-6700: queuosine biosynthesis	0.784	< 0.001	(0.368, 1.199)
TT	PWY0-1319: CDP-diacylglycerol biosynthesis II	0.881	< 0.001	(0.468, 1.294)
TT	SALVADEHYPOX-PWY: adenosine nucleotides degradation II	0.706	0.002	(0.282, 1.129)

Supplementary Figure 4. 2: Sensitivity analysis PCoA plots of bacterial 16S V4-V5 rRNA sequencing Bray-Curtis dissimilarity for PCP and TT

a. PCP (PERMANOVA P = 0.688)



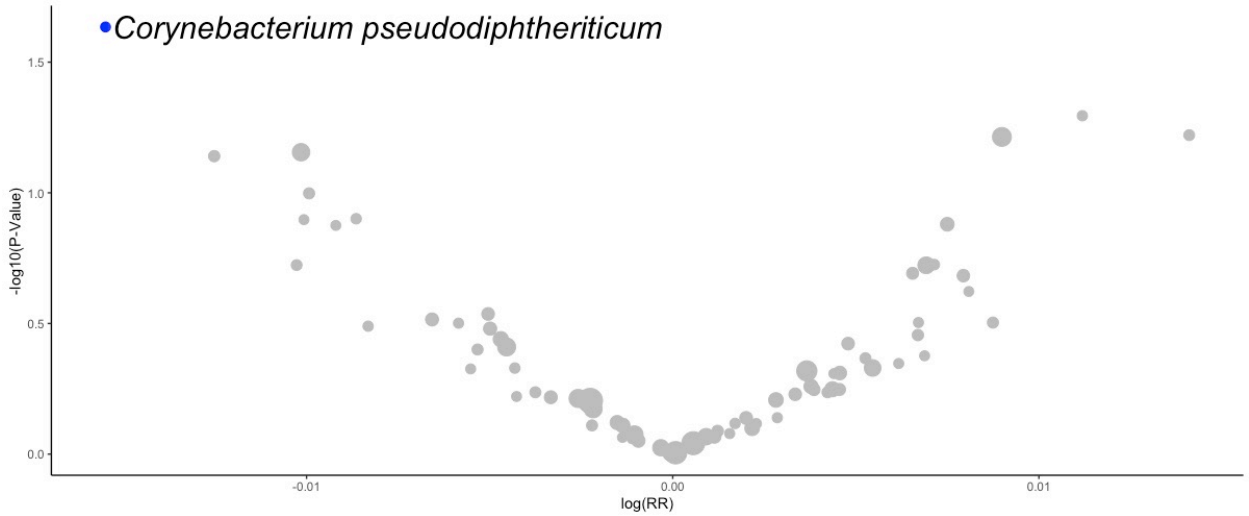
b. TT (PERMANOVA P = 0.112)



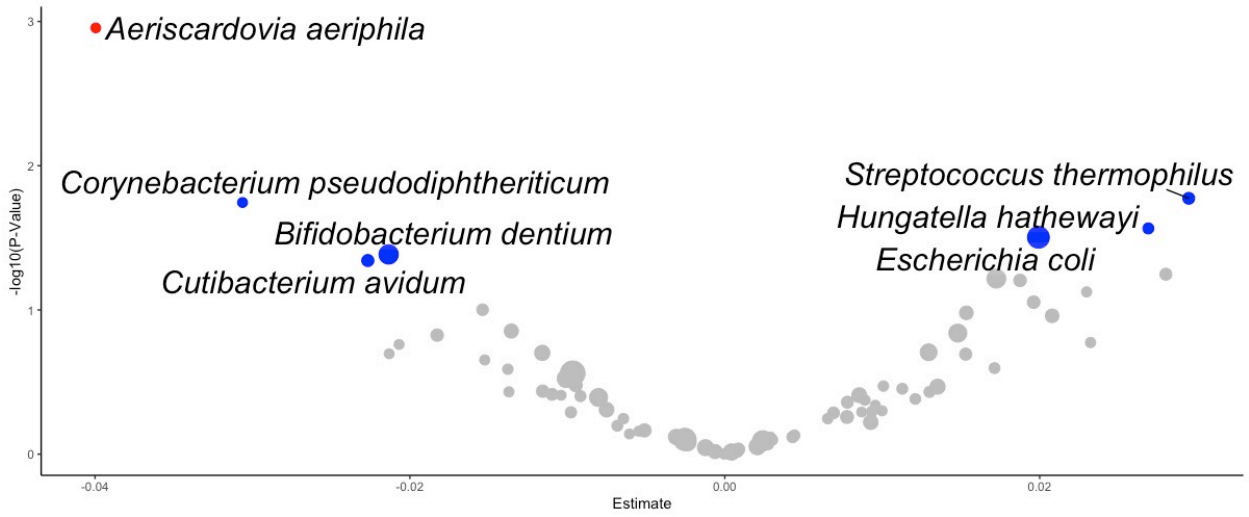
PCP groups assigned by median PCP IgG concentration threshold. TT groups assigned by preferred protection threshold of 0.1IU/mL. Percentages on the X and Y axis of plots represent percentage of variance explained by first two eigenvectors.

Supplementary Figure 4. 3: Sensitivity analysis associations between metagenomics bacterial species and vaccine response.

