Annals of Epidemiology 26 (2016) 336–341

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

The Microbiome and Epidemiology

Challenges for case-control studies with microbiome data

J. Paul Brooks PhD *

Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond
Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond
Department of Supply Chain Management and Analytics, Virginia Commonwealth University, Richmond

ARTICLE INFO

Article history:
Received 19 November 2015
Accepted 25 March 2016
Available online 7 April 2016

Keywords:
Microbiome
Control experiments
Rare taxa
Bias
Normalization

ABSTRACT

Purpose: In case-control studies of the human microbiome, the goal is to evaluate whether cases differ from controls in the microbiome composition of a particular body habitat and which taxa are responsible for the differences. These studies leverage sequencing technology and spectroscopy that provide new measurements of the microbiome.
Methods: Three challenges in conducting reproducible microbiome research using a case-control design are compensating for differences in observed and actual microbial community composition, detecting "rare" taxa in microbial communities, and choosing properly powered analysis methods. The significance of each challenge, evaluation of commonly held views, analysis of unanswered questions, and suggestions of strategies for solutions are discussed.
Results: Understanding the effects of these choices on case-control analyses has been underappreciated, with an implicit assumption that further advances in technology will address all the current shortcomings.
Conclusions: It is recommended that research on the human microbiome include positive and negative control experiments to provide insight into bias, contamination, and technical variation. Research protocols such as these may afford a better opportunity to make quantitative and qualitative adjustments to data, thereby reducing the risk of falsely positive results, increasing power to discover true disease determinants, and enhancing interpretation across studies.

Introduction

Advances in sequencing technology and spectroscopy allow deeper insight into the composition and function of the human microbiome. New findings may have epidemiologic and clinical implications. To translate these results to public health policy and clinical practice, a precise understanding of our biological measurements that ensures reproducibility is necessary. Failure to account for differences in protocols can confound results from different experiments and prevent generalizable discoveries [1].

In case-control experiments involving microbiome measurements, the following questions are of interest to epidemiologists in generating further hypotheses about disease mechanisms and treatment:

1. Is there a difference in the α-diversity (within-sample diversity) of bacteria observed in samples from cases versus those of controls?
2. Do the observed quantities of bacteria in samples from cases arise from a different multivariate probability distribution than those from controls? If so, which bacteria are found in different quantities in cases versus controls?

Choices in sample processing protocols and bioinformatics protocols for measuring the microbiome can affect the ability to find answers to these questions. In addition, differences across laboratories in these choices can lead to contradictory conclusions.

Laboratories make different choices in microbiome measurement protocols based on their particular environment of interest. For example, in our measurements of the vaginal microbiome using 16S rRNA sequencing, we designed polymerase chain reaction (PCR) primers that were sure to amplify Chlamydia and other key taxa [2,3]. Doing so makes comparing our results to other studies using different primers difficult, but the ability to detect particular taxa of interest can be a higher priority. Contrary to claims from many

* Corresponding author. Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, P.O. Box 843083, Richmond, VA. Tel.: +1 804-828-5853; fax: +1 804-828-8785.
E-mail address: jpbrooks@vcu.edu.

http://dx.doi.org/10.1016/j.annepidem.2016.03.009
1047-2797/

microbiome researchers [1], not all laboratories should use the same protocols for biological measurements, but should aim for the best characterization of the environment and disease(s) of interest. Understanding the effects of different choices in protocols, and publishing details of protocols [4], then becomes of paramount importance so that results from multiple studies can be reproduced and assimilated.

Because host-microbiome interactions are dynamic and complex, clear signals are often difficult to detect in case-control studies involving microbiome measurements. Most studies to date focus on only a single 'omic type at a single time point, thereby providing an incomplete picture of the environment. Therefore, a careful treatment of data is important for reproducible results.

I discuss three challenges in conducting reproducible microbiome research for case-control studies. For each, we present the significance of the challenge, evaluate commonly held views, pose unanswered questions, and suggest strategies for overcoming them. The challenges are as follows:

1. Differences in observed and actual microbial community composition.
2. Detecting "rare" taxa in microbial communities.
3. Choosing properly powered analysis methods.

