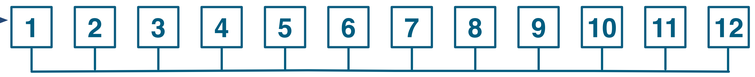
Uncovering microbiome correlates of upper and lower respiratory tract infections in infants using sequential Bayesian analysis

Jyoti Shankar[1], Leilani T. Nillos[2], Stephanie Mounaud[1], Edelwisa Segubre-Mercado[2], Aileen M. R. Mojica[2], Brian Bacay[2], Sherry Vi N. Sarra[2], Xenia D. Geraldino[2], Herbert Guinto[2], Vernoni E. Dulalia[2], Karen E. Nelson[1], Marilla G. Lucero[2] and William C. Nierman[1]. Author affiliations: [1]J. Craig Venter Institute, USA; [2]Research Institute for Tropical Medicine, Philippines

Study design

N = 135 healthy Philippine Infants recruited at birth and followed up every month for a year.



12 clinical evaluations + 12 nasopharyngeal samples

Clinical variables

135 × **~22 variables**

Microbiome proportions

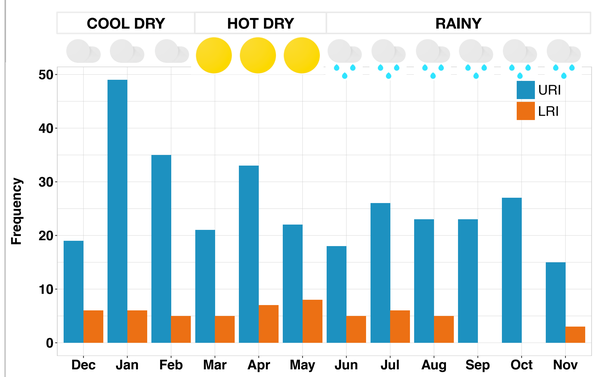
135 × **63 dominant genera** × 12

Bayesian mixed-effects models

$$\log\left(\frac{\text{Prob.}(\text{Diagnosis} = d \in \{\text{URI, LRI}\}\,|\boldsymbol{X})}{\text{Prob.}(\text{Diagnosis} = \text{No resp. symptoms}\,|\boldsymbol{X})}\right) = \alpha_{\text{infant}} + \boldsymbol{\beta}\boldsymbol{X}^{(c,\,m)} + \boldsymbol{\beta}^{(\text{interact.})}\boldsymbol{X}^{(\text{season}\times m)}$$

Why mixed-effects models with interactions?

Each infant contributes from 4 to 12 samples to the analysis and the distribution of URI and LRI is seasonal in nature.



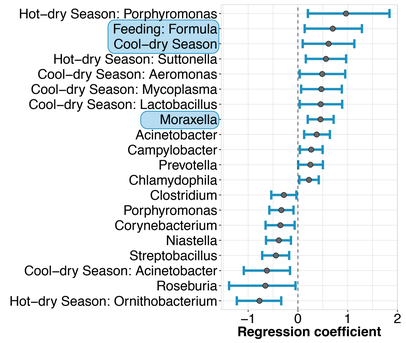
Model specifications:
Multinomial logistic regression with mixed-effects and interaction terms

Sample size: 1474 samples from 135 infants.
Response variable: URI, LRI, None (Categorical)
Covariates: 63 dominant genera, 22 clinical variables

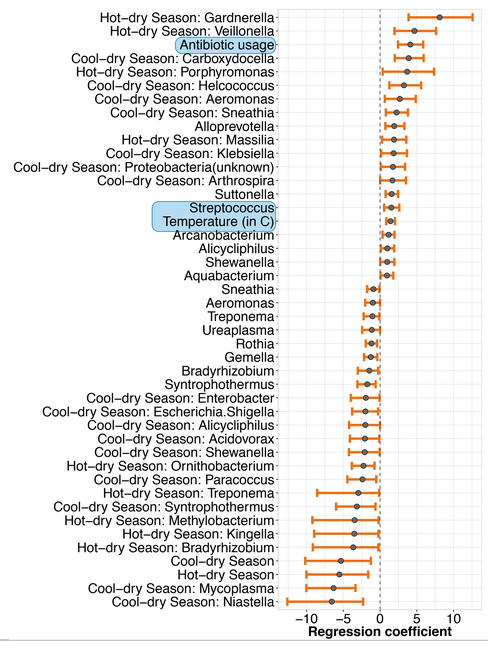
- Mixed-effects model provides a "cross-sectional" view of a time-series dataset while adjusting for repeated measurements from infants.
- Detected known associations (e.g. known clinical correlates of URI and LRI and known pathogen-containing genera) that served as a sanity check.
- Identified several seasonal-interactions of pathogen-containing as well as commensal microbe groups. Could be mediators of observed seasonal variation in URI/LRI and possible future therapeutic targets.
- Provided estimates of URI and LRI probability for each visit. These estimates were used as the response variable in the time-series models.

