

RESEARCH ARTICLE

An empirically derived method for measuring human gut microbiome alpha diversity: Demonstrated utility in predicting health-related outcomes among a human clinical sample

Sarah L. Hagerty^{1,2*}, Kent E. Hutchison¹, Christopher A. Lowry^{3,4,5}, Angela D. Bryan¹

1 Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, United States of America, **2** Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, United States of America, **3** Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, United States of America, **4** Center for Neuroscience, University of Colorado Boulder, Boulder, CO, United States of America, **5** Center for Microbial Exploration, University of Colorado Boulder, Boulder, CO, United States of America

* sarah.hagerty@colorado.edu



OPEN ACCESS

Citation: Hagerty SL, Hutchison KE, Lowry CA, Bryan AD (2020) An empirically derived method for measuring human gut microbiome alpha diversity: Demonstrated utility in predicting health-related outcomes among a human clinical sample. PLoS ONE 15(3): e0229204. <https://doi.org/10.1371/journal.pone.0229204>

Editor: Pratibha V. Nerurkar, University of Hawai'i at Manoa College of Tropical Agriculture and Human Resources, UNITED STATES

Received: August 22, 2019

Accepted: January 31, 2020

Published: March 2, 2020

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files

Funding: This work was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1144083 awarded to Sarah Hagerty: <https://www.nsfgrfp.org/> National Institute on Alcohol Abuse and Alcoholism under Grant No. R01 AA024632/AA/NIAAA

Abstract

The human gut microbiome has emerged as a potential key factor involved in the manifestation of physical and mental health. Despite an explosion of cross-disciplinary interest in researching the gut microbiome, there remains to be a gold-standard method for operationalizing gut microbiome alpha diversity. Given researchers' interest in examining the relationships among gut microbiome alpha diversity and health-related outcomes of interest, a way of operationalizing the microbiome that yields a numeric value, which could be used in common statistical approaches, is needed. Thus, the current study aims to provide methodological guidance for how to operationalize microbiome alpha diversity. Findings suggest that alpha diversity of the human gut microbiome is comprised of two sub-constructs (richness and evenness), and we propose a step-by-step method of creating alpha diversity composite measures based on this key insight. Finally, we demonstrate that our empirically derived richness and evenness composite measures are significantly associated with health-related variables of interest (alcohol use, symptoms of depression) among a human clinical sample.

Introduction

Over the past two decades, the gut microbiome has received extensive attention from research communities due to its apparent involvement in a host of medical and psychiatric disorders [1]. Some researchers have begun to think of it akin to another body organ, which interacts with the host to promote health and, in some instances, initiate disease. As a point of nomenclature, the microbial species that inhabit the gastrointestinal tract are known as *gut microbiota*, while the genomes of the microbiota are known as the *gut microbiome*. A common, measurable outcome used in microbiome research is the diversity of microbes within a given body habitat, which has been defined as the number, abundance, and distribution of organism

awarded to Kent Hutchison. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

types [2]. Relatedly, the microbiome can characterize the abundance and diversity of species of gut microbiota [3]. Methodologically, human feces are representative of these interpersonal differences [4], making fecal samples a viable sample to collect for such investigations.

Current objectives of microbiome research include understanding the mechanistic role of microbial diversity in disease etiology and human health. Towards these ends, researchers have outlined factors that drive variation in microbiome environments. Among these factors are age, such that microbiome environments are volatile in human infancy and become more stable after the first three years of the lifespan [5]. Furthermore, genetics, environment, and lifestyle factors (e.g., diet, alcohol consumption) appear to influence the human gut microbiome [5,6]. For example, alcohol consumption is one such lifestyle factor that seems to be associated with composition of the human gut microbiome and gut health more broadly. Specifically, alcohol consumption is associated with greater intestinal permeability [7]. In the context of alcohol use disorder, intestinal permeability has been linked to increased levels of proinflammatory signaling molecules [8] and gut microbiome dysbiosis (an imbalance of microbial species within the gut) [7]. Though the causal relationship between gut microbiome composition and inflammation are still being characterized, these initial studies support a theoretical model in which gut microbiome composition underlies the manifestation of disease states via interactions with other biological systems, such as the immune system. Alcohol use disorder is just one of many areas where the microbiome is emerging as an important biological mechanism that could help explain the causes and consequences of various disease states.

A better understanding of the human gut microbiome, and the factors that drive variability with respect to microbiome composition, could lead to important insights in the areas of mental and physical health. A compelling body of research suggests that the gut-brain axis, which refers to the bidirectional influence between the gastrointestinal tract and the brain [9], might underlie the manifestation of disease and health dysfunction [10]. The composition of the gut microbiome plays a fundamental role in the gut-brain axis [11]; therefore, studying gut microbiome composition is of principle importance in order to better comprehend the etiology and consequences of physical and mental health dysfunction and refine maximally effective treatments.

Despite the relevance of gut microbiome composition to human health, there is not yet a gold standard, empirically based operationalization of gut microbiome alpha diversity, which captures the diversity of species within a given sample [12–14]. In the literature to date, many existing studies relating alpha diversity to human outcomes have selected individual alpha diversity indices to use in statistical models. For example, Vogt and colleagues [15] separately tested whether two alpha diversity metrics, abundance-based coverage estimator (Ace) and Chao1, were associated with diagnostic (i.e., presence of Alzheimer's disease) and biological (i.e., cerebrospinal spinal fluid) outcomes. Furthermore, a recently published study tested whether three alpha diversity metrics (observed operational taxonomic units (OTUs), Shannon, and Chao1) were associated with attention-deficit/hyperactivity disorder (ADHD) [16]. Additionally, Jiang and colleagues [17] found that Shannon diversity differed among patients who met criteria for major depressive disorder and healthy controls. As exemplified by these studies, a common strategy employed in the existing literature is to select *a priori* individual alpha diversity indices to represent gut microbiome diversity. An examination of the literature [18], including the mathematical expressions that are used to derive these alpha diversity measures [19], suggest that the construct of alpha diversity comprises both *richness* (i.e., how many different kinds of species are present within a given community) as well as *evenness* (i.e., the extent to which species within a community are held in even abundance with one another). Importantly, no published study to date has empirically tested whether these richness/evenness sub-constructs of alpha diversity emerge within human samples.

There exist a number of documented alpha diversity indices, thus providing an opportunity to empirically derive a composite measure based on multiple indices of alpha diversity. One commonly used method for building composite measures is to run an exploratory factor analysis and create composite variables based on factor loadings [20]. One particular study investigated the statistical benefits of using composite measures as compared to individual indicator variables. The results of this study suggest that model fit is improved when using composites compared to treating all indicator variables individually, which is especially true at smaller sample sizes [20]. Another study evaluated the reliability of composite measures of pain compared to individual pain measures. Findings suggest that the composite pain measures demonstrated superior reliability (as defined by stability over time throughout the follow-up period) compared to the individual measures [21]. These findings are consistent with what is known about the statistical properties of multi-item scales. That is, writings on quantitative methodology argue that composite measures reduce measurement error, increase reliability, and result in improved normality [22] thereby improving the overall robustness of the measure.

Therefore, the aims of the current study were twofold: 1) to empirically derive a robust and reliable measure of gut microbiome alpha diversity; and 2) to demonstrate the utility of the derived gut microbiome metric with regards to statistically predicting measures of human anthropometrics (i.e., age), behavior (alcohol use and fruit/vegetable consumption), and health characteristics (depressive symptoms).

Materials and methods

Study design

The present set of analyses employed data from a larger parent study, which tested the effectiveness of behavioral interventions among a sample of heavy alcohol drinkers who reported wanting to quit or reduce their drinking. The parent study was approved by the University of Colorado Boulder Institutional Review Board, and all participants provided written informed consent at the beginning of the baseline appointment. Only elements of the parent study that are relevant to the current set of analyses will be outlined here. The current investigation includes data from the baseline assessment prior to patients engaging in the intervention. At the time of the baseline assessment, patients completed self-report and researcher-administered assessments. Additionally, participants were provided a stool sample collection kit and were instructed to return a viable sample to the lab at the time of their first intervention session.