a. PCP

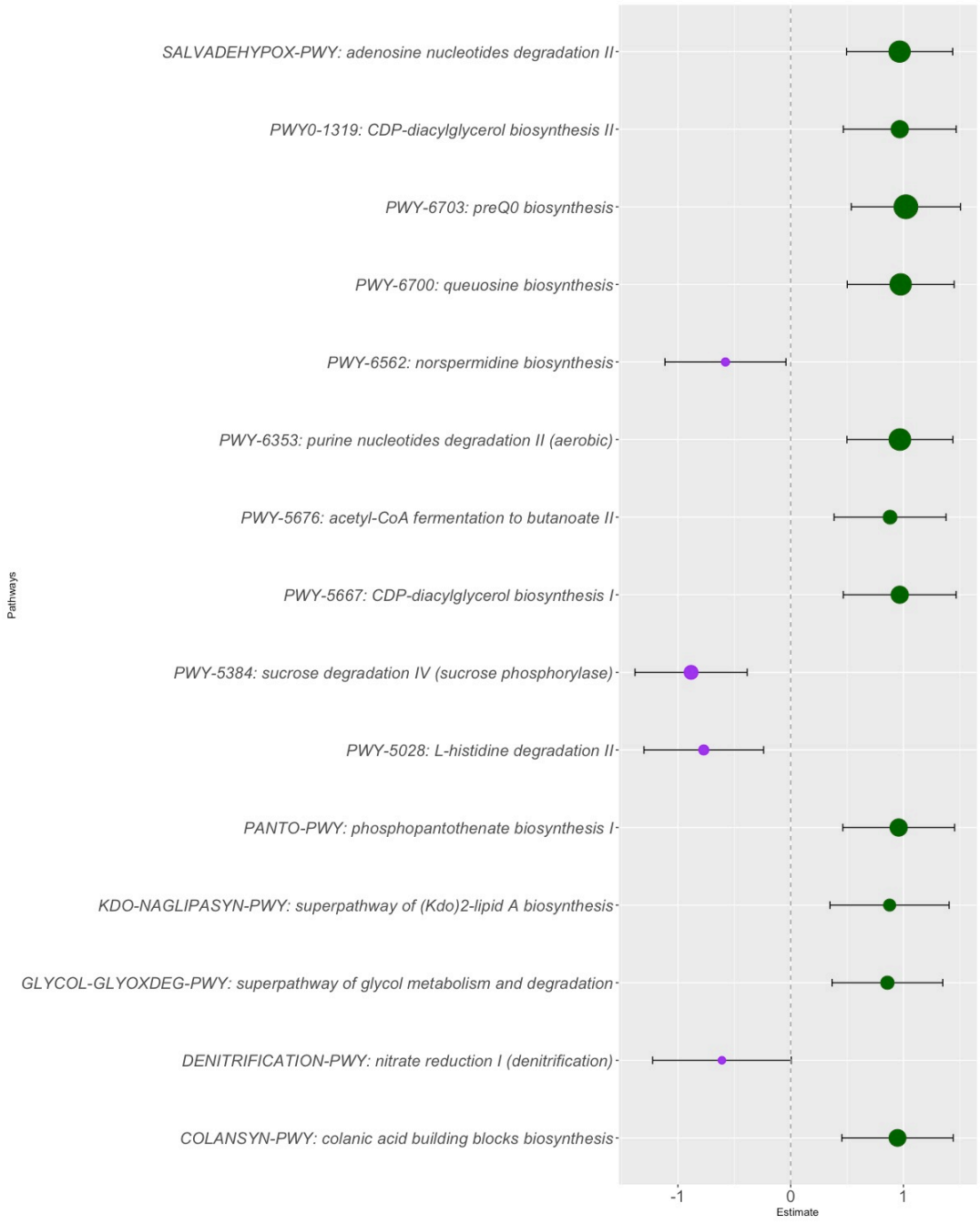


b. TT



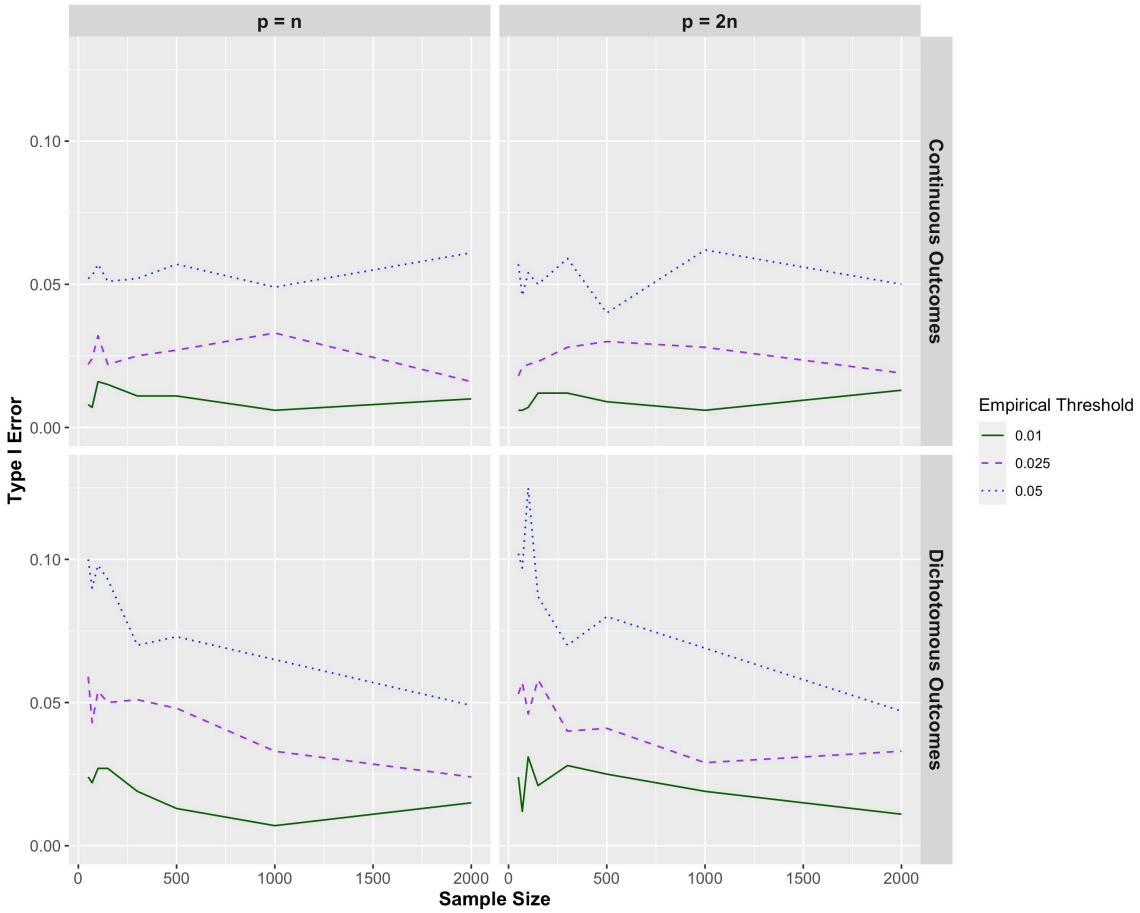
Dots indicate bacterial species, and size of dots vary by mean abundance. Blue indicates species with p-value < 0.05. Red indicates species with p-values <0.05 and meet FDR correction.

Supplementary Figure 4. 4: Sensitivity analysis associations between elastic-net and metabolic pathways and TT response.



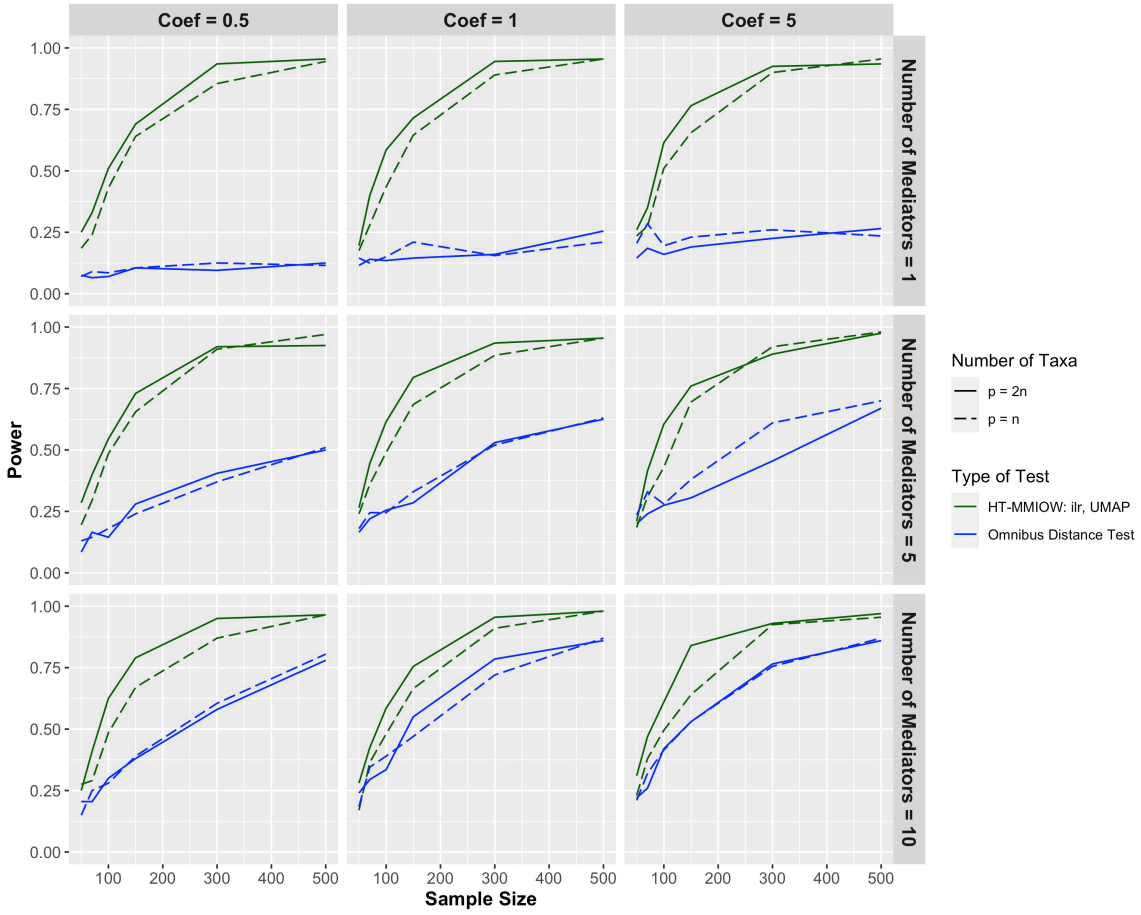
No associations were found for PCP response. Dots indicates effect size, and horizontal bands indicate 95% CI. Green dots represent positive association, while purple dots represent negative association. Size of dots vary by p-value: larger dot indicates smaller p-value. Only pathways selected by elastic net and FDR correction shown here.

