

Modeling time evolution and therapeutic effect in human microbiota.

J. D. Silverman¹, S. Mukherjee², L. A. David³, and V. Pawlowsky-Glahn⁴, J. J. Egozcue⁵

¹Duke University, NC, USA; *justin.silverman@duke.edu*

²Duke University, NC, USA; *sayan@stat.duke.edu*

³Duke University, NC, USA; *lawrence.david@duke.edu*

⁴ Universitat de Girona, Spain; *vera.pawlowsky@udg.edu*

⁵Universidad Politècnica de Catalunya, Barcelona, Spain; *juan.jose.egozcue@upc.edu*

Abstract

Within the biomedical community there is an increasing recognition of the importance that host-associated microbes play in both human health and disease. Moreover, there has been much excitement over the insights that can be obtained from longitudinal measurements of these microbial communities; however, due to statistical limitations appropriate models have been lacking. Host microbiota are typically measured using high-throughput DNA sequencing which results in counts for different species. Relative abundances, assumed compositional, are then estimated from these counts. In addition, due to technological limitations the total number of counts per sample is often small compared to the distribution of species relative abundances leading to datasets with many zero or small counts. With such data, maximum likelihood estimates of sample proportions are biased and models that incorporate the sampling variability are essential. To accommodate time-series modeling of host microbiota, a multinomial-normal-on-the-simplex generalized dynamic linear model has been developed. Using a combination of both real and simulated datasets we demonstrate that this modeling framework enables accurate inference of the effects of prebiotic treatments in microbiota time-series. In addition, the proposed model provides a way of carrying a posteriori multiple imputation, which accounts for the uncertainty caused by the low-count sampling and the effects of the time dependence. The model parameters, estimated in this way can be interpreted as a linear evolution of the community thus enabling novel biological insights.