Participant selection and eligibility

Advertisements in mass media outlets (e.g., newspaper, radio), social media (e.g., Facebook), and direct mailings targeting the greater Front Range metropolitan area, which includes Boulder County, Denver, and suburbs of Denver, were the primary sources of participants. Participants who were deemed eligible following a phone screening procedure scheduled their baseline appointment. For the current set of analyses, participants included a final sample of 64 individuals who were selected for inclusion according to the following criteria: 1) must have been between the ages of 21 and 60 and provided informed consent; 2) had a primary DSM-5 diagnosis of alcohol use disorder; 3) were within 10 days of last drink at the time of the baseline assessment; 4) had been drinking heavily (criteria dependent upon individual's age, gender, and BMI) for a consistent period of time; 5) had a breath alcohol level of 0 at screening (to sign consent form); 6) were not currently taking any medications for the treatment of psychiatric disorders, including substance use disorders, mood disorders, and psychosis; 7) female participants were not pregnant, as indicated by a pregnancy test administered at baseline; 8) tested negatively for sedatives, opiates, cocaine, or amphetamine on drug screen at baseline; 9)

did not meet criteria for a psychotic disorder or bipolar disorder; 10) had a Clinical Institute Withdrawal Assessment (CIWA) score less than 8 (indicating no need for medical detox); 11) had expressed a desire during their initial screen to reduce the number of drinks they regularly consume; 12) had their baseline stool sample sequenced for analysis of microbiome composition by the initiation of data analysis.

Sample characteristics

All 64 participants completed each of the study procedures relevant to the current set of analyses outlined below. Behavioral and biological data for these 64 individuals were collected between September 2016 and December 2018. A summary of relevant sample characteristics at baseline can be found in [Table 1](#). The sample includes middle-aged adults (mean = 46.3 years), and is broadly representative of the Boulder/Denver metropolitan areas with regards to race (92.2% Caucasian). On average, participants reported moderate depression based on responses to the Beck Depression Inventory (BDI; mean = 12.5; [21]). A validated self-report measure of fruit and vegetable consumption [23] suggested that, at baseline, participants consumed 2.6 daily servings of fruits and vegetables on average.

Given the nature of the parent study, particular emphasis was placed on characterizing the sample with regards to alcohol use. At baseline, participants reported hazardous or harmful alcohol use indicative of dependence based on AUDIT score (mean = 19.4; [24]). Generally, an AUDIT score of eight or above is considered harmful use, while an AUDIT score of 11 or more is indicative of alcohol dependence [25]. The TLFB [26] was employed to quantify alcohol and other substance use across the preceding 30 days. At baseline, participants reported drinking approximately 23 days in the preceding month, and an average of 4.5 drinks on each drinking day (see [Table 1](#)).

Study procedures

Relevant to the current set of analyses, after arriving at the lab for their baseline appointment, qualified participants: 1) completed an informed consent form; 2) filled out baseline self-report questionnaires; 3) completed researcher-administered assessments; and 4) were also given a gut microbiome sample collection kit at the end of the session to be returned during their first scheduled intervention appointment.

Table 1. Characteristics of sample at baseline.

<u>Measure</u>	<u>Mean</u>	<u>Standard Deviation</u>
Age	46.34	9.30
AUDIT	19.39	8.07
Drinking Days/30 Days	23.27	5.47
Total Drinks/30 Days	103.82	62.26
Drinks/Drinking Day	4.47	2.37
BDI	12.45	7.54
Fruit and Vegetable Servings	2.57	2.00

AUDIT = Alcohol Use Disorders Identification Test; Drinking Days/30 = number of days over the past month alcohol was consumed as reported on the 30-day Timeline Follow Back (TLFB); Total Drinks/30 Days = total number of drinks reported on the TLFB over the preceding 30 days; Drinks/Drinking Day = average number of drinks per days alcohol was consumed as reported on the TLFB; BDI = Beck Depression Inventory; Fruit and Vegetable Servings = average number of fruit and vegetable servings consumed per day over the month preceding the baseline session according to Subar et al. (2001).

<https://doi.org/10.1371/journal.pone.0229204.t001>

The self-report measures that were administered included a number of individual difference measures, such as a demographics questionnaire that was used to collect information on participants' race, age, and sex. Participants also completed the Alcohol Use Disorders Identification Test (AUDIT) [24], which was used to assess symptoms of alcohol dependence. In order to assess the quality of participants' diet, participants completed The Fruits and Vegetable Screener [23], a 10-item measure that was developed by the National Cancer Institute to assess how many times in the previous month participants consumed different types of fruits and vegetables, including portion size questions for every food item. This measure was shown to be a useful estimate of median intakes of fruit and vegetable servings in U.S. populations, and has been recognized as a proxy measure for quality of diet. In addition to the self-report measures described above, participants worked with a research assistant to complete a TLFB assessment. The TLFB is a calendar-based assessment of daily alcohol and drug use [27] that has been shown to have good psychometric characteristics among a variety of drinker groups and can generate variables that provide a wide range of information about an individual's drinking (e.g., pattern, variability, and magnitude of drinking).

Gut microbiome sample collection

As mentioned, participants were provided with a fecal sample collection kit during the baseline assessment, and they were instructed to return the sample collection kit to the lab at the time of their first therapy appointment. Participants provided a stool sample by using a BD BBL CultureSwab Sterile, Media-free Swab (Cat. No. 220145, Becton, Dickinson and Company, Sparks, MD, USA) and a piece of used toilet paper. Participants were instructed to collect a sample within a day of providing it to the study staff. The sample was stored at -80°C until molecular processing.

Molecular processing

DNA was extracted using the PowerSoil DNA extraction kit (Cat No. 12888–100 & 12955–4, MoBio Laboratories, Carlsbad, CA, USA) according to the manufacturer's instructions. Marker genes in isolated DNA were PCR-amplified using HotStarTaq Master Mix (Cat No. 203433, Qiagen, Valencia, CA, USA) and 515 F (5' -GTGCCAGCMGCCGCGGTAA-3') 806 R (5' -GGACTACHVGGGTWTCTAAT-3') primer pair (Integrated DNA Technologies, Coralville, IA, USA) targeting the V4 hypervariable region of the 16S rRNA gene modified with a unique 12-base sequence identifier for each sample and the Illumina adapter as previously described in Caporaso et al., 2012 [28]. The thermocycling program consisted of an initial step at 94°C for 3 min followed by 35 cycles (94°C for 45 sec, 55°C for 1 min, and 72°C for 1.5 min), and a final extension at 72°C for 10 min. PCR reactions were run in duplicate and the products from the duplicate reactions were pooled and visualized on an agarose gel to ensure successful amplification. PCR products were cleaned and normalized using a SequalPrep Normalization Kit (Cat. No. A1051001, ThermoFisher, Waltham, MA, USA) following manufacturer's instructions. The normalized amplicon pool was sequenced on an Illumina MiSeq run by using V3 chemistry and 600 cycle, 2 x 300-bp paired-end sequencing. All sequencing and library preparation were conducted at the University of Colorado Boulder BioFrontiers Next-Gen Sequencing core facility, <https://bfccores.colorado.edu/sequencing-lab>.

Microbial data processing and preliminary analysis

After the raw sequencing files were generated, data were processed using Quantitative Insights Into Microbial Ecology (QIIME 2 version 2018.11) open source software [29]. A metadata file for each run was validated using the Keemei Google Sheets add-on [30]. Raw sequencing files

were demultiplexed and subsequently analyzed for quality by using the QIIME visualizer to examine the interactive quality plots. De-noising parameters (i.e., trimming sequence read lengths) were specified for each run according to the following: 1) demonstrated a minimum of 150 nucleotides in both forward and reverse reads; and 2) maintained quality scores at or above 30. Rarefaction was conducted according to [31]. A tree was generated for downstream phylogenetic diversity analyses by aligning sequences, generating an ‘unrooted’ tree, and then ‘rooting’ the tree. Taxonomy was assigned using the Ribosomal Database Project classifier [32] trained on the Greengenes 13_8 16S rRNA gene database [33]. Several alpha diversity metrics were calculated in QIIME 2 (see Table 2 for full list of alpha diversity measures that were calculated).