J. Craig Venter® INSTITUTE

URI associations



LRI associations



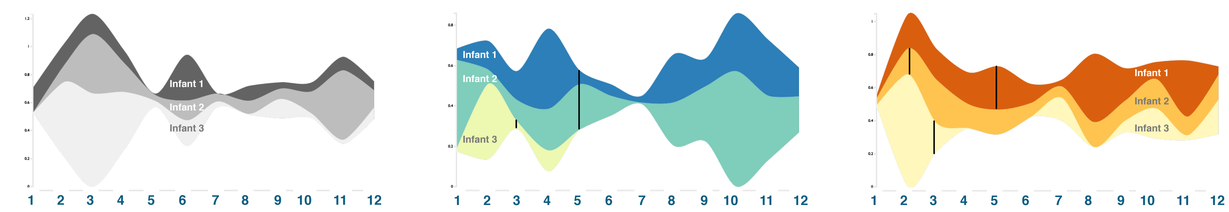
Background and Objectives

- **Upper and lower respiratory tract infections (URI and LRI)** are among **top illnesses** among infants in **first year** of life.
- The **Nasopharynx** is the most common pathway of entry of pathogens into the respiratory tract
- **To what extent** does the microbial community (***the microbiome***) in the nasopharynx contribute to the establishment and proliferation of pathogens and the **occurrence of URI and LRI**?

Bayesian structural time-series models

Why infant-level time-series models?

Infant-level heterogeneity in microbe levels is staggering. Below is a tiny example: **Streptococcus levels** across the **12 visits** in 3 representative infants in each group (Healthy, URI, LRI) are shown. Vertical lines represent URI and LRI diagnoses.

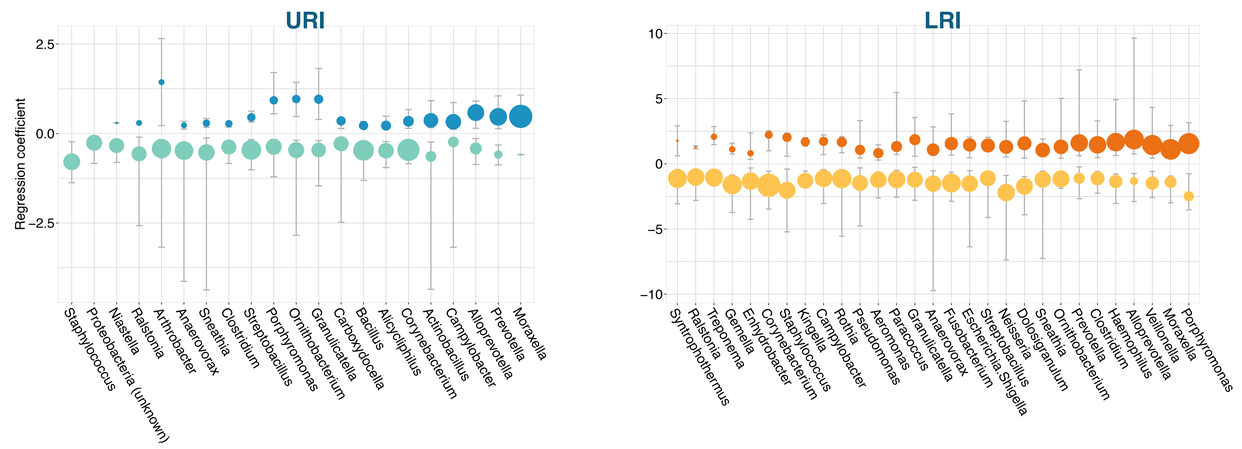


Each structural time-series equation modeled the URI/LRI probability across visits as a function of 63 microbial genera and 13 clinical variables (excluding signs and symptoms of URI/LRI). The URI/LRI probability was estimated by the mixed-effects model

$$Y_t = s_t + \boldsymbol{\beta}\boldsymbol{X}_t^{(\text{clinical, microbiome})} + v_t;\, v_t \sim \mathcal{N}(0, \sigma_v^2) \;\ldots \textit{(Observation equation)}$$

$$s_t = s_{t-1} + w_t;\, w_t \sim \mathcal{N}(0, \sigma_w^2) \;\ldots \textit{(State equation)}$$

Across 131 infant-specific time-series models, microbial genera significantly associated with either increased or decreased probability of URI and LRI, in at least 20% (≥26) infants are summarized below. URI associations appear to be much more consistent across infants than LRI associations. A handful of prominent pathogen-containing genera and several commensal genera modulate risk of URI. LRI associations are evenly distributed across several genera.



Size of each circle corresponds to number of infants with the association. Bars are 95%CI of regression coefficients across these infants.

Future directions

- Improve mixed-model estimates of URI/LRI probability and re-estimate the structural time-series models with these improved estimates.
- Adapt survival modeling framework to examine time-to-URI and time-to-LRI events.
- Investigate determinants of recurrent URI and LRI. (Primary challenge: Very sparse data.)

We most gratefully acknowledge funding from the Bill and Melinda Gates Foundation.

DEPARTMENT OF HEALTH · RESEARCH INSTITUTE FOR TROPICAL MEDICINE