Supplementary Figure 5. 1: Type I error calculations for varying sample size, number of taxa, and type of outcome.



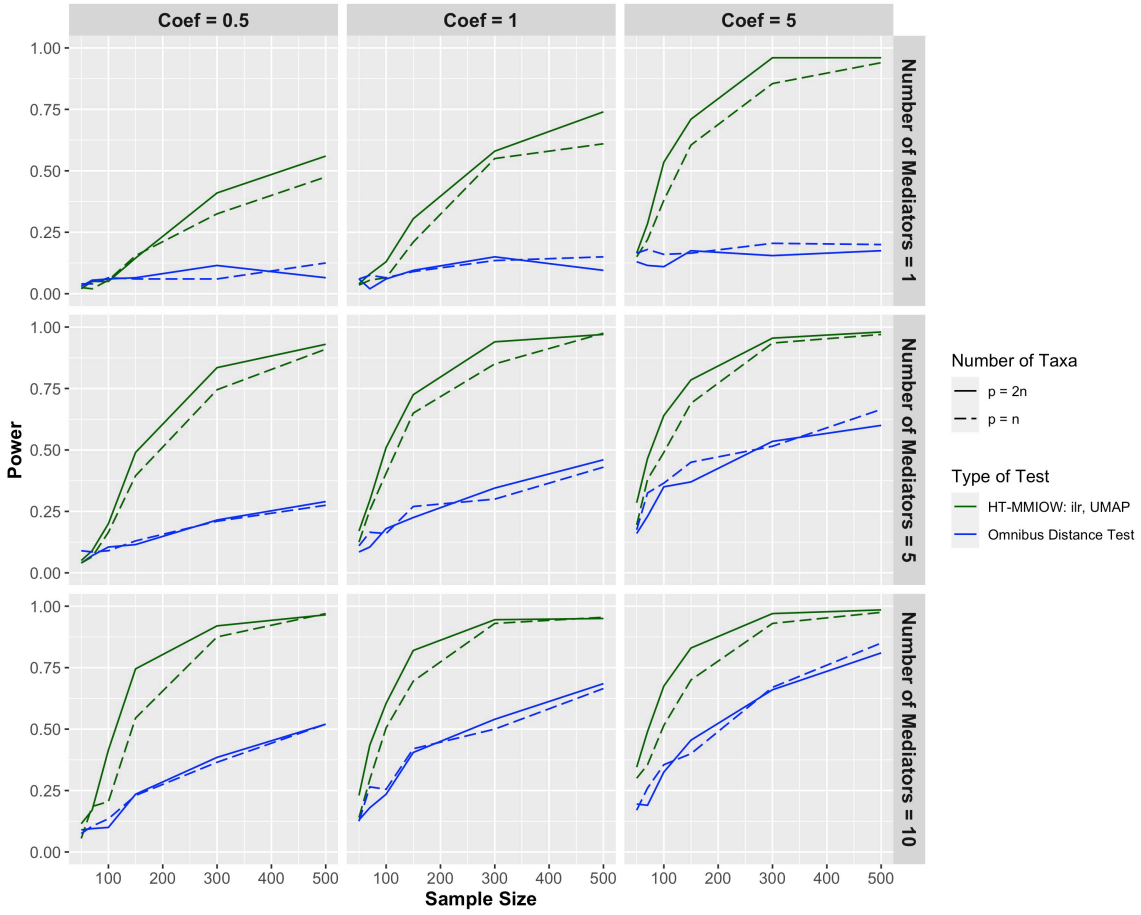
Results are based on 1000 simulations. Each line represents type I error at varying empirical thresholds.

Supplementary Figure 5. 2: Power calculations for HT-MMIOW with a 0.01 threshold and the omnibus distance test for continuous outcomes, with varying sample size, effect size, and number of true mediators and number of taxa.



Results are based on 200 simulations. Solid lines indicate $p = 2n$, and dashed lines indicate $p = n$. Green lines represent HT-MMIOW with a 0.01 empirical threshold, and blue lines represent the omnibus distance test.

Supplementary Figure 5. 3: Power calculations for HT-MMIOW with a 0.01 threshold and the omnibus distance test for dichotomous outcomes, with varying sample size, effect size, and number of true mediators and number of taxa.



Results are based on 200 simulations. Solid lines indicate $p = 2n$, and dashed lines indicate $p = n$. Green lines represent HT-MMIOW with a 0.01 empirical threshold, and blue lines represent the omnibus distance test.

References

1. Gupta, R. *et al.* The Economic Impact of Childhood Food Allergy in the United States. *JAMA Pediatrics* **167**, 1026–1031 (2013).
2. Amand, C., Tong, S., Kieffer, A. & Kyaw, M. H. Healthcare resource use and economic burden attributable to respiratory syncytial virus in the United States: a claims database analysis. *BMC Health Services Research* **18**, 294 (2018).
3. Ladomenou, F., Moschandreas, J., Kafatos, A., Tselentis, Y. & Galanakis, E. Protective effect of exclusive breastfeeding against infections during infancy: a prospective study. *Archives of Disease in Childhood* **95**, 1004–1008 (2010).
4. Duijts, L., Jaddoe, V. W. V., Hofman, A. & Moll, H. A. Prolonged and Exclusive Breastfeeding Reduces the Risk of Infectious Diseases in Infancy. *Pediatrics* **126**, e18–e25 (2010).
5. Christensen, N., Søndergaard, J., Christesen, H. T., Fisker, N. & Husby, S. Association Between Mode of Delivery and Risk of Infection in Early Childhood: A Cohort Study. *The Pediatric Infectious Disease Journal* **37**, 316–323 (2018).
6. Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: a pooled analysis. *The Lancet* **355**, 451–455 (2000).
7. Victora, C. G. *et al.* Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet* **387**, 475–490 (2016).
8. Martino, D. & Prescott, S. Epigenetics and Prenatal Influences on Asthma and Allergic Airways Disease. *Chest* **139**, 640–647 (2011).