Statistical analysis

Aim 1: Deriving a composite gut microbiome diversity metric. *Selecting viable alpha diversity metrics.* Alpha diversity of species within the gut microbiome is a common outcome of interest that is examined in microbiome research. Analysis of the composition of the microbial community involves sequencing of the microbial community DNA, and matching the sequences to known functional genes [5]. Aim 1 of the current study involved empirically deriving a composite numerical metric that represents microbiome alpha diversity, which could be used as a predictor or outcome of interest in common statistical models. As discussed

Table 2. Alpha diversity measures and descriptions.

<u>Alpha Diversity Measure</u>	<u>Description</u>	<u>Citations</u>
Ace	An estimate of species richness using a correction factor	[18,34]
Observed OTUs	The number of different OTUs per sample	[16]
Chao1	A measure of species richness, which gives more weight to rare species	[34]
Margalef	Measure of species richness in a given area or community	[35,36]
Fisher alpha	Relationship between the number of species and the relative abundance of each species; referred to as a measure of richness	[37,38]
Faith pd	Sum of OTU branch lengths; takes into account phylogenetic distance between OTUs; the greater the number of unique OTUs, the higher this index will be	[18,39]
Brillouin d	Measures the diversity of the species present	[18]
Shannon	Calculation of richness and evenness using a natural logarithm; accounts for both relative abundance and evenness of the taxa present; referred to as an attempt for one metric to capture both richness and relative evenness	[18,40]
Enspie (ENS)	Measure of the number of equally abundant species; if the relative abundance of all species in a community were exactly identical, ENS would simply be the total number of species in that community	[41]
Menhinick	The ratio of the number of taxa to the square root of the sample size; referenced as "species richness"	[18,35]
Mcintosh e	A measure of how evenly/unevenly abundant taxa are in the sample	[18,42]
Simpson	Measure of relative abundance of the different species making up the sample richness; referred to as an attempt for one metric to capture both richness and relative abundance	[40,43]
Berger Parker d	A measure of relative richness of the abundant species; higher values indicate relative abundance disturbance	[18,44,45]
Strong	Assesses species relative abundance unevenness/dominance concentration	[46]
Simpson e	A measure of evenness based on the number of organisms and number of species	[43]
Pielou e	Measure of relative evenness of species richness	[12,42,47]
Heip e	A measure of evenness	[18,42]
Lladser pe	A measure of how much of the environment contains unsampled taxa	[18,48]

<https://doi.org/10.1371/journal.pone.0229204.t002>

above, a composite measure comprised of alpha diversity metrics could be a useful way to operationalize gut microbiome alpha diversity.

Alpha diversity metrics from the baseline microbiome samples output from QIIME 2 [29] were evaluated for inclusion in an exploratory factor analysis (EFA). Briefly, EFA is a statistical technique where the input is a correlation matrix of observed variables and the output is a summary of how each observed variable maps onto latent, unobserved “factors.” For further comprehensive descriptions of factor analytic methods see [49–52]. Because the number of latent factors is typically far fewer than the number of observed input variables, factor analysis is often considered to be a useful data reduction technique. Importantly, there are other statistical techniques that allow researchers to explore the underlying relationships among variables, including principal component analysis (PCA). Mathematically, the difference between PCA and EFA concerns the variance that is analyzed among the observed variables, such that all the variance in the observed variables is analyzed in PCA and only the shared variance among the observed variables is analyzed in EFA. Theoretically, the difference between PCA and EFA lies in the assumptions made about the nature of the relationship between the set of observed variables and the latent, unmeasured factor(s)/component(s). That is, factor analysis supposes that the measured indicator variables are caused by latent, unmeasured forces (i.e., the latent factors). As articulated in a useful chapter by [53], factor analysis explores the underlying processes that could have produced correlations among a given set of variables. In contrast, PCA supposes that the individual indicator variables cause the component(s). As such, components derived from a PCA analysis can be thought of simply as aggregates of the set of indicator variables. In other words, variables that comprise a component are merely empirically associated and are not necessarily thought to reflect an underlying process.

In the case of the present analysis, our choice to use EFA was based on our theoretical understanding of the alpha diversity construct. That is, using the EFA approach is consistent with our implicit hypothesis that an underlying process(es) (e.g., biological, environmental) cause(s) certain indicator variables to be correlated with one another. Therefore, our choice to use EFA over PCA was made due to the directionality of our theory-based understanding of the relationship among the alpha diversity construct and measurable indicator variables thought to measure alpha diversity. In summary, the choice to use EFA instead of PCA was an a priori, theory-driven decision. However, it is important to note that there is an ongoing debate among statistical theorists and researchers regarding when to use EFA and PCA [54]. When electing to use an EFA approach, we also had to determine whether EFA or confirmatory factor analysis (CFA) was most appropriate. While some have noted that alpha diversity of the gut microbiome may be comprised of richness and evenness processes (i.e., two proposed factors), we believe that it is still prudent to continue to explore whether this hypothesized structure of underlying processes is supported by empirical data in an exploratory fashion. Therefore, we elected to specify an EFA as opposed to testing the two-factor CFA model.

Prior to inclusion in the EFA, some alpha diversity metrics were excluded on the basis of being highly correlated with other alpha diversity metrics. Rather than selecting an ad-hoc cut-off for correlation strength, our strategy was to include the greatest number of observed variables as possible while simultaneously attempting to remove redundancy. Towards these ends, we evaluated the pattern of correlations between observed variables, and retained metrics that seemed to represent unique aspects of alpha diversity based on the correlation matrix. For example, ace, observed otus, chao1, margalef, and fisher alpha are correlated with one another at $r > .98$, therefore, we elected to retain one representative metric from that cluster of four metrics (fisher alpha). Some instances of high correlations were addressed by removing metrics due to elevated skewness and kurtosis. For example, removing mcintosh e because of

elevated skewness and kurtosis took care of the high correlation between mcintosh e and Shannon ($r = -.94$) and between mcintosh e and simpson ($r = -.96$).

Next, alpha diversity metrics were evaluated for inclusion in the EFA based on whether there was an indication that the variable did not meet the assumption of normality. Specifically, some alpha diversity metrics were excluded from the EFA if they demonstrated elevated skewness and/or kurtosis, which are numerical indicators of whether a given variable is normally distributed. We chose to exclude indicator variables on the basis of their skewness and/or kurtosis because non-normality can affect the integrity of some statistical tests, including factor analysis. Many statistical tests are based on the normality assumption [53]. Skewness and kurtosis are measures of normality such that a skewness value different from 0 indicates that the distribution deviates from symmetry, and a kurtosis value different from 0 indicates that the distribution deviates from normality in tail mass and shoulder [55]. Cain and colleagues (2016) evaluated the impact of skewness and kurtosis in the context of several statistical tests, including CFA. Specifically, authors simulated data with a known factor structure, manipulated the skewness and kurtosis of the data, and evaluated the Type I error rate of rejecting the factor model using the chi-squared-goodness of fit test. Results suggest that elevated skewness and kurtosis result in Type I error rate of 30% in factor analysis [56]. Authors suggest that researchers should report skewness and kurtosis so that results can be evaluated in the context of the underlying variables meeting the assumptions of the normal distribution. Additionally, authors suggest that researchers should consider taking steps to modify their analyses to mitigate the impact of non-normally distributed variables on their analyses. Based on the methodological literature discussed above, we chose to eliminate indicator variables that demonstrated elevated skewness and kurtosis values. While some methodological research has highlighted the consequences of including variables with non-normal distributions in univariate and multivariate statistical analyses [56–58], there does not yet exist any concrete guidance in terms of numerical thresholds for indices that measure normality. Specifically, there is no agreed upon answer to the question: “how high of a kurtosis/skewness value is *too high*?” Therefore, we decided to eliminate indicator variables that demonstrated elevated skewness and kurtosis relative to the other indicator variables. See Table 3 for a summary of alpha diversity metrics that were evaluated, and reason for exclusion from the EFA (if applicable).

Exploratory factor analysis. The final set of nine viable alpha diversity metrics was advanced to the factor analysis stage. First, a scree plot was constructed in R version 3.5.3 in order to determine how many factors likely underlie the set of diversity indicators. After evaluating the scree plot and eigenvalues associated with each indicator variable, an exploratory factor analysis was specified with the baseline microbiome data in R [59]. Factor analytic approaches demonstrate considerable utility as a data reduction tool by reducing many variables down to fewer factors assuming a latent factor structure. Such approaches summarize the patterns of correlations in the observed correlation matrix to yield a more parsimonious and robust structure that is often more reliable than the original variables [53].

