Published in final edited form as:

Eur J Clin Invest. 2013 July ; 43(7): . doi:10.1111/eci.12100.

The microbiota and human health: Beyond exploration

James J. Goedert, M.D.

Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics,
National Cancer Institute, 6120 Executive Boulevard, Room 7068, Rockville, MD 20852, Ph:
301-435-4724

James J. Goedert: goedertj@mail.nih.gov

Exploratory studies suggest that the human microbiota contributes to clinical conditions that range from the obvious (antibiotic-induced diarrhea, bacterial vaginosis, periodontitis), to the plausible (obesity, inflammatory bowel disease) and the surprising (autism, breast cancer). To accelerate translation to human health in the next phase of microbiota research, challenges and potential confounders need to be carefully considered (Table 1). Avoiding biases and false leads that deter progress requires a clearly stated hypothesis that is rigorously tested in a suitable study design. Case-control and other cross-sectional studies may identify differences that lead to improved diagnostic methods. However, cross-sectional studies provide little insight on whether alterations or differences in the microbiota contribute to the development of or recovery from a disease, because the disease may have altered the microbiota (which is termed reverse causality). For causality, differences or changes in the microbiota must precede the onset or change in the condition. This generally mandates a prospective design, ideally with repeated specimen sampling, to obtain insight on the time during which a microbial difference predicts disease or recovery.

Homogeneous, well validated cases are important. For example, alteration of the gut microbiota with Crohn's disease of the terminal ileum has been noted, whereas alterations with colonic Crohn's disease and ulcerative colitis have been less apparent.¹ Kwashiorkor was associated with an altered fecal microbiota in infants who were fed a protein-deficient diet.² Marasmus and less severe malnutrition have not been evaluated.

Appropriate control subjects are equally important. Ideal controls are not pristine. Rather, they are identical to the cases, except for the condition of interest. As in a randomized clinical trial, that ideal can be approximated in a cohort in which the participants are free of the condition at baseline and from whom specimens are collected at intervals during follow-up. Even with a cohort, specimens selected for microbiota testing should be restricted to those that address the hypothesis, as it relates to improving diagnostic methods, etiology, pathogenesis, or something else.

Statistical power to address the primary hypothesis should determine the size of a study. Large studies have more power, but rigorously conducted, small studies are occasionally sufficient.³ Hypotheses based on metagenomic pathways or the abundance of rare taxa, which are difficult to quantify,^{4,5} will have lower statistical power. If the parameter is postulated *a priori*, multiple comparisons with other parameters can be considered as purely exploratory, *a posteriori* analyses.

Known errors related to specimen collection, handling and analysis can be controlled by adherence to publicly available protocols (http://www.hmpdacc.org/tools_protocols/tools_protocols.php). Briefly, adequate specimen material must be collected, which is difficult for sites with a relatively low density of microbes such as the bronchus, bladder and skin. If local changes in the microbiota are postulated to contribute to disease, then sampling

of sub-sites, such as subgingival plaque or colonic mucosa is required. Broader microbial profiling in surrogates such as saliva or feces may be appropriate for other hypotheses.^{1,4} Contamination by adjacent body sites, fluids, and the clinic environment must be avoided. Collected specimens must be rapidly frozen or chemically stabilized. Otherwise, large population shifts arise from differential *ex vivo* growth, for example by aerobes versus anaerobes. Inhibitors of polymerase chain reaction (PCR) or DNA sequencing, as may be encountered in bile and formalin-fixed paraffin-embedded tissues, as well as day-to-day and batch-to-batch variation in assays, must be assessed by inclusion of positive controls. RNA expression and protein functions are rather labile, whereas DNA sequences are highly reproducible when specimens have been stored at or below -20°C . Repeated thaws, however, can cause sequence artifacts including loss of minor taxa. Further losses can occur with washing to remove collection media, incomplete lysis of the microbial cells, and inefficient DNA extraction. Profiling of the microbiota generally targets 16S rRNA genes, because they are present in all bacteria, have highly conserved regions for PCR amplification, and have highly variable regions to estimate diversity. However, because not all 16S rRNA genes are amplified with equal efficiency, the PCR primers must be carefully selected to match the taxa that are anticipated. DNA bar coding each specimen's amplicons allows pooling of dozens to hundreds of specimens for multiplexed sequencing. This increases efficiency and reduces cost, but such profiling may not detect minor taxa including pathogenic or antibiotic resistant variants.^{4,5} Careful inspection and editing are needed to remove low-quality sequences and amplification artifacts termed chimeras. Sequences deemed to be valid are clustered as operational taxonomic units (OTUs), typically with 97% identity that approximates species. Phylogenetic classification of OTUs relies on large, but still incomplete, public databases. Rare taxa may be retained to test hypotheses of differences in diversity, but rare taxa should be excluded when accurate taxonomy is essential.³ Standards to estimate absolute levels of human microbes have not been established.