9. Lack, G. Update on risk factors for food allergy. *Journal of Allergy and Clinical Immunology* **129**, 1187–1197 (2012).
10. van Neerven, R. J. J. & Savelkoul, H. Nutrition and Allergic Diseases. *Nutrients* **9**, 762 (2017).
11. Risnes, K. R., Belanger, K., Murk, W. & Bracken, M. B. Antibiotic Exposure by 6 Months and Asthma and Allergy at 6 Years: Findings in a Cohort of 1,401 US Children. *American Journal of Epidemiology* **173**, 310–318 (2011).
12. Kronman, M. P., Zaoutis, T. E., Haynes, K., Feng, R. & Coffin, S. E. Antibiotic Exposure and IBD Development Among Children: A Population-Based Cohort Study. *Pediatrics* **130**, e794–e803 (2012).
13. Hoskin-Parr, L., Teyhan, A., Blocker, A. & Henderson, A. J. W. 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* **24**, 762–771 (2013).
14. Metsälä, J. *et al.* Mother's and Offspring's Use of Antibiotics and Infant Allergy to Cow's Milk. *Epidemiology* **24**, 303–309 (2013).
15. Miller, J. E. *et al.* Maternal antibiotic exposure during pregnancy and hospitalization with infection in offspring: a population-based cohort study. *International Journal of Epidemiology* **47**, 561–571 (2018).
16. Baron, R. *et al.* The relationship of prenatal antibiotic exposure and infant antibiotic administration with childhood allergies: a systematic review. *BMC Pediatrics* **20**, 312 (2020).

17. Mozaffarian, D. & Rimm, E. B. Fish Intake, Contaminants, and Human Health: Evaluating the Risks and the Benefits. *JAMA* **296**, 1885–1899 (2006).
18. Järup, L. Hazards of heavy metal contamination. *Br Med Bull* **68**, 167–182 (2003).
19. Williams, P. N. *et al.* Greatly Enhanced Arsenic Shoot Assimilation in Rice Leads to Elevated Grain Levels Compared to Wheat and Barley. *Environ. Sci. Technol.* **41**, 6854–6859 (2007).
20. O'Donovan, S. M. *et al.* Adherence with early infant feeding and complementary feeding guidelines in the Cork BASELINE Birth Cohort Study. *Public Health Nutr* **18**, 2864–2873 (2015).
21. Karagas, M. R. *et al.* Association of Rice and Rice-Product Consumption With Arsenic Exposure Early in Life. *JAMA Pediatr* **170**, 609–616 (2016).
22. Dangleben, N. L., Skibola, C. F. & Smith, M. T. Arsenic immunotoxicity: a review. *Environ Health* **12**, 73 (2013).
23. Farzan, S. F. *et al.* Infant Infections and Respiratory Symptoms in Relation to in Utero Arsenic Exposure in a U.S. Cohort. *Environ Health Perspect* **124**, 840–847 (2016).
24. Signes-Pastor, A. J. *et al.* Infants' dietary arsenic exposure during transition to solid food. *Scientific Reports* **8**, 7114 (2018).
25. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science* **352**, 539–544 (2016).
26. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Research* **30**, 492–506 (2020).

27. Coker, M. & Madan, J. C. The microbiome and immune system development. in *The Developing Microbiome* 43–66 (Academic Press, 2020).
28. Gaboriau-Routhiau, V. *et al.* The Key Role of Segmented Filamentous Bacteria in the Coordinated Maturation of Gut Helper T Cell Responses. *Immunity* **31**, 677–689 (2009).
29. Chung, H. *et al.* Gut Immune Maturation Depends on Colonization with a Host-Specific Microbiota. *Cell* **149**, 1578–1593 (2012).
30. Umesaki, Y., Setoyama, H., Matsumoto, S., Imaoka, A. & Itoh, K. Differential Roles of Segmented Filamentous Bacteria and Clostridia in Development of the Intestinal Immune System. *Infection and Immunity* **67**, 3504–3511 (1999).
31. Schnupf, P., Gaboriau-Routhiau, V. & Cerf-Bensussan, N. Host interactions with Segmented Filamentous Bacteria: An unusual trade-off that drives the post-natal maturation of the gut immune system. *Seminars in Immunology* **25**, 342–351 (2013).
32. Atarashi, K. *et al.* Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. *Science* **331**, 337–341 (2011).
33. Tan, T. G. *et al.* Identifying species of symbiont bacteria from the human gut that, alone, can induce intestinal Th17 cells in mice. *Proc Natl Acad Sci U S A* **113**, E8141–E8150 (2016).
34. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of RORγ+ regulatory T cells. *Science* **349**, 993–997 (2015).
35. Round, J. L. & Mazmanian, S. K. The gut microbiome shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313–323 (2009).

36. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).

37. Sokol, H. *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731–16736 (2008).

38. Cervantes-Barragan, L. *et al.* Lactobacillus reuteri induces gut intraepithelial CD4+CD8αα+ T cells. *Science* **357**, 806–810 (2017).

39. Das, G. *et al.* An important regulatory role for CD4+CD8αα T cells in the intestinal epithelial layer in the prevention of inflammatory bowel disease. *PNAS* **100**, 5324–5329 (2003).

40. Khosravi, A. *et al.* Gut microbiota promotes hematopoiesis tocontrol bacterial infection. *Cell Host Microbe* **15**, 374–381 (2014).

41. Belkaid, Y. & Hand, T. Role of the Microbiota in Immunity and inflammation. *Cell* **157**, 121–141 (2014).

42. Milani, C. *et al.* The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews* (2017) doi:10.1128/MMBR.00036-17.

43. Prince, B. T., Mandel, M. J., Nadeau, K. & Singh, A. M. Gut Microbiome and the Development of Food Allergy and Allergic Disease. *Pediatr Clin North Am* **62**, 1479–1492 (2015).

44. Cuello-Garcia, C. A. *et al.* Probiotics for the prevention of allergy: A systematic review and meta-analysis of randomized controlled trials. *Journal of Allergy and Clinical Immunology* **136**, 952–961 (2015).

45. Chang, Y.-S. *et al.* Synbiotics for Prevention and Treatment of Atopic Dermatitis: A Meta-analysis of Randomized Clinical Trials. *JAMA Pediatrics* **170**, 236–242 (2016).

46. Ren, X. *et al.* The distribution characteristics of intestinal microbiota in children with community-acquired pneumonia under five Years of age. *Microbial Pathogenesis* **142**, 104062 (2020).

47. Reyman, M. *et al.* Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature Communications* **10**, 4997 (2019).

48. Hasegawa, K. *et al.* The Fecal Microbiota Profile and Bronchiolitis in Infants. *Pediatrics* **138**, (2016).

49. Huda, M. N. *et al.* Stool Microbiota and Vaccine Responses of Infants. *Pediatrics* **134**, e362–e372 (2014).

50. Harris, V. C. *et al.* Significant Correlation Between the Infant Gut Microbiome and Rotavirus Vaccine Response in Rural Ghana. *The Journal of Infectious Diseases* **215**, 34–41 (2017).

51. Harris, V. *et al.* Rotavirus vaccine response correlates with the infant gut microbiota composition in Pakistan. *Gut Microbes* **9**, 93–101 (2018).