Specifically, all nine viable alpha diversity metrics were input into an EFA function specified with a varimax rotation and maximum likelihood factoring method. Based on the results from the EFA (see results section), a two-factor solution was adopted consistent with the theoretically relevant richness and evenness factors. Based on the pattern of factor loadings, searching the literature to derive a theoretical understanding of each alpha diversity measure, and examining the mathematical equations that are used to calculate each measure [19], the two factors were labeled “richness” and “evenness.” While we allowed the results of the EFA to guide our decision to retain two factors we used our knowledge of the literature to both characterize those two factors as “richness” and “evenness” and to guide our decisions in an iterative way throughout the model building process, which is considered to be an accepted practice in the

Table 3. Summary of alpha diversity metrics.

<u>Alpha Diversity Measure</u>	<u>Included in EFA</u>	<u>Reason for exclusion</u>
Ace	No	High inter-correlation with other indicators
Observed OTUs	No	High inter-correlation with other indicators
Chao1	No	High inter-correlation with other indicators
Margalef	No	High inter-correlation with other indicators
Fisher alpha	Yes	
Faith pd	Yes	
Brillouin d	No	High inter-correlation with other indicators
Shannon	Yes	
Enspie (ENS)	Yes	
Menhinick	Yes	
Mcintosh e	No	Elevated Skewness/Kurtosis
Simpson	No	Elevated Skewness/Kurtosis
Berger Parker d	No	Elevated Skewness/Kurtosis
Strong	Yes	
Simpson e	Yes	
Pielou e	No	Elevated Skewness/Kurtosis
Heip e	Yes	
Lladser pe	Yes	

OTUs = Operational taxonomic units; See Fig 1 for the correlation value between each pair of alpha diversity indices evaluated.

<https://doi.org/10.1371/journal.pone.0229204.t003>

context of EFA [53]. That is, when the scree plot suggested a two-factor solution, we hypothesized that the EFA results would support a “richness” and “evenness” factor structure based on our knowledge of the alpha diversity literature. Because we hypothesized that the two-factor solution may relate to “richness” and “evenness” factors, we chose to specify the model with a varimax rotation. Varimax, the most commonly used rotation, is a variance maximizing procedure that maximizes the variance of factor loadings (i.e., makes high loadings higher and low loadings lower) [53]. Varimax assumes, and mathematically encourages, an orthogonal factor structure [60,61]. Importantly, an orthogonal factor structure is consistent with the working theoretical understanding of how the richness and evenness sub-categorizations of alpha diversity measures relate to one another. That is, Mason and colleagues [62] argue that richness and evenness are orthogonal to one another (i.e., vary independently of one another). Thus, specifying the EFA with a varimax rotation is consistent with a theory-driven interpretation of the factor structure.

In order to derive composite scores for each factor, each alpha diversity metric was transformed into a z-score. This standardization step was crucial given that each alpha diversity metric is measured on a different scale. Individual alpha diversity metrics were reverse scored according to theory prior to being converted into a standardized scale. Next, alpha diversity z-scores were averaged together according to the factor structure. This process resulted in a “richness score” and an “evenness score” for each participant. These composite measures were then used in subsequent statistical models (see below).

Aim 2: Association between derived diversity metric and human variables of interest.

Aim 2 of the study set out to assess the utility of the gut microbiome diversity metrics derived in Aim 1 in the context of statistically predicting human outcomes. Towards these ends, we set out to evaluate the association between gut microbiome alpha diversity, age, diet, depressive symptoms, and recent alcohol consumption. These four human outcome variables were

selected due to their apparent association with gut microbiome composition and their broad relevance to human health [63,64].

Simple regression models tested the relationships between the microbiome richness and evenness composite measures and the four human health-related variables of interest. Diet was operationalized as the average number of fruits and vegetables consumed over the previous 30 days per the Fruits and Vegetable Screener [23]. Alcohol use was operationalized as the total number of standard alcoholic drinks consumed over the 30 days preceding the baseline appointment according to the TLFB assessment [27]. Participant age was based on their response on the demographics questionnaire. Finally, symptoms of depression was operationalized as participants' scores on the BDI self-report measure [65].

First, baseline gut microbiota richness and evenness were regressed on baseline drinking behavior in separate models. Second, baseline gut microbiota richness and evenness were regressed on participant age in separate models. Third, baseline gut microbiota richness and evenness were regressed on average daily fruit and vegetable consumption, which is a proxy measure for quality of diet. Finally, baseline gut microbiota richness and evenness were regressed on BDI scores, which is a measure of participants' current depressive symptoms.

In addition to testing the association between the human outcomes and the derived composite measures of microbiome diversity, simple regression models testing the association between alcohol use, age, diet, BDI scores and individual alpha diversity metrics were also specified. These models were identical to the models shown above, except instead of regressing the *composite* diversity measures on the human health and behavior variables of interest, *individual* alpha diversity measures (Faith pd, Strong) were regressed on each human outcome variable in separate models. This was meant to determine whether the composite measures derived in Aim 1 performed differently in their ability to statistically predict common human health and behavior variables of interest compared to individual diversity metrics (e.g., Faith pd). We chose to run regression models for only two individual measures of alpha diversity in an attempt to be sensitive to the issue of multiple comparisons because as the number of tests increase, so does the Type I error rate. Thus, we selected one "richness" measure and one "evenness" measure a priori in order to explore the question of whether the composite metrics provide predictive power over and above individual metrics of alpha diversity.

Results

Aim 1: Deriving a gut microbiome diversity metric

Selecting viable alpha diversity metrics. The full list of alpha diversity measures that were output by QIIME are shown and described in Table 2. The correlations among these alpha diversity measures are shown in Fig 1. Prior to inclusion in the EFA, some alpha diversity metrics were excluded on the basis of being highly correlated with other alpha diversity metrics. Rather than selecting an ad-hoc cutoff for correlation strength, our strategy was to include the greatest number of observed variables as possible while simultaneously attempting to remove redundancy. Towards these ends, we evaluated the pattern of correlations between observed variables, and retained metrics that seemed to represent unique aspects of alpha diversity based on the correlation matrix. For example, ace, observed otus, chao1, margalef, fisher alpha are correlated with one another at $r > .98$, therefore, we elected to retain one representative metric from that cluster of four metrics (fisher alpha). Some instances of high correlations were addressed by removing metrics due to elevated skewness and kurtosis. For example, removing mcintosh e because of elevated skewness and kurtosis took care of the high correlation between mcintosh e and Shannon ($r = -.94$) and between mcintosh e and simpson ($r = -.96$). See Fig 1 for the correlation value between each pair of alpha diversity indices evaluated.