Antibiotic use is the major confounder of microbiota research. In practice, this risk is especially high for clinical conditions in which the case patients are likely to have received various types, schedules and doses of antibiotics. Unless antibiotic exposure is equivalent in the case and control groups, an observed association between the microbiota and a clinical condition may be impossible to interpret. Data from specimens collected within one month of an antibiotic exposure provide little insight on the disease. Although less profound than antibiotics, other medications have been associated with different profiles, and potentially important functions, of the microbiota.^{1,5}

Potential confounding by other exogenous factors, such as bathing, dental cleaning, diet and weight change, should be considered. Day-to-day variation in the gut microbiota is surprisingly small, and this can be reduced further by repeated sampling.^{1,4} Probiotics and prebiotics have thus far had negligible effects on the human fecal microbiota. In contrast, substantial change in weight, as well as vegetarian, gluten-free and other major dietary restrictions, must be considered in the selection of participants for gut microbiota studies.^{1,2} Age is important, whereas sex apparently is not.^{1,4,5} Ethnic differences have been seen but may reflect diet, personal hygiene, or other confounders.⁴ There is little indication that human genetics confounds or interacts with the microbiota. For example, monozygotic and dizygotic twin pairs have equivalent similarities in their fecal microbiotas.¹⁻³ Obesity and Kwashiorkor, despite strong links to differences in the fecal microbiota, are unrelated to zygosity.^{2,3}

Increasingly rapid and cost-effective DNA sequencing methods, corresponding information management and analytic pipeline systems, and substantial funding have led to a quantum leap in understanding of the microbiotas that inhabit the human body.^{1,4} Human health

probably is impacted by differences or alterations in the microbiota, but few, if any, studies have unambiguously demonstrated causality. Animal models can deepen understanding and help to translate microbiota research to human health, especially when the experiments are closely tied to and derived from observational research in humans.^{1,2} However, to accelerate understanding of when, where and especially how the microbiota affects human health, there is need for more randomized clinical trials and other rigorous study designs, particularly cohort studies. Actionable translation to human health will not follow unless and until hypotheses are formally tested with careful planning and rigorous execution.

References

1. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012; 489:220–30. [PubMed: 22972295]
2. Smith MI, Yatsunenko T, Manary MJ, et al. Gut microbiomes of Malawian twin pairs discordant for Kwashiorkor. *Science*. 2013
3. Cozen W, Yu G, Gail MH, et al. Fecal microbiota diversity in survivors of adolescent/young adult Hodgkin lymphoma: A study of twins. *Brit J Cancer*. 2013
4. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486:207–14. [PubMed: 22699609]
5. Flores R, Shi J, Gail MH, Gajer P, Ravel J, Goedert JJ. Assessment of the human faecal microbiota: II. Reproducibility and associations of 16S rRNA pyrosequences. *European journal of clinical investigation*. 2012; 42:855–63. [PubMed: 22385292]

Table 1

Challenges and potential confounders in translational microbiota research.

<u>Study design</u>
Hypothesis
Homogeneous, validated cases
Appropriate controls
Statistical power
<u>Specimen collection</u>
Body site, sub-site or surrogate
Quantity
Contamination
Amplification inhibitors
Stabilization
<u>Specimen handling and analysis</u>
Storage
Processing, lysis, extraction
Primer selection
Sequencing depth
Data inspection, editing
Inclusion, assignment of taxa
<u>Confounders</u>
Antibiotics
Other medications
Personal hygiene
Diet, probiotics
Weight change