52. Huda, M. N. *et al.* Bifidobacterium Abundance in Early Infancy and Vaccine Response at 2 Years of Age. *Pediatrics* **143**, (2019).

53. Hasan, N. & Yang, H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* **7**, e7502 (2019).

54. Claus, S. P., Guillou, H. & Ellero-Simatos, S. The gut microbiota: a major player in the toxicity of environmental pollutants? *npj Biofilms Microbiomes* **2**, 1–11 (2016).

55. Coryell, M., Roggenbeck, B. A. & Walk, S. T. The Human Gut Microbiome's Influence on Arsenic Toxicity. *Curr Pharmacol Rep* **5**, 491–504 (2019).

56. Baldassarre, M. E. *et al.* Dysbiosis and Prematurity: Is There a Role for Probiotics? *Nutrients* **11**, 1273 (2019).

57. Depner, M. *et al.* Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. *Nat Med* **26**, 1766–1775 (2020).

58. Tun, H. M. *et al.* Roles of Birth Mode and Infant Gut Microbiota in Intergenerational Transmission of Overweight and Obesity From Mother to Offspring. *JAMA Pediatrics* **172**, 368–377 (2018).

59. Lynn, M. A. *et al.* Early-Life Antibiotic-Driven Dysbiosis Leads to Dysregulated Vaccine Immune Responses in Mice. *Cell Host & Microbe* **23**, 653-660.e5 (2018).

60. Gonzalez-Perez, G. *et al.* Maternal Antibiotic Treatment Impacts Development of the Neonatal Intestinal Microbiome and Antiviral Immunity. *The Journal of Immunology* **196**, 3768–3779 (2016).

61. Patrick, D. M. *et al.* Decreasing antibiotic use, the gut microbiota, and asthma incidence in children: evidence from population-based and prospective cohort studies. *The Lancet Respiratory Medicine* **8**, 1094–1105 (2020).

62. Zeng, P., Shao, Z. & Zhou, X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Computational and Structural Biotechnology Journal* **19**, 3209–3224 (2021).
63. Baron, R. M. & Kenny, D. A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* **51**, 1173–1182 (1986).
64. MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G. & Sheets, V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* **7**, 83–104 (2002).
65. Robins, J. M. & Greenland, S. Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology* **3**, 143–155 (1992).
66. Pearl, J. Direct and indirect effects. in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence* 411–420 (Morgan Kaufmann Publishers Inc., 2001).
67. Rubin, D. B. Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association* **100**, 322–331 (2005).
68. Imai, K. & Yamamoto, T. Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. *Political Analysis* **21**, 141–171 (2013).
69. VanderWeele, T. & Vansteelandt, S. Mediation Analysis with Multiple Mediators. *Epidemiologic Methods* **2**, 95–115 (2014).

70. Huang, Y.-T. & Pan, W.-C. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics* **72**, 402–413 (2016).
71. Chén, O. Y. *et al.* High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics* **19**, 121–136 (2018).
72. Zhang, Q. High-Dimensional Mediation Analysis with Applications to Causal Gene Identification. *Stat Biosci* (2021) doi:10.1007/s12561-021-09328-0.
73. Zhao, Y., Lindquist, M. A. & Caffo, B. S. Sparse principal component based high-dimensional mediation analysis. *Computational Statistics & Data Analysis* **142**, 106835 (2020).
74. Boca, S. M., Sinha, R., Cross, A. J., Moore, S. C. & Sampson, J. N. Testing multiple biological mediators simultaneously. *Bioinformatics* **30**, 214–220 (2014).
75. Millstein, J., Chen, G. K. & Breton, C. V. cit: hypothesis testing software for mediation analysis in genomic applications. *Bioinformatics* **32**, 2364–2365 (2016).
76. Sampson, J. N., Boca, S. M., Moore, S. C. & Heller, R. FWER and FDR control when testing multiple mediators. *Bioinformatics* **34**, 2418–2424 (2018).
77. Djordjilović, V. *et al.* Global test for high-dimensional mediation: Testing groups of potential mediators. *Statistics in Medicine* **38**, 3346–3360 (2019).
78. Liu, Z. *et al.* Large-Scale Hypothesis Testing for Causal Mediation Effects with Applications in Genome-wide Epigenetic Studies. *Journal of the American Statistical Association* **117**, 67–81 (2022).

79. Dai, J. Y., Stanford, J. L. & LeBlanc, M. A Multiple-Testing Procedure for High-Dimensional Mediation Hypotheses. *Journal of the American Statistical Association* **117**, 198–213 (2022).

80. Gao, Y. *et al.* Testing Mediation Effects in High-Dimensional Epigenetic Studies. *Frontiers in Genetics* **10**, (2019).

81. Song, Y. *et al.* Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. *Biometrics* **76**, 700–710 (2020).

82. Sohn, M. B. & Li, H. Compositional mediation analysis for microbiome studies. *The Annals of Applied Statistics* **13**, 661–681 (2019).

83. Zhang, J., Wei, Z. & Chen, J. A distance-based approach for testing the mediation effect of the human microbiome. *Bioinformatics* **34**, 1875–1883 (2018).

84. Wang, C., Hu, J., Blaser, M. J. & Li, H. Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data. *Bioinformatics* **36**, 347–355 (2020).

85. Zhang, H. *et al.* Mediation effect selection in high-dimensional and compositional microbiome data. *Statistics in Medicine* **40**, 885–896 (2021).

86. Zhang, H., Chen, J., Li, Z. & Liu, L. Testing for Mediation Effect with Application to Human Microbiome Data. *Stat Biosci* **13**, 313–328 (2021).

87. Wu, Q. *et al.* MedZIM: Mediation analysis for Zero-Inflated Mediators with applications to microbiome data. *arXiv:1906.09175 [stat]* (2021).

88. Sohn, M. B., Lu, J. & Li, H. A compositional mediation model for a binary outcome: Application to microbiome studies. *Bioinformatics* **38**, 16–21 (2022).

89. Simon, A. K., Hollander, G. A. & McMichael, A. Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci* **282**, (2015).

90. Holt, P. G. & Jones, C. A. The development of the immune system during pregnancy and early life. *Allergy* **55**, 688–697 (2000).

91. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4340604/.

92. Mennella, J. A., Ziegler, P., Briefel, R. & Novak, T. Feeding Infants and Toddlers Study: The Types of Foods Fed to Hispanic Infants and Toddlers. *Journal of the American Dietetic Association* **106**, 96–106 (2006).

93. Meharg, A. A. *et al.* Inorganic arsenic levels in baby rice are of concern. *Environmental Pollution* **152**, 746–749 (2008).

94. Lai, P. Y., Cottingham, K. L., Steinmaus, C., Karagas, M. R. & Miller, M. D. Arsenic and Rice: Translating Research to Address Health Care Providers' Needs. *J Pediatr* **167**, 797–803 (2015).

95. Zhao, F.-J., McGrath, S. P. & Meharg, A. A. Arsenic as a Food Chain Contaminant: Mechanisms of Plant Uptake and Metabolism and Mitigation Strategies. *Annual Review of Plant Biology* **61**, 535–559 (2010).

96. Punshon, T. *et al.* Understanding arsenic dynamics in agronomic systems to predict and prevent uptake by crop plants. *Science of The Total Environment* **581–582**, 209–220 (2017).