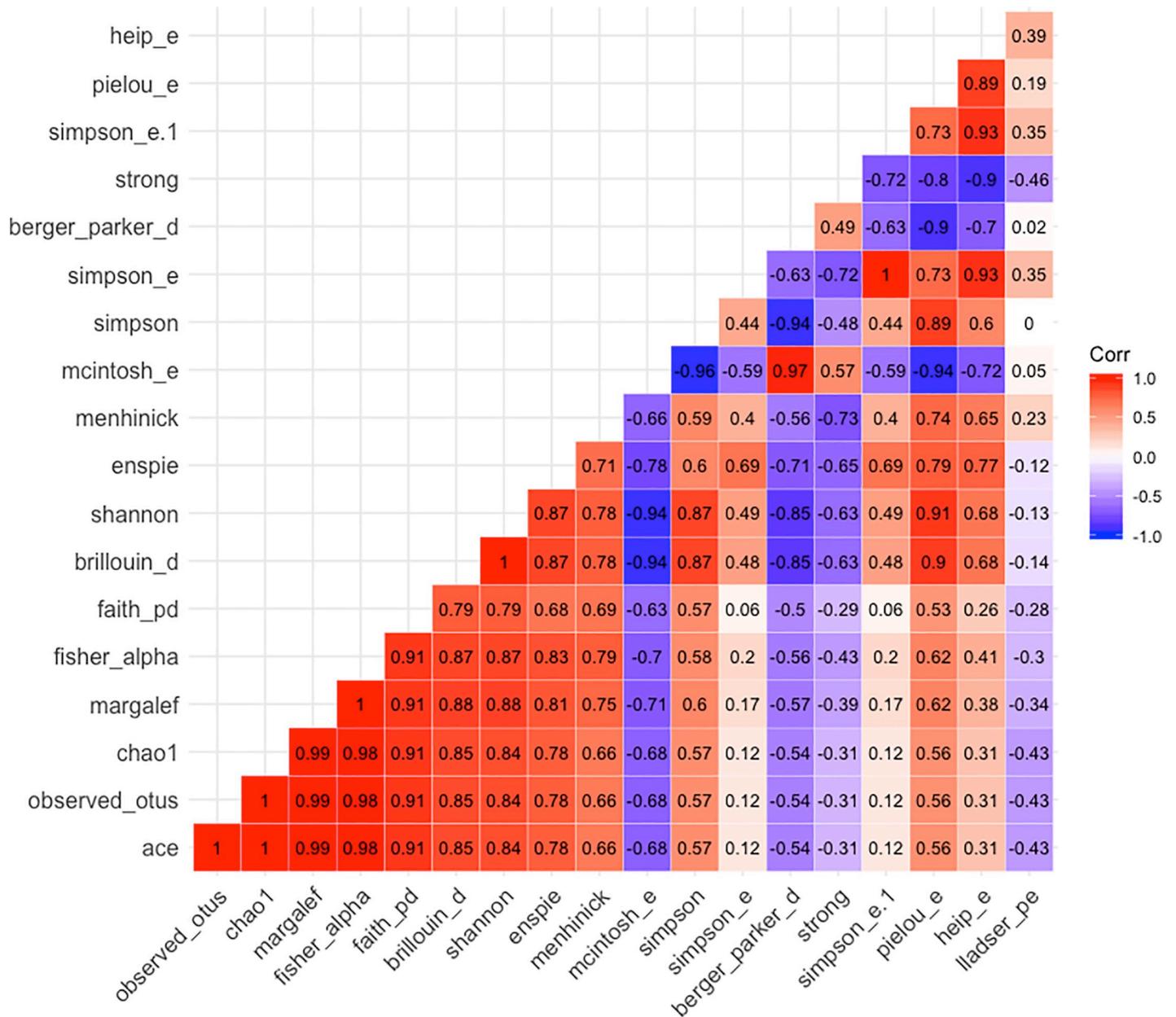


Fig 1. Correlations among alpha diversity indices. Correlations among all calculated alpha diversity indices are displayed. Positive correlations displayed in blue, and negative correlations displayed in red. Stronger correlations denoted by greater opaqueness.

<https://doi.org/10.1371/journal.pone.0229204.g001>

In addition to examining correlations among the set of alpha diversity indicator variables, the distributions of the indicator variables were also examined in order to determine viability for inclusion in the EFA. Specifically, skewness and kurtosis were evaluated for each alpha diversity indicator variable. At this stage, 4 metrics were excluded on the basis of demonstrating elevated skewness/kurtosis, as evidenced by their skewness and kurtosis: Simpson (skewness = -5.03, kurtosis = 31.97), Pielou e (skewness = -1.84, kurtosis = 7.05), Mcintosh e (skewness = 2.71, kurtosis = 12.12), and Berger Parker d (skewness = 2.78, kurtosis = 11.54).

Exploratory factor analysis. The final set of viable alpha diversity indicator variables were included in an EFA (see Table 3). In order to determine the likely number of factors

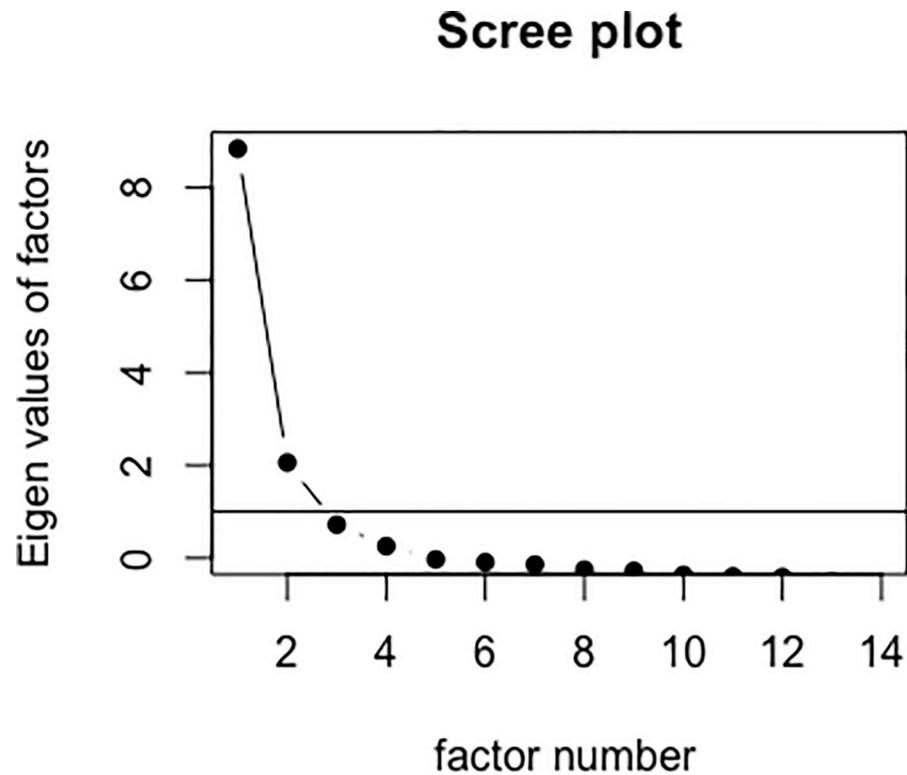


Fig 2. Scree plot of alpha diversity indices. Scree plot showing likely number of factors underlying the relationships among alpha diversity indices.

<https://doi.org/10.1371/journal.pone.0229204.g002>

among the sample of diversity indices, a scree plot was evaluated. Interpretation of the resulting scree plot (see Fig 2) suggests a two-factor solution. Specifically, two factors demonstrate eigenvalues greater than or equal to one, and common approaches to factor analysis suggest that eigenvalues greater than one are large enough to be retained [53]. An examination of the literature on diversity suggests that measures of alpha diversity capture two sub-constructs of diversity: richness and evenness [62,66–68]. Therefore, the scree plot, eigenvalues, and theory based on past literature all indicate a two-factor solution.

A summary of the exploratory factor analysis results is found in Table 4. As shown in Table 4, the factor loading pattern again supports a two-factor solution, as each alpha diversity indicator variable loads cleanly onto either of the two factors (i.e., factor loading $>.5$). Importantly, none of the alpha diversity indicator variables show a cross-loading pattern. That is, there are no indicator variables where the factor loading is $>.5$ for both factors (see Table 4). To ensure that these decisive loading patterns were consistent with theory, relevant literature/each measure's mathematical expression [19] was consulted before categorizing a given alpha diversity measure as a measure of "richness" or "evenness." For example, Faith pd loaded strongly onto the factor that has been labeled "richness" (.91 loading). Consultation of the literature confirms that it is appropriate to categorize Faith pd as a richness measure given that it is a measure of cumulative OTU branch lengths, and the greater the number of OTUs, the higher the Faith pd measure will be. This type of analysis was repeated for each alpha diversity measure.

Finally, in one case the factor loading was negative (i.e., Strong). In this case, an evaluation of the literature suggests that this negative factor loading is theoretically meaningful. The literature and evaluation of the mathematical equation associated with strong suggest that this measure does not measure evenness, but rather it measures *unevenness*, or species dominance.

Table 4. Results of the exploratory factor analysis.

<u>Alpha Diversity Measure</u>	<u>“Richness” Factor Loading</u>	<u>“Evenness” Factor Loading</u>	<u>h2</u>	<u>Final Categorization</u>
Fisher alpha	1.0		1.0	Richness
Faith pd	.91		.84	Richness
Shannon	.86		.89	Richness
Enspie (ENS)	.81		.90	Richness
Menhinick	.77		.74	Richness
Strong*		-.80	.81	Evenness
Simpson e		.93	.90	Evenness
Heip e		.92	1.0	Evenness
Lladser pe		.55	.40	Evenness

h2 = proportion of variance of each measure explained by the factors

* denotes measure that was “reverse coded” prior to z-score conversion on the basis of direction of factor loading and evaluation of relevant literature and mathematical expression; factor loadings > .5 shown.

