See corresponding article on page 432.

Can data-driven approaches for dietary pattern assessment improve microbiome epidemiology research?

Abigail J Johnson[1] *and Noel T Mueller*[2]

[1]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA; and [2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

A growing body of research implicates the gut microbiome—the collection of microbes and their genetic elements that inhabit the human gastrointestinal tract—as a potential mediator or modifier of diet's effects on a variety of health-related outcomes, including variable blood glucose response to foods (1), lipid profiles (2), and blood pressure (3). Gut microbes survive by consuming indigestible dietary constituents, like fiber, and other byproducts of digestion. They also consume constitutively available nutrients such as mucin, lactate, and glucose. Ultimately, microbial production of metabolites from diet and nondiet precursors can impact host physiology (4). Because diet is the main source of fuel for gut microbes, it is appealing to conclude that dietary composition should account for a majority of microbiome variation. However, the amount of interindividual microbiome variation attributed to dietary features has thus far been relatively small (5), potentially owing to suboptimal diet measurement.

In this issue of the Journal, Cotillard and colleagues (6) explored dietary pattern–driven variation in the gut microbiome using data from the American Gut Project (AGP). Dietary pattern assessment typically uses 1 of 2 approaches: a priori (predefined) or a posteriori (data-driven). To our knowledge, no prior studies had compared how a priori and a posteriori approaches associate with gut microbiome features or examined the extent to which each approach explains variation in overall microbiome composition.

A priori dietary indices measure adherence to a dietary pattern that is defined by prior evidence or expert opinion. The benefit of this approach is that dietary pattern associations can be more easily compared across populations because they are not derived from study population-specific dietary data. However, because predefined indices often depend on expert recommendations, culturally specific foods, or outdated research, the dietary patterns may fit specific populations less well. In their study, Cotillard and colleagues (6) chose first to explore relationships between microbiome features and adherence to the 2010 Dietary Guidelines for Americans in 744 individuals using the predefined Healthy Eating Index (HEI-2010). They found a minimal but statistically significant association between HEI-2010 scores and gut microbiome beta diversity (the measure of overall microbiome similarity between two samples) after controlling for the influence of confounding variables (age, sex, and BMI). This finding agrees with other research that has shown that the HEI-2010 is associated, albeit modestly, with gut microbiome beta-diversity variance (7, 8). The weak association of HEI-2010 with overall microbiome composition is not entirely surprising given how the HEI-2010 index quantifies and incorporates dietary components (e.g., total vegetables, total fruit, whole grains, etc.), which may inadequately account for the heterogeneous microbiome effects of the individual foods within these food groups.

Cotillard and colleagues (6) then examined whether an a posteriori, data-driven dietary pattern would afford greater explanatory power for microbiome variation. Data-driven strategies use dimensionality reduction techniques to define dietary patterns for a specific study population (9). Because these approaches are study-population specific, the results may not be generalizable to other populations. The strength of data-driven dietary patterns is that they should best reflect the dietary patterns of the study population. However, these dietary patterns may also be overfit to the study population, and are unlikely to be replicated in cultures with distinct culinary practices. In the microbiome space, data-driven dietary patterns built from food-frequency questionnaire (FFQ) data (10) and daily 24-hour recall data (11) have been shown to covary significantly with gut microbiome beta diversity. In their study, Cotillard et al. used energy-adjusted food groups to develop dietary clusters for 620 individuals (a subset of the 744 used for a priori analysis). The authors described the dietary patterns of the resulting clusters with 2 prudent-like patterns, "Plant-based" and "Flexitarian"; 2 Western-like patterns, "Health-Conscious Western" and "Standard Western"; and 1 "Exclusion" pattern characterized by low intake of carbohydrates. Unlike the HEI-2010 or other diet factors examined, the data-driven dietary patterns were significantly associated with

AJJ is supported by the National Institutes of Health (P30CA077598 and R21DK125933). NTM is supported by the National Institutes of Health (K01HL141589). The funding sources had no involvement in the collection, analysis, or interpretation of data; the writing of the report; or the decision to submit the article for publication.

Address correspondence to AJJ (e-mail: abbyj@umn.edu).

First published online January 5, 2022; doi: https://doi.org/10.1093/ajcn/nqab386.

microbiome alpha diversity (microbial diversity within a sample). For example, the "Standard Western" pattern was associated with lower gut microbiome alpha diversity as compared with both the "Flexitarian" and "Plant-based" patterns, which is consistent with prior work showing that greater fruit and vegetable intake is a driver of microbial diversity (12). The data-driven dietary patterns were also significantly associated with beta diversity, however, they only explained a marginally greater percent of variance than the HEI-2010 (1.3% vs. 1%, respectively). Unfortunately, the predefined and data-driven approaches were not applied to the exact same subset of individuals, making it difficult to directly compare the associations and variance explained by the different methods.