Normalization methods have been proposed that address one or more of these challenges based on certain assumptions regarding sample processing. Normalization refers to modifications of processed data, and its common use of the term sometimes confounds the three issues. In this commentary, I tease out these challenges, call into question our knowledge regarding typical assumptions for normalization methods, and propose experiments that will provide answers.

I focus here on 16S rRNA surveys, but the principles apply to other omics technologies. The aim is not to disparage 16S technology as unusable. To the contrary, microbiome researchers know the most about 16S measurements, and the challenges that I discuss here are likely surmountable. Many of the other omics technologies such as whole metagenome shotgun sequencing (WMGSS), cytokine assays, and untargeted metabolomics suffer from many of the same challenges as 16S rRNA surveys, and yet have additional layers of difficulties that must be addressed.

A common myth in the microbiome research community is that problems with microbiome measurements and differences in choices among protocols can only be solved by further advances in technology. The goal of this commentary is to dispel this myth and demonstrate that many of the challenges in reproducibility of microbiome research can be addressed with current technology.

Differences in observed and actual microbial community compositions

Sources of bias

Bias is a difference in observed and actual microbial community compositions. The presence of bias in 16S [5–8], WMGSS [1,9], and whole genome amplification [10] studies is no secret. Bias refers to differences in the observed and actual quantities in a measurement. In 16S rRNA surveys, samples are collected from an environment, filtered from nonorganic material (if needed), subjected to DNA extraction, PCR amplification, sequencing, and taxonomic classification. The end result is a table of counts of DNA fragments (reads) assigned to each taxon or operational taxonomic unit. Each step in this process can introduce bias and alter the signal so that the observed community composition is different from what is actually there and affect conclusions drawn from case-control studies.

Evidence for bias

As an example, consider the samples in Figure 1. In five replicates, the observed proportions of each bacterium exhibit low variation, but the difference with the true mixing concentration is substantial. Although *Fusobacterium* species comprised only 20% of the mixture, the observed proportions were all between 50% and 70%. The observed proportions of *Fusobacterium*, *Prevotella*, *Lactobacillus*, and *Staphylococcus* were larger than the actual proportions. The observed proportions of *Gardnerella* and *Enterococcus* were consistently smaller than the actual proportions. The composition of the positive control mock community is described in the Online supplement.

Similar issues arise in the processing of samples for WMGSS, metabolomics profiling, and cytokine assays. Compared to these measurement technologies, addressing this issue for 16S rRNA surveys is less complicated because the available reference databases are more complete, sample degradation is typically not an issue, constructing mock communities with known compositions is straightforward, and technical variation is lower.

Current practices and open questions regarding bias

Some confuse variation with bias, and mistakenly assume that the small differences between replicates means that there are no problems with bias. In addition, control samples of mock communities with a precise quantitation of the bacteria are rare. Quality control is sometimes performed with only environmental replicates. Control experiments are not standardized and typically are not published.

The current *modus operandi* is to either use only the presence and/or absence of bacteria or to simply ignore the issue of bias. A common view is that though the observed abundances in 16S rRNA surveys are distorted, rank abundances are preserved. In other words, if more reads are observed from bacterium A than bacterium B, then there is more bacterium A than B in the original sample. In reality, whether rank abundance is preserved or not remains an open question.

Additional open questions regarding bias include

1. How much bias is due to each processing step?
2. Is bias an independent effect for each organism?
3. Can bias be modeled to estimate the true community composition?

We recently published results of an experiment that provides preliminary information about these three questions [8] but does not address the rank abundance issue. The experiment involved processing mixtures consisting of one to seven vaginally relevant bacteria. By studying mock communities with known compositions, we were able to accurately quantify overall bias. The median total bias for *Lactobacillus iners* was 38.3% across the samples in which it was included, meaning that more *L. iners* was typically observed than was actually present. The median bias for *Gardnerella vaginalis* was −30.6%, meaning that less was typically observed than was actually present.

By conducting three experiments mixing equal amounts of cells, DNA, and PCR product, and comparing the results, we quantified the contribution to bias of each processing step. We found that DNA extraction and PCR amplification contributed the most to bias, but sequencing and classification contributed very little, indicating that advances in sequencing technology will not address the issue. The median bias due to DNA extraction ranged from −21.6% for *Atopobium vaginae* to 17.6% for *Lactobacillus crispatus*. The median bias due to PCR amplification ranged from −20.8% for *L. crispatus* to