<https://doi.org/10.1371/journal.pone.0229204.t004>

In other words, a higher score on the Strong measure suggests that the community being measured is more dominated by a fewer number of species [19]. Therefore, the negative factor loading in concert with an evaluation of the relevant literature and mathematical expression suggest that this measure should be reverse coded prior to including it in a composite measure. The Strong metric demonstrated observed values that ranged from 0–1. Therefore, in order to reverse code this measure, the inverse of the original measure was estimated by the following formula: 1-original measure. This reverse coding process was completed prior to Strong being converted to a z-score. The culmination of this set of analyses resulted in empirically derived and theoretically interpretable alpha diversity composite metrics (i.e., richness and evenness), which were then used to represent microbiome composition in statistical models.

Aim 2: Association between derived diversity metric and human variables of interest

In order to assess the utility of the microbiome composition metrics derived in Aim 1, simple regression models tested the association between the richness and evenness composite measures and several variables of interest related to human health. Specifically, the associations between richness, evenness, age, recent alcohol consumption, symptoms of depression, and recent fruit and vegetable consumption were tested in separate models according to the models outlined in the statistical analysis section above. Results from each of these models are summarized in Table 5.

Table 5. Associations between microbiome composites and variables of interest.

<u>Variable of Interest</u>	<u>beta coefficient (p-value)</u>			
	<u>Richness Composite</u>	<u>Evenness Composite</u>	<u>Faith pd</u>	<u>Strong</u>
Drinks/30 days	-.004 (.037)	.00091 (.302)	-.012 (.02)	.0011 (.257)
Fruit/veggie servings	.008 (.913)	.045 (.13)	-.074 (.72)	.0038 (.287)
Age	-.014 (.251)	-.0025 (.677)	-.047 (.173)	.000065 (.92)
BDI	.0023 (.881)	.021 (.0032)	-.028 (.515)	-.0011 (.185)

BDI = Beck Depression Inventory; beta values and *p*-values correspond to the parameter estimate and *p*-value corresponding with the variable of interest in each row regression models specified in the statistical analysis section (where richness/evenness composites were regressed on each variable of interest in separate models); statistically significant parameter estimates and corresponding *p*-values at an $\alpha < .05$ two-tailed threshold are shown in boldface type.

<https://doi.org/10.1371/journal.pone.0229204.t005>

Given the aims of the current investigation, we were not only concerned with whether the composite measures of microbiome composition were significantly associated with the variables of interest. We were interested in what the pattern of results suggests regarding the utility of the derived microbiome composite measures. Importantly, the pattern of results lends support to what we determined in Aim 1: richness and evenness are two subcomponents of microbiome composition. In fact, richness and evenness seem to relate differently to human behavior. That is, richness, but not evenness, of species within the human gut microbiome was associated with recent alcohol consumption ($b = -.004$, $p = .037$, 95% CI: -0.0074, -0.00024) such that greater alcohol consumption is associated with less richness of species within the microbiome (as expected based on our working theoretical model). Additionally, richness and evenness of the gut microbiome appear to relate differently to current depressive symptoms. That is, evenness, but not richness, is significantly associated with participants' scores on the BDI ($b = .021$, $p = .0032$, 95% CI: 0.0075, 0.035), such that greater depressive symptoms is associated with less evenness of microbial species. Though only two of the human variables of interest demonstrated differing relationships with richness and evenness based on meeting significance thresholds ($p < .05$), the pattern of results more generally (e.g., directionality and magnitude of parameter estimates) suggests that richness and evenness might differ with regard to their associations with human behavior (e.g., diet, alcohol consumption) and human characteristics (e.g., age, mental health).

Also of note is the finding that the gut microbiome composite measures detected significant associations when individual alpha diversity measures did not. Specifically, model results suggest that evenness of microbial species, operationalized by the evenness composite, is significantly associated with symptoms of depression. However, this effect was not detected when microbiome evenness was operationalized by the Strong metric (an individual alpha diversity metric thought to capture (un)evenness of microbial species). This suggests that the microbiome composite measures may be a more robust and powerful way to capture composition of the gut microbiome.

The impact of analytic decision making on results. Throughout the analytic process, decisions were made that could have impacted the results. Specifically, we made the decision to exclude some observed alpha diversity metrics from the EFA if a given metric demonstrated elevated skewness/kurtosis or if a given metric was redundant (i.e., highly correlated) with other alpha diversity metrics. In doing so, we intended to include a representative set of alpha diversity indices that contributed unique information and did not violate the assumption of normality. Though the intention behind these decisions was to enhance the robustness and integrity of the analysis, we acknowledge that these decisions could have impacted the results. In an effort to determine the possible impact of these decisions, we re-ran the scree plot and EFA including all alpha diversity metrics.

Consistent with the primary results, the scree plot in which all alpha diversity indices were included again indicated a two-factor solution. Thus, an EFA was specified with two factors. The results of this model are summarized in [Table 6](#). The results are broadly consistent with the EFA results from the model that included the more exclusive list of alpha diversity indices. That is, the factor loadings support a two-factor solution. Each of the alpha diversity indices that was included in the more exclusive model demonstrates consistency between the two models in terms of the pattern of factor loading. For example, Shannon was loaded strongly onto the "richness" factor in the exclusive EFA (factor loading = .86), and Shannon loaded strongly onto the "richness" factor in the all-inclusive model (factor loading = .81). Interestingly, several alpha diversity metrics demonstrate a cross-loading pattern (the metric loads $> .5$ onto both the "richness" factor and the "evenness" factor) in the context of the model that includes the full set of alpha diversity metrics. Despite these cross-loadings, the pattern of

Table 6. Results of the exploratory factor analysis with all observed alpha diversity metrics.

Alpha Diversity Measure	“Richness” Factor Loading	“Evenness” Factor Loading	h ²
Ace	.99		.99
Observed otus	.99		.99
Chao1	.99		.99
Margalef	.98		.98
Fisher alpha	.96		.96
Faith pd	.91		.83
Brillouin d	.82	.56	.98
Shannon	.81	.57	.98
Enspie (ENS)	.68	.61	.84
Menhinick	.61	.52	.64
Mcintosh e	-.65	-.67	.86
Simpson	.57	.58	.65
Simpson e		.92	.85
Berger parker d	-.51	-.68	.73
Strong*		-.81	.70
Pielou e		.85	.96
Heip e		.98	.99
Lladser pe			.40

h² = proportion of variance of each measure explained by the factors; factor loadings > .5 shown

<https://doi.org/10.1371/journal.pone.0229204.t006>

factor loadings suggests that one factor represents “richness” and the other factor represents “evenness.” Given that the interpretation of results was consistent across the two models, it is reasonable to suggest that the decisions made regarding which factors to ultimately include in the EFA did not influence the results in a way that impacts the broad conclusions drawn.

Discussion

Summary of results

The current study aimed to make a novel methodological contribution to the literature by developing an empirically-derived composite metric(s) that captures gut microbiome alpha diversity, which could be used in common statistical modeling approaches (Aim 1). Alpha diversity emerged as a relevant construct due to the fact that alpha diversity broadly captures within-participant microbial diversity. Additionally, several alpha diversity metrics have been developed and documented [19], thus providing an opportunity to derive a robust composite measure from a host of existing metrics.

Results provide key insights into the broad construct of microbial diversity, and offer a novel method for operationalizing gut microbiome diversity. Results of the EFA empirically confirm what has been noted in the literature. That is, existing alpha diversity metrics capture two facets of microbial composition: “richness” and “evenness.” The current study provides one possible method by which to calculate these composite measures. The method suggested here is to reverse code metrics as needed, convert all metrics to standardized z-scores, and average together metrics that are characterized as “evenness” metrics and average together metrics that are characterized as “richness” metrics.

Results from Aim 2 of the current investigation demonstrate the utility of this methodological approach with regard to investigating the relationship between microbiome composition and variables related to human health. Results suggest that the microbiome composition

composite metrics that were derived using the method described in the manuscript demonstrate significant associations with key variables of interest. Specifically, in this sample, microbiome richness was significantly associated with recent alcohol consumption, and microbiome evenness was significantly associated with self-reported symptoms of depression.