When interpreting the results of this study, readers must consider the following caveats. First, there was no external replication of the data-driven approaches. Data-driven dietary patterns are, by nature, best fit to the study population in which they were derived. However, they perform less well when tested in external populations with different distributions of diet data. In contrast, predefined dietary patterns, such as the HEI-2010, may reflect a given population's dietary patterns less well because they are externally derived. Second, the AGP cohort is a unique study population that skews toward older individuals, with a higher proportion of women, who have higher income, higher-quality diets, lower BMI, and a higher incidence of gastrointestinal complaints as compared to a representative sample of US adults. This limits the external validity of these results. Third, the cross-sectional nature of this study limits the ability to discern temporality, let alone causality of the observed associations. Moreover, this study design is subject to potential residual and unmeasured confounding. For example, we cannot conclude whether the association of the "Exclusion" dietary pattern with low *Bifidobacteria* levels is due to the restricted dietary pattern itself or to the reason that individuals chose to adhere to a restricted diet, namely gastrointestinal disorders, which were more common in the restricted dietary pattern group. Finally, information bias is also a concern in this study, which used self-reported dietary data collected using VioScreen (Viocare, Inc.), a visual FFQ. Underlying health status can influence the accuracy of self-reported diet; for example, adults with higher body weight may systematically underreport their dietary intake.

Putting these issues aside, this research begs the question, do these and similar findings have implications for how we explore causal associations with the gut microbiome in epidemiologic studies? Specifically, we ask whether data-driven approaches are helpful when aiming to control for diet in observational studies that seek to connect the microbiome with health outcomes? While data-driven dietary patterns could help to control for dietary variation in microbiome studies, in the current study the differences in variation explained in gut microbiome beta diversity by the different dietary patterns were modest (approximately 1.3% for the data-driven patterns and less than 1% for the predefined patterns) and unlikely to materially improve model fit or reduce residual confounding by diet. We may never be able to control for all external factors that contribute to microbiome variation. Still, if, in the future, data-driven approaches can help to capture these difficult-to-measure factors, particularly those correlated with diet, they would be a most-welcome addition to the microbiome-epidemiology tool belt. It is important to note that in addition to helping to control for confounding by diet, data-driven dietary patterns can also facilitate the discovery of novel microbiome-centric dietary patterns to test in randomized trials. However, the first step would be identify unique dietary patterns that strongly associate with both microbiome features and health outcomes, and then replicate these data-driven dietary patterns in external cohorts.

In summary, Cotillard and colleagues provide evidence that data-driven dietary patterns are more strongly associated with several gut microbiome outcomes, including alpha-diversity, than HEI-2010, but they only explain slightly more variance in overall microbiome composition (i.e. beta diversity). This research has implications for observational studies hoping to minimize confounding by dietary patterns on microbiome–outcome associations. However, the amount of variation explained by dietary patterns built from FFQ-collected data, regardless of the method used, remains relatively small, potentially suggesting higher resolution dietary data is needed for data driven approaches (11). Also, while the authors do not directly compare the microbiome variance explained by dietary patterns with that of other variables like age, sex, and BMI, we might posit that those factors correlate with diet and remain important to control for in microbiome research. More broadly, while identifying dietary pattern–microbiome associations in observational studies may eventually lead to translatable microbiota interventions that can be leveraged to improve health, conclusive results in this research area will ultimately require adequately powered randomized trials.

The authors' responsibilities were as follows—AJJ: wrote the first draft of the paper; NTM: provided additional content and edits; and both authors: hold equal responsibility for the content and read and approved the final manuscript. The authors report no conflicts of interest.

References

1. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015;163(5):1079–94.
2. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, Leeming E, Gibson R, Le Roy C, Khatib HA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med 2021;27(2):321–32.
3. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, Wu S, Liu W, Cui Q, Geng B, et al. Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome 2017;5(1):14.
4. Laursen MF, Sakanaka M, von Burg N, Mörbe U, Andersen D, Moll JM, Pekmez CT, Rivollier A, Michaelsen KF, Mølgaard C, et al. Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in the infant gut. Nature Microbiology 2021;6(11):1367–82.
5. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature 2018;555(7695):210–5.
6. Cotillard A, Cartier-Meheust A, Litwin NS, Chaumont S, Saccareau M, Lejzerowicz F, Tap J, Koutnikova H, Lopez DG, McDonald D, et al. A posteriori dietary patterns better explain variations of the gut microbiome than individual markers in the American Gut Project. Am J Clin Nutr 2022;115(2):432–43.
7. Bowyer RCE, Jackson MA, Pallister T, Skinner J, Spector TD, Welch AA, Steves CJ. Use of dietary indices to control for diet in human gut microbiota studies. Microbiome 2018;6(1):77.
8. Maskarinec G, Hullar MAJ, Monroe KR, Shepherd JA, Hunt J, Randolph TW, Wilkens LR, Boushey CJ, Le Marchand L, Lim U, et al. Fecal microbial diversity and structure are associated with diet

quality in the Multiethnic Cohort Adiposity Phenotype Study. J Nutr 2019;149(9):1575–84.
9. Zhao J, Li Z, Gao Q, Zhao H, Chen S, Huang L, Wang W, Wang T. A review of statistical methods for dietary pattern analysis. Nutr J 2021;20(1):37.
10. Claesson MJ, Jeffery IB, Conde S, Power SE, O'Connor EM, Cusack S, Harris HMB, Coakley M, Lakshminarayanan B, O'Sullivan O, et al. Gut microbiota composition correlates with diet and health in the elderly. Nature 2012;488(7410):178–84.
11. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Kim AD, Shmagel AK, Syed AN, Walter J, et al. Daily sampling reveals personalized diet-microbiome associations in humans. Cell Host Microbe 2019;25(6):789–802, e5.
12. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, et al. American Gut: an open platform for Citizen Science Microbiome Research mSystems 2018;3:e00031–18.