97. US EPA. Monosodium Methanearsonate (MSMA), an Organic Arsenical. *US EPA* https://www.epa.gov/ingredients-used-pesticide-products/monosodium-methanearsonate-msma-organic-arsenical (2015).

98. Dickens, R. & Hiltbold, A. E. Movement and Persistence of Methanearsonates in Soil. *Weeds* **15**, 299–304 (1967).

99. U.S. FDA. FDA Makes Available Results from Testing of Infant Rice Cereal for Inorganic Arsenic. *FDA* http://www.fda.gov/food/cfsan-constituent-updates/fda-makes-available-results-testing-infant-rice-cereal-inorganic-arsenic (2020).

100. Naujokas, M. F. *et al.* The Broad Scope of Health Effects from Chronic Arsenic Exposure: Update on a Worldwide Public Health Problem. *Environ Health Perspect* **121**, 295–302 (2013).

101. Ahmed, S. *et al.* In Utero Arsenic Exposure Is Associated With Impaired Thymic Function in Newborns Possibly Via Oxidative Stress and Apoptosis. *Toxicol Sci* **129**, 305–314 (2012).

102. Tregoning, J. S. & Schwarze, J. Respiratory Viral Infections in Infants: Causes, Clinical Symptoms, Virology, and Immunology. *Clinical Microbiology Reviews* **23**, 74–98 (2010).

103. Rahman, A., Vahter, M., Ekström, E.-C. & Persson, L.-Å. Arsenic Exposure in Pregnancy Increases the Risk of Lower Respiratory Tract Infection and Diarrhea during Infancy in Bangladesh. *Environmental Health Perspectives* **119**, 719–724 (2011).

104. Raqib, R. *et al.* Effects of in utero arsenic exposure on child immunity and morbidity in rural Bangladesh. *Toxicol. Lett.* **185**, 197–202 (2009).

105. Winterbottom, E. F. *et al.* Prenatal arsenic exposure alters the placental expression of multiple epigenetic regulators in a sex-dependent manner. *Environmental Health* **18**, 18 (2019).

106. Nygaard, U. C. *et al.* Cord blood T cell subpopulations and associations with maternal cadmium and arsenic exposures. *PLOS ONE* **12**, e0179606 (2017).

107. Hoen, A. G. *et al.* Sex-specific associations of infants' gut microbiome with arsenic exposure in a US population. *Scientific Reports* **8**, 12627 (2018).

108. Nadeau, K. C. *et al.* In utero arsenic exposure and fetal immune repertoire in a US pregnancy cohort. *Clin Immunol* **155**, 188–197 (2014).

109. Belderbos, M., Levy, O. & Bont, L. Neonatal innate immunity in allergy development. *Current Opinion in Pediatrics* **21**, 762–769 (2009).

110. Martino, D. J. & Prescott, S. L. Silent mysteries: epigenetic paradigms could hold the key to conquering the epidemic of allergy and immune disease. *Allergy* **65**, 7–15 (2010).

111. Deo, S. S., Mistry, K. J., Kakade, A. M. & Niphadkar, P. V. Role played by Th2 type cytokines in IgE mediated allergy and asthma. *Lung India* **27**, 66–71 (2010).

112. Smith, A. H. *et al.* Chronic respiratory symptoms in children following in utero and early life exposure to arsenic in drinking water in Bangladesh. *Int J Epidemiol* **42**, 1077–1086 (2013).

113. George, C. M. *et al.* Arsenic exposure is associated with pediatric pneumonia in rural Bangladesh: a case control study. *Environ Health* **14**, (2015).

114. Karagas, M. R. *et al.* Rice Intake and Emerging Concerns on Arsenic in Rice: a Review of the Human Evidence and Methodologic Challenges. *Curr Envir Health Rpt* (2019) doi:10.1007/s40572-019-00249-1.

115. European Food Safety Authority. Dietary exposure to inorganic arsenic in the European population. *EFSA Journal* **12**, 3597 (2014).

116. *Commission Regulation (EU) 2015/1006 of 25 June 2015 amending Regulation (EC) No 1881/2006 as regards maximum levels of inorganic arsenic in foodstuffs (Text with EEA relevance)*. *OJ L* vol. 161 (2015).

117. U.S. FDA. FDA proposes limit for inorganic arsenic in infant rice cereal. *FDA* http://www.fda.gov/news-events/press-announcements/fda-proposes-limit-inorganic-arsenic-infant-rice-cereal (2016).

118. *Food Safety: Federal Efforts to Manage the Risk of Arsenic in Rice*. https://www.gao.gov/assets/700/690701.pdf (2018).

119. U.S. FDA. A Quantitative Assessment of Inorganic Arsenic in Apple Juice. (2013).

120. US EPA, O. National Primary Drinking Water Regulations. *US EPA* https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations (2015).

121. NJDEP-News Release 04/130 - DEP Adopts New Mercury, Arsenic Standards. https://www.nj.gov/dep/newsrel/2004/04_0130.htm.

122. New Hampshire HB261 | 2019 | Regular Session. *LegiScan* https://legiscan.com/NH/text/HB261/id/1964222.

123. Gilbert-Diamond, D. *et al.* Rice consumption contributes to arsenic exposure in US women. *Proc Natl Acad Sci U S A* **108**, 20656–20660 (2011).

124. Madan, J. C. *et al.* Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants. *JAMA Pediatr* **170**, 212 (2016).

125. Carignan, C. C. *et al.* Estimated Exposure to Arsenic in Breastfed and Formula-Fed Infants in a United States Cohort. *Environ Health Perspect* **123**, 500–506 (2015).

126. Zou, G. & Donner, A. Extension of the modified Poisson regression model to prospective studies with correlated binary data. *Stat Methods Med Res* **22**, 661–670 (2013).

127. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* **30**, 377–399 (2011).

128. Yorita, K. L., Holman, R. C., Sejvar, J. J., Steiner, C. A. & Schonberger, L. B. Infectious disease hospitalizations among infants in the United States. *Pathologica* **121**, 244–252 (2008).

129. Jackson, K. D., Howie, L. D. & Akinbami, L. J. Trends in allergic conditions among children: United States, 1997-2011. *NCHS Data Brief* 1–8 (2013).

130. National Center for Health Statistics. Allergies and Hay Fever. *Centeror Disease Control and Prevention* https://www.cdc.gov/nchs/fastats/allergies.htm (2019).

131. Wilson, A. C. *et al.* Relation of infant diet to childhood health: seven year follow up of cohort of children in Dundee infant feeding study. *BMJ* **316**, 21–25 (1998).

132. Wright, C. M., Parkinson, K. N. & Drewett, R. F. Why are babies weaned early? Data from a prospective population based cohort study. *Archives of Disease in Childhood* **89**, 813–816 (2004).

133. Poole, J. A. *et al.* Timing of Initial Exposure to Cereal Grains and the Risk of Wheat Allergy. *Pediatrics* **117**, 2175–2182 (2006).

134. Bellamri, N., Morzadec, C., Fardel, O. & Vernhet, L. Arsenic and the immune system. *Current Opinion in Toxicology* **10**, 60–68 (2018).

135. World Health Organization. *Guidelines for Drinking-Water Quality*. (1993).

136. World Health Organization. Preventing disease through healthy environments: exposure to arsenic: a major public health concern. (2019).

137. Signes-Pastor, A. J. *et al.* Levels of infants' urinary arsenic metabolites related to formula feeding and weaning with rice products exceeding the EU inorganic arsenic standard. *PLoS One* **12**, (2017).