Methodological considerations

The current study employs samples from human fecal samples, which were sequenced in order to extract gut microbiome data. The sample is a methodological strength and enhances the novelty of the current investigation for two main reasons. First, the sample is a clinical sample comprised of participants who meet criteria for a common health condition (alcohol use disorder), thus providing an opportunity to examine the relationship between the gut microbiome and variables related to human health using a clinical sample of treatment-seeking individuals. Second, the current study employed one of the largest sample sizes to date ($N = 64$) compared to existing studies on the human gut microbiome and alcohol use. For example, one study that investigated the effect of diet on the human gut microbiome was published in the journal *Nature* in 2014 with a sample size of $N = 10$ [6]. The existing authoritative studies on human gut microbiome composition and alcohol have sample sizes ranging between $N = 13$ –66 [7,69,70]. Thus, the current study makes a robust addition to the literature at the intersection of clinical health and the human gut microbiome. Additionally, our analytical approach represents a methodological strength, as it is the first known application of factor analysis to study the underlying structure of the alpha diversity construct in relation to the human gut microbiome, thus enhancing the novelty and rigor of our analysis. Ultimately, the current investigation provides one of the first sources of empirical evidence of key alpha diversity sub-constructs (richness and evenness) within the human gut microbiome.

In addition to the methodological strengths discussed above, the study also has a number of limitations, and our results should be interpreted in the context of these considerations. First, though the sample size is moderate to large compared to previous published studies employing fecal samples from human clinical samples in the context of alcohol use, the current sample size is considered small in the context of factor analytic approaches, as discussed in Tabachnick & Fidell [53]. Authors note, however, that the sample size guidelines might be relaxed if the factor analysis solution include several high loading marker variables ($< .8$), as is the case in the current set of analyses. Additionally, the health-related variables of interest included in the current set of analyses (fruit and vegetable consumption, alcohol use behavior, depression symptomatology) are based on self-report data. Though each of the respective self-report measures that were used to operationalize our variables of interest are validated, psychometrically sound measures, future studies should consider ways of integrating objective measures of behavior.

Furthermore, the statistical analyses conducted in this study included several statistical tests. That is, 16 individual regression models were run in order to test the associations between health-related variables of interest (i.e., age, alcohol consumption, fruit and vegetable servings, BDI) and alpha diversity measures (i.e., richness composite, evenness composite, Faith pd, Strong). Type I error rate, which is the likelihood of getting a “false positive” result in the absence of a “true” effect, increases as the number of statistical tests increases. Given that 16 statistical tests were run, it is possible that some of the significant effects reported in the present investigation were the result of Type I error. Importantly, efforts were made to reduce the threat of Type I error. Specifically, we planned our statistical tests a priori, such that we selected the four health-related variables of interest and the two individual alpha diversity metrics prior to the start of data analysis. Finally, these data are cross-sectional in nature because

these variables were measured at one single time point. Ultimately, our interest is in characterizing the causal relationships between manifestations of health and biological mechanisms. Therefore, future studies should examine microbiome composition over time, and how stability and/or changes in the human gut microbiome relate to changes in other biological and behavioral variables of interest over time.

Directions for future research

While we feel that our study represents a meaningful contribution to the literature, our work also highlights the many important questions that were not addressed in the present investigation. Future investigations should examine the impact of moderating factors on gut microbiome alpha diversity. The study sample in the current investigation was 92.19% Caucasian and 49% female. Due to the sample size ($n = 64$ participants) we were not sufficiently powered to evaluate the impacts of race and gender, two potentially important moderating factors, on the gut microbiome. Therefore, it is not possible to determine if the results here extend to non-Caucasian individuals, and both male and female gender. Future studies with larger sample sizes and higher racial diversity should be conducted to address these issues. Additionally, future studies should seek to validate the factor structure found in the present investigation. The two-factor solution and the “richness” and “evenness” factor names are consistent with theory and with the interpretation of the EFA results in this study. It will be important to replicate these results using different samples. Towards these ends, researchers could run exploratory and/or confirmatory factor analyses using other datasets comprised of human fecal microbiome samples.

Significance and concluding remarks

Based on the current investigation, results suggest that gut microbiome alpha diversity is a multi-dimensional construct, and thus researchers should consider operationalizing the gut microbiome alpha diversity with this complexity in mind. As demonstrated by the results of Aim 2 of the current investigation, different facets of gut microbiome complexity (i.e., richness vs. evenness) show different relationships with key variables of interest. Failing to account for the heterogeneity of the gut microbiome construct could mask true, significant associations between facets of microbiome alpha diversity and key variables related to human health. With this in mind, we are making two concrete methodological recommendations:

1. We suggest that the operationalization of microbiome alpha diversity should be comprised of separate “richness” and “evenness” measures. We make this suggestion because we found evidence that “richness” and “evenness” are two distinct sub-constructs of the broader alpha diversity construct. One method by which this can be accomplished is for researchers to create composite measures comprised of the individual alpha diversity indices that we used to construct our richness/evenness measures (i.e., richness = Fisher alpha, Faith pd, Shannon, Enspie, Menhinick; evenness = Strong, Simpson e, Heip e, Lladser pe). In order to further explore and justify this proposed factor structure, we also suggest that researchers test this factor structure by specifying CFA models.
2. When investigating the relationship between human health and behavioral variables of interest and microbiome diversity, researchers should use composite measures, comprised of multiple individual alpha diversity indices that measure the same alpha diversity sub-construct (i.e., richness, evenness). We make this suggestion because our results suggest that composite measures can provide information over and above individual alpha diversity metrics.

Based on our analysis, these recommendations should be considered for the following reasons: 1) composite measures are more robust and reliable compared to individual indices; 2) the richness and evenness sub-constructs are empirically derived based on analysis of a human sample; and 3) these composite measures demonstrate utility in predicting key health-related variables of interest. It is our hope that this investigation serves as a methodological guide for researchers across a host of disciplines.

Supporting information

S1 Data.
(CSV)

Acknowledgments

We are grateful to Jared D. Heinze for conducting the molecular processing of samples. We thank the University of Colorado Boulder BioFrontiers Institute Next-Gen Sequencing Core Facility, which performed the Illumina sequencing.

Author Contributions

Conceptualization: Sarah L. Hagerty.

Formal analysis: Sarah L. Hagerty, Angela D. Bryan.

Funding acquisition: Kent E. Hutchison.

Methodology: Sarah L. Hagerty, Angela D. Bryan.

Resources: Kent E. Hutchison, Christopher A. Lowry.

Supervision: Kent E. Hutchison, Christopher A. Lowry, Angela D. Bryan.

Writing – original draft: Sarah L. Hagerty.

Writing – review & editing: Kent E. Hutchison, Christopher A. Lowry, Angela D. Bryan.

References

1. Hagerty SL, Ellingson JM, Helmuth TB, Bidwell LC, Hutchison KE, Bryan AD. An Overview and Proposed Research Framework for Studying Co-Occurring Mental- and Physical-Health Dysfunction. *Perspect Psychol Sci.* 2019 Jul 1; 14(4):633–45. <https://doi.org/10.1177/1745691619827010> PMID: 31173535
2. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. *PLOS Biol.* 2012 Aug 14; 10(8): e1001377. <https://doi.org/10.1371/journal.pbio.1001377> PMID: 22904687
3. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project [Internet]. *Nature.* 2007 [cited 2018 Apr 26]. Available from: <https://www.nature.com/articles/nature06244>
4. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. *Nature.* 2006 Dec 21; 444(7122):1022–3. <https://doi.org/10.1038/4441022a> PMID: 17183309
5. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature.* 2012; 489(7415):220. <https://doi.org/10.1038/nature11550> PMID: 22972295
6. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature.* 2014 Jan 23; 505(7484):559–63. <https://doi.org/10.1038/nature12820> PMID: 24336217
7. Mutlu EA, Gillevet PM, Rangwala H, Sikaroodi M, Naqvi A, Engen PA, et al. Colonic microbiome is altered in alcoholism. *Am J Physiol—Gastrointest Liver Physiol.* 2012 May 1; 302(9):G966–78. <https://doi.org/10.1152/ajpgi.00380.2011> PMID: 22241860

