

Figure 1. Violin plot of intraclass correlation coefficients calculated from Human Microbiome Project (7) samples collected 6 months apart, across 3 body sites.

this data structure and, thus, provide misleading distance measures and spurious correlations that invalidate statistical interpretations (2). Alternative transformations that properly account for the compositional data structure, like centered (4), additive (4), isometric (8), and phylogenetic isometric (9) log-ratios, should be used in conjunction with standard statistical tests in the analysis of microbiome data.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Sinha R, Goedert JJ, Vogtmann E, et al. Quantification of human microbiome stability over 6 months: implications for epidemiologic studies. *Am J Epidemiol*. 2018;187(6):1282–1290.
2. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, et al. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224.
3. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol*. 2016;26(5):330–335.
4. Aitchison J. *The Statistical Analysis of Compositional Data*. London, United Kingdom: Chapman and Hall Ltd; 1986.
5. Aitchison J, Greenacre M. Biplots of compositional data. *J R Stat Soc Ser C Appl Stat*. 2002;51(4):375–392.
6. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207–214.
7. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012;486(7402):215–221.
8. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, et al. Isometric logratio transformations for compositional data analysis. *Math Geol*. 2003;35(3):279–300.
9. Silverman JD, Washburne AD, Mukherjee S, et al. A phylogenetic transform enhances analysis of compositional microbiota data. *ELife*. 2017;6:e21887.

Daniel Ian McSkimming^{1,2}, Hailey R. Banack³, Robert Genco⁴, Jean Wactawski-Wende³, and Michael J. LaMonte³ (e-mail: dim@buffalo.edu)
¹ *Genome, Environment and Microbiome (GEM) Community of Excellence, University at Buffalo, Buffalo, NY*
² *Department of Biochemistry, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY*
³ *Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, NY*
⁴ *Department of Oral Biology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY*

DOI: 10.1093/aje/kwz021; Advance Access publication: February 25, 2019

Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health 2019. This work is written by (a) US Government employee(s) and is in the public domain in the US.

THREE AUTHORS REPLY

We thank McSkimming et al. (1) for their comments on our paper (2), which provide an opportunity for us to expand the discussion of quantification of temporal instability of the human microbiome and the implications for large-scale prospective epidemiology studies.

In our original paper (2), we reported relatively low intraclass correlation coefficients (ICCs) for phylum-level relative abundance (RA) data from the Human Microbiome Project (HMP) (3, 4). McSkimming et al. reported much higher ICC values after centered log-ratio (CLR) transformation (5).

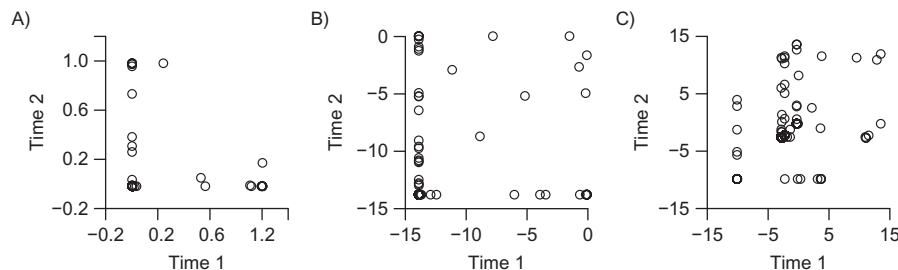


Figure 1. Intraclass correlation coefficient values for an uncommon taxon using naive estimates of relative abundance (RA) vector (A), log transformation (B), and centered log-ratio transformation (C). The RA vectors at 2 data points for 1,000 subjects were generated for 5 taxa based on the data generative model with $\Lambda = (0.1, 1, 1, 1, 5)$ and $\theta = 0.1$ (suggesting large over-time instability). The average RA for the first taxon is $0.1/(0.1 + 1 + 1 + 1 + 5) = 1.2\%$.

Building upon this, we developed a data-generative model to investigate whether CLR improved ICC values only for the HMP data or consistently for other scenarios. To proceed, we assume that the underlying long-term RA composition vector for K taxa is $\Lambda = (\lambda_1, \dots, \lambda_K)$. The time-specific RA vector $P_t = (p_{t1}, \dots, p_{tK})$ follows a Dirichlet distribution $D(\theta\lambda_1, \dots, \theta\lambda_K)$. Here, the parameter θ models the over-time instability, with a large value of θ implying smaller variability and thus a higher ICC. We used this model to simulate RA vectors for 1,000 subjects at 2 time points to evaluate ICC numerically. We found that naive RA estimates without transformation typically had the lowest ICC values and that CLR transformation consistently improved ICC for individual taxa, particularly for rare taxa. In Figure 1, using the naive estimate for uncommon taxa (average RA = 1.2%), the ICC is practically zero; log transformation modestly improves ICC; CLR improves ICC to 0.83. While it is mathematically very complicated, if not impossible, to explicitly derive a formula for ICC after CLR, numerical examination suggested that the high value of ICC after CLR was driven primarily by other, more common taxa that have relatively high ICC values.

The motivation for calculating ICC metrics is to evaluate the temporal-instability power loss to detect associations. More specifically, an ICC value is relevant only to the specific disease model. If the risk of developing a disease depends linearly on the RA of a taxon, the relevant ICC should be evaluated using the naive RA of the taxon. If the disease risk depends on $\log(\text{RA})$, the relevant ICC should be evaluated using the log-transformed RA. The observed high ICC after CLR transformation in the HMP data would suggest a potentially small power loss due to over-time variability. However, effect size change due to the CLR transformation is another important factor that affects statistical power. CLR-based ICC might be overly optimistic for power estimates if effect-size change is ignored. Thus, careful investigation of transformation is warranted under different disease models by jointly considering ICC and effect sizes. Given that very few prospective studies have been performed to suggest an unambiguous disease model, we would recommend evaluating ICCs under different transformations that do or do not consider the compositional feature of the microbiome (i.e., log

transformation, CLR (5), and isometric transformation (6), as suggested by McSkimming et al. (1)). In conclusion, transformation of microbiome data should be compatible not only with compositional structure of the data but also with the postulated disease model.

ACKNOWLEDGMENTS

This work was funded by the Intramural Research Program of the National Cancer Institute.

Conflict of interest: none declared.

REFERENCES

1. McSkimming DI, Banack HR, Genco R, et al. Re: "Quantification of human microbiome stability over 6 months: implications for epidemiologic studies". *Am J Epidemiol.* 2019; 188(4):808–809.
2. Sinha R, Goedert JJ, Vogtmann E, et al. Quantification of human microbiome stability over 6 months: implications for epidemiologic studies. *Am J Epidemiol.* 2018;187(6): 1282–1290.
3. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature.* 2012;486(7402):215–221.
4. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* 2012; 486(7402):207–214.
5. Aitchison J. *The Statistical Analysis of Compositional Data.* London, United Kingdom: Chapman and Hall Ltd; 1986.
6. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, et al. Isometric logratio transformations for compositional data analysis. *Math Geol.* 2003;35(3):279–300.

Jianxin Shi, Rashmi Sinha, and James J. Goedert
(e-mail: sinhar@nih.gov)

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD

DOI: 10.1093/aje/kwz022; Advance Access publication: February 25, 2019