138. U.S. FDA. *Arsenic in Rice and Rice Products Risk Assessment Report*. https://www.fda.gov/media/96071/download (2016).

139. Korioth, T. Parent Plus: Limit infants' exposure to arsenic by feeding a variety of grains. *AAP News* (2019).

140. Islam, S., Rahman, M. M., Islam, M. R. & Naidu, R. Arsenic accumulation in rice: Consequences of rice genotypes and management practices to reduce human health risk. *Environment International* **96**, 139–155 (2016).

141. Cubadda, F., Jackson, B. P., Cottingham, K. L., Van Horne, Y. O. & Kurzius-Spencer, M. Human exposure to dietary inorganic arsenic and other arsenic species: State of knowledge, gaps and uncertainties. *Sci Total Environ* **579**, 1228–1239 (2017).

142. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **380**, 2095–2128 (2012).

143. Wang, Y. *et al.* 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *The ISME Journal* **3**, 944–954 (2009).

144. Antosca, K. M. *et al.* Altered Stool Microbiota of Infants with Cystic Fibrosis Shows a Reduction in Genera Associated with Immune Programming from Birth. *Journal of Bacteriology* **201**, (2019).

145. Hoen, A. G. *et al.* Associations between Gut Microbial Colonization in Early Life and Respiratory Outcomes in Cystic Fibrosis. *The Journal of Pediatrics* **167**, 138-147.e3 (2015).

146. Stewart, C. J. *et al.* The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. *Acta Paediatrica* **101**, 1121–1127 (2012).

147. Stewart, C. J. *et al.* Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. *Microbiome* **5**, 75 (2017).

148. Jakobsson, H. E. *et al.* Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by Caesarean section. *Gut* **63**, 559–566 (2014).

149. Bäckhed, F. *et al.* Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host & Microbe* **17**, 690–703 (2015).

150. Schwartz, S. *et al.* A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome Biology* **13**, r32 (2012).

151. Bosch, A. A. T. M. *et al.* Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. *EBioMedicine* **9**, 336–345 (2016).

152. Laubereau, B. *et al.* Caesarean section and gastrointestinal symptoms, atopic dermatitis, and sensitisation during the first year of life. *Archives of Disease in Childhood* **89**, 993–997 (2004).

153. Dieterich, C. M., Felice, J. P., O'Sullivan, E. & Rasmussen, K. M. Breastfeeding and Health Outcomes for the Mother-Infant Dyad. *Pediatr Clin North Am* **60**, 31–48 (2013).

154. Zhang, M. *et al.* Association of prenatal antibiotics with measures of infant adiposity and the gut microbiome. *Ann Clin Microbiol Antimicrob* **18**, 18 (2019).

155. Zou, Z.-H. *et al.* Prenatal and postnatal antibiotic exposure influences the gut microbiota of preterm infants in neonatal intensive care units. *Ann Clin Microbiol Antimicrob* **17**, 9 (2018).

156. Coker, M. *et al.* Specific class of intrapartum antibiotics relates to maturation of the infant gut microbiota: a prospective cohort study. *BJOG* **127**, 217–227 (2020).

157. Hemarajata, P. & Versalovic, J. Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation. *Therap Adv Gastroenterol* **6**, 39–51 (2013).

158. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat Med* **22**, 713–722 (2016).

159. Harding, J. N. *et al.* Altered gut microbiota in infants is associated with respiratory syncytial virus disease severity. *BMC Microbiology* **20**, 140 (2020).

160. Lundgren, S. N. *et al.* Maternal diet during pregnancy is related with the infant stool microbiome in a delivery mode-dependent manner. *Microbiome* **6**, 109 (2018).

161. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581–583 (2016).

162. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**, 7188–7196 (2007).

163. McIver, L. J. *et al.* bioBakery: a meta'omic analysis environment. *Bioinformatics* **34**, 1235–1237 (2018).

164. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* **9**, 811–814 (2012).

165. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).

166. Rosas-Salazar, C. *et al.* Exclusive Breastfeeding, the Early-life Microbiome and Immune Response, and Common Childhood Respiratory Illnesses. *Journal of Allergy and Clinical Immunology* **0**, (2022).

167. Poppleton, D. I. *et al.* Outer Membrane Proteome of Veillonella parvula: A Diderm Firmicute of the Human Microbiome. *Frontiers in Microbiology* **8**, (2017).

168. Reichardt, N. *et al.* Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J* **8**, 1323–1335 (2014).

169. Shimizu, J. *et al.* Propionate-producing bacteria in the intestine may associate with skewed responses of IL10-producing regulatory T cells in patients with relapsing polychondritis. *PLoS One* **13**, (2018).

170. van den Bogert, B., Meijerink, M., Zoetendal, E. G., Wells, J. M. & Kleerebezem, M. Immunomodulatory Properties of Streptococcus and Veillonella Isolates from the Human Small Intestine Microbiota. *PLoS One* **9**, (2014).

171. Yang, K., Kruse, R. L., Lin, W. V. & Musher, D. M. Corynebacteria as a cause of pulmonary infection: a case series and literature review. *Pneumonia* **10**, 10 (2018).

172. Manzella, J. P., Kellogg, J. A. & Parsey, K. S. Corynebacterium pseudodiphtheriticum : A Respiratory Tract Pathogen in Adults. *Clinical Infectious Diseases* **20**, 37–40 (1995).

173. Ahmed, K. *et al.* Corynebacterium pseudodiphtheriticum: A Respiratory Tract Pathogen. *Clinical Infectious Diseases* **20**, 41–46 (1995).

174. Edouard, S. *et al.* The nasopharyngeal microbiota in patients with viral respiratory tract infections is enriched in bacterial pathogens. *Eur J Clin Microbiol Infect Dis* **37**, 1725–1733 (2018).

175. Hayashi, A. *et al.* A single strain of Clostridium butyricum induces intestinal IL-10-producing macrophages to suppress acute experimental colitis in mice. *Cell Host Microbe* **13**, 711–722 (2013).

176. Atarashi, K. *et al.* T reg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).

177. Lozupone, C. *et al.* Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles of human gut symbionts. *Genome Res.* **22**, 1974–1984 (2012).

178. Mégraud, F., Bébéar, C., Dabernat, H. & Delmas, C. Haemophilus species in the human gastrointestinal tract. *Eur. J. Clin. Microbiol. Infect. Dis.* **7**, 437–438 (1988).

179. Zhang, D. *et al.* The Cross-Talk Between Gut Microbiota and Lungs in Common Lung Diseases. *Front Microbiol* **11**, 301 (2020).

180. Yang, I. *et al.* The Infant Microbiome: Implications for Infant Health and Neurocognitive Development. *Nursing Research* **65**, 76–88 (2016).

181. Arrieta, M.-C. *et al.* Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. *J. Allergy Clin. Immunol.* **142**, 424-434.e10 (2018).

182. Arrieta, M.-C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* **7**, 307ra152 (2015).

183. Stiemsma, L. T. *et al.* Shifts in Lachnospira and Clostridium sp. in the 3-month stool microbiome are associated with preschool age asthma. *Clinical Science* **130**, 2199–2207 (2016).

184. Abrahamsson, T. R. *et al.* Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin. Exp. Allergy* **44**, 842–850 (2014).