8. Wang HJ, Zakhari S, Jung MK. Alcohol, inflammation, and gut-liver-brain interactions in tissue damage and disease development. *World J Gastroenterol WJG*. 2010 Mar 21; 16(11):1304–13. <https://doi.org/10.3748/wjg.v16.i11.1304> PMID: 20238396
9. Carabotti M, Scirocco A, Maselli MA, Severi C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann Gastroenterol Q Publ Hell Soc Gastroenterol*. 2015; 28(2):203–9.
10. Grenham S, Clarke G, Cryan JF, Dinan TG. Brain–Gut–Microbe Communication in Health and Disease. *Front Physiol* [Internet]. 2011 Dec 7 [cited 2017 Nov 29];2. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232439/>
11. Chen X, D'Souza R, Hong S-T. The role of gut microbiota in the gut-brain axis: current challenges and perspectives. *Protein Cell*. 2013 Jun 1; 4(6):403–14. <https://doi.org/10.1007/s13238-013-3017-x> PMID: 23686721
12. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology*. 2007; 88(10):2427–2439. <https://doi.org/10.1890/06-1736.1> PMID: 18027744
13. Peet RK. The measurement of species diversity. *Annu Rev Ecol Syst*. 1974; 5(1):285–307.
14. The Human Microbiome Project Consortium, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, et al. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012 Jun; 486(7402):207–14. <https://doi.org/10.1038/nature11234> PMID: 22699609
15. Vogt NM, Kerby RL, Dill-McFarland KA, Harding SJ, Merluzzi AP, Johnson SC, et al. Gut microbiome alterations in Alzheimer's disease. *Sci Rep*. 2017 Oct 19; 7(1):13537. <https://doi.org/10.1038/s41598-017-13601-y> PMID: 29051531
16. Prehn-Kristensen A, Zimmermann A, Tittmann L, Lieb W, Schreiber S, Baving L, et al. Reduced microbiome alpha diversity in young patients with ADHD. *PLOS ONE*. 2018 Jul 12; 13(7):e0200728. <https://doi.org/10.1371/journal.pone.0200728> PMID: 30001426
17. Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, et al. Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav Immun*. 2015 Aug 1; 48:186–94. <https://doi.org/10.1016/j.bbi.2015.03.016> PMID: 25882912
18. Alpha and Beta Diversity Explanations and Commands [Internet]. QIIME 2 Forum. 2017 [cited 2019 Mar 26]. Available from: <https://forum.qiime2.org/t/alpha-and-beta-diversity-explanations-and-commands/2282>
19. Chiarucci A, Bacaro G, Scheiner SM. Old and new challenges in using species diversity for assessing biodiversity. *Philos Trans R Soc B Biol Sci*. 2011; 366(1576):2426–2437.
20. Landis RS, Beal DJ, Tesluk PE. A Comparison of Approaches to Forming Composite Measures in Structural Equation Models. *Organ Res Methods*. 2000 Apr 1; 3(2):186–207.
21. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain*. 1999; 2(83):157–62.
22. Yuan K-H, Bentler PM, Kano Y. On averaging variables in a confirmatory factor analysis model. *Behaviormetrika*. 1997; 24(1):71–83.
23. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol*. 2001 Dec 15; 154(12):1089–99. <https://doi.org/10.1093/aje/154.12.1089> PMID: 11744511
24. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption—II. *Addict Abingdon Engl*. 1993 Jun; 88(6):791–804.
25. Tsai M-C, Tsai Y-F, Chen C-Y, Liu C-Y. Alcohol Use Disorders Identification Test (AUDIT): establishment of cut-off scores in a hospitalized Chinese population. *Alcohol Clin Exp Res*. 2005 Jan; 29(1):53–7. <https://doi.org/10.1097/01.alc.0000151986.96710.e0> PMID: 15654291
26. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol*. 1982; 13:290–312.
27. Sobell LC, Sobell MB. Timeline Follow-Back. In: *Measuring Alcohol Consumption* [Internet]. Humana Press, Totowa, NJ; 1992 [cited 2018 Mar 8]. p. 41–72. Available from: https://link.springer.com/chapter/10.1007/978-1-4612-0357-5_3
28. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012 Aug; 6(8):1621–4. <https://doi.org/10.1038/ismej.2012.8> PMID: 22402401
29. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010 May; 7(5):335–6. <https://doi.org/10.1038/nmeth.f.303> PMID: 20383131

30. Rideout JR, Chase JH, Bolyen E, Ackermann G, González A, Knight R, et al. Keemei: cloud-based validation of tabular bioinformatics file formats in Google Sheets. *GigaScience* [Internet]. 2016 Dec 1 [cited 2019 Mar 27]; 5(1). Available from: <https://academic.oup.com/gigascience/article/5/1/s13742-016-0133-6/2720989>
31. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*. 2017; 5(1):27. <https://doi.org/10.1186/s40168-017-0237-y> PMID: 28253908
32. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Env Microbiol*. 2007; 73(16):5261–5267.
33. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved GreenGenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2012; 6(3):610. <https://doi.org/10.1038/ismej.2011.139> PMID: 22134646
34. Chao A. Nonparametric estimation of the number of classes in a population. *Scand J Stat*. 1984;265–270.
35. Magurran AE. *Measuring biological diversity*. John Wiley & Sons; 2013.
36. Magurran AE. *Measuring biological diversity*. 2004. Malden Blackwell. 2004;
37. Fisher RA, Corbet AS, Williams CB. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *J Anim Ecol*. 1943 May; 12(1):42.
38. Willis A. Rarefaction, alpha diversity, and statistics. *bioRxiv* [Internet]. 2017 Dec 11 [cited 2019 Mar 16]; Available from: <http://biorxiv.org/lookup/doi/10.1101/231878>
39. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv*. 1992 Jan 1; 61(1):1–10.
40. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol Evol*. 2014; 4(18):3514–24. <https://doi.org/10.1002/ece3.1155> PMID: 25478144
41. Chase JM, Knight TM. Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecol Lett*. 2013; 16(s1):17–26.
42. Heip C. A New Index Measuring Evenness. *J Mar Biol Assoc U K*. 1974 Aug; 54(3):555–7.
43. Simpson EH. Measurement of diversity. *Nature*. 1949; 163(4148):688.
44. Berger WH, Parker FL. Diversity of Planktonic Foraminifera in Deep-Sea Sediments. *Science*. 1970 Jun 12; 168(3937):1345–7. <https://doi.org/10.1126/science.168.3937.1345> PMID: 17731043
45. Caruso T, Pigino G, Bernini F, Bargagli R, Migliorini M. The Berger–Parker index as an effective tool for monitoring the biodiversity of disturbed soils: a case study on Mediterranean oribatid (Acari: Oribatida) assemblages. In: *Biodiversity and conservation in Europe*. Springer; 2006. p. 35–43.
46. Strong WL. Assessing species abundance unevenness within and between plant communities. *Community Ecol*. 2002; 3(2):237–246.
47. Pielou EC. The measurement of diversity in different types of biological collections. *J Theor Biol*. 1966; 13:131–144.
48. Lladser ME, Gouet R, Reeder J. Extrapolation of urn models via poissonization: accurate measurements of the microbial unknown. *PLoS One*. 2011; 6(6):e21105.
49. Child D. *The essentials of factor analysis*. 2nd ed. Cassell Educational; 1990.
50. Thompson B. *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association; 2004.
51. Williams B, Onsmann A, Brown T. *Exploratory factor analysis: A five-step guide for novices*. *Aust J Paramed*. 2010; 8(3).
52. Yong A, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutor Quant Methods Psychol*. 2013; 9(2):79–94.
53. Tabachnick BG, Fidell LS. Chapter 13 principal components and factor analysis. *Using Multivar Stat Lond Pearson*. 2012;
54. Costello A, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assess Res Eval*. 2005; 10(7).
55. DeCarlo L. On the meaning and use of kurtosis. *Psychol Methods*. 1997; 2(3):292.
56. Cain M, Zhang Z, Yuan K-H. Univariate and multivariate skewness and kurtosis for measuring non-normality: Prevalence, influence and estimation. *Behav Res Methods*. 2016; 49(5):1716–1735.
57. Blanca M, Arnau J, López-Montiel D, Bono R, Bendayan R. Skewness and Kurtosis in Real Data Samples Skewness and kurtosis in real data samples. *Methodol Eur J Res Methods Behav