185. Krishnamoorthy, S., Coetzee, V., Kruger, J., Potgieter, H. & Buys, E. M. Dysbiosis Signatures of Fecal Microbiota in South African Infants with Respiratory, Gastrointestinal, and Other Diseases. *The Journal of Pediatrics* **218**, 106-113.e3 (2020).

186. Liu, H. *et al.* Diagnosing and tracing the pathogens of infantile infectious diarrhea by amplicon sequencing. *Gut Pathogens* **11**, 12 (2019).

187. Pop, M. *et al.* Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biology* **15**, R76 (2014).

188. Monira, S. *et al.* Metagenomic profile of gut microbiota in children during cholera and recovery. *Gut Pathog* **5**, 1–9 (2013).

189. Kaci, G. *et al.* Anti-Inflammatory Properties of Streptococcus salivarius, a Commensal Bacterium of the Oral Cavity and Digestive Tract. *Appl Environ Microbiol* **80**, 928–934 (2014).

190. Achievements in Public Health, 1900-1999 Impact of Vaccines Universally Recommended for Children -- United States, 1990-1998. https://www.cdc.gov/mmwr/preview/mmwrhtml/00056803.htm.

191. Chard, A. N. Routine Vaccination Coverage — Worldwide, 2019. *MMWR Morb Mortal Wkly Rep* **69**, (2020).

192. Hill, H. A. Vaccination Coverage by Age 24 Months Among Children Born in 2016 and 2017 — National Immunization Survey-Child, United States, 2017–2019. *MMWR Morb Mortal Wkly Rep* **69**, (2020).

193. Zimmermann, P. & Curtis, N. Factors That Influence the Immune Response to Vaccination. *Clinical Microbiology Reviews* **32**, e00084-18.

194. Zhang, Y. *et al.* Composition of the murine gut microbiome impacts humoral immunity induced by rabies vaccines. *Clin Transl Med* **10**, e161 (2020).

195. Uchiyama, R., Chassaing, B., Zhang, B. & Gewirtz, A. T. Antibiotic Treatment Suppresses Rotavirus Infection and Enhances Specific Humoral Immunity. *The Journal of Infectious Diseases* **210**, 171–182 (2014).

196. Lamousé-Smith, E. S., Tzeng, A. & Starnbach, M. N. The Intestinal Flora Is Required to Support Antibody Responses to Systemic Immunization in Infant and Germ Free Mice. *PLOS ONE* **6**, e27662 (2011).

197. Oh, J. Z. *et al.* TLR5-Mediated Sensing of Gut Microbiota Is Necessary for Antibody Responses to Seasonal Influenza Vaccination. *Immunity* **41**, 478–492 (2014).

198. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).

199. Jódar, L. *et al.* Serological criteria for evaluation and licensure of new pneumococcal conjugate vaccine formulations for use in infants. *Vaccine* **21**, 3265–3272 (2003).

200. Liang, J. L. Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* **67**, (2018).

201. Birth-18 Years Immunization Schedule | CDC. https://www.cdc.gov/vaccines/schedules/hcp/imz/child-adolescent.html (2021).

202. Olarte, L. & Jackson, M. A. Streptococcus pneumoniae. *Pediatrics In Review* **42**, 349–359 (2021).

203. McCool, T. L., Harding, C. V., Greenspan, N. S. & Schreiber, J. R. B- and T-Cell Immune Responses to Pneumococcal Conjugate Vaccines: Divergence between Carrier- and Polysaccharide-Specific Immunogenicity. *Infection and Immunity* **67**, 4862–4869 (1999).

204. Genova, G. D., Roddick, J., McNicholl, F. & Stevenson, F. K. Vaccination of human subjects expands both specific and bystander memory T cells but antibody production remains vaccine specific. *Blood* **107**, 2806–2813 (2006).

205. Zhao, T. *et al.* Influence of gut microbiota on mucosal IgA antibody response to the polio vaccine. *npj Vaccines* **5**, 1–9 (2020).

206. Robertson, R. C. *et al.* The fecal microbiome and rotavirus vaccine immunogenicity in rural Zimbabwean infants. *Vaccine* **39**, 5391–5400 (2021).

207. Fix, J. *et al.* Association between Gut Microbiome Composition and Rotavirus Vaccine Response among Nicaraguan Infants. *The American Journal of Tropical Medicine and Hygiene* **102**, 213–219 (2019).

208. Mullié, C. *et al.* Increased Poliovirus-Specific Intestinal Antibody Response Coincides with Promotion of Bifidobacterium longum-infantis and Bifidobacterium breve in Infants: A Randomized, Double-Blind, Placebo-Controlled Trial. *Pediatr Res* **56**, 791–795 (2004).

209. Kukkonen, K., Nieminen, T., Poussa, T., Savilahti, E. & Kuitunen, M. Effect of probiotics on vaccine antibody responses in infancy – a randomized placebo-controlled double-blind trial. *Pediatric Allergy and Immunology* **17**, 416–421 (2006).

210. Licciardi, P. *et al.* Maternal Supplementation with LGG Reduces Vaccine-Specific Immune Responses in Infants at High-Risk of Developing Allergic Disease. *Frontiers in Immunology* **4**, 381 (2013).

211. Lucas-Hourani, M. *et al.* Inhibition of Pyrimidine Biosynthesis Pathway Suppresses Viral Growth through Innate Immunity. *PLOS Pathogens* **9**, e1003678 (2013).

212. Dimitrova, P. *et al.* Restriction of De Novo Pyrimidine Biosynthesis Inhibits Th1 Cell Activation and Promotes Th2 Cell Differentiation. *The Journal of Immunology* **169**, 3392–3399 (2002).

213. Hubler, M. J. & Kennedy, A. J. Role of lipids in the metabolism and activation of immune cells. *The Journal of Nutritional Biochemistry* **34**, 1–7 (2016).

214. Milani, C. *et al.* The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and Molecular Biology Reviews* **81**, e00036-17 (2017).

215. Coker, M. O. *et al.* Infant Feeding Alters the Longitudinal Impact of Birth Mode on the Development of the Gut Microbiota in the First Year of Life. *Frontiers in Microbiology* **12**, (2021).

216. VanderWeele, T. J. & Vansteelandt, S. Odds Ratios for Mediation Analysis for a Dichotomous Outcome. *American Journal of Epidemiology* **172**, 1339–1348 (2010).

217. Valeri, L. & Vanderweele, T. J. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods* **18**, 137–150 (2013).

218. VanderWeele, T. J. A unification of mediation and interaction: a four-way decomposition. *Epidemiology* **25**, 749–761 (2014).

219. Tchetgen Tchetgen, E. J. Inverse Odds Ratio-Weighted Estimation for Causal Mediation Analysis. *Stat Med* **32**, 4567–4580 (2013).

220. Nguyen, Q. C., Osypuk, T. L., Schmidt, N. M., Glymour, M. M. & Tchetgen Tchetgen, E. J. Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting. *American Journal of Epidemiology* **181**, 349–356 (2015).

221. Sobel, M. E. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology* **13**, 290–312 (1982).

222. Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G. & Barceló-Vidal, C. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology* **35**, 279–300 (2003).

223. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2020).

224. Ma, S. *et al.* A statistical model for describing and simulating microbial community profiles. *PLOS Computational Biology* **17**, e1008913 (2021).

225. Dierikx, T. H. *et al.* The influence of prenatal and intrapartum antibiotics on intestinal microbiota colonisation in infants: A systematic review. *Journal of Infection* **81**, 190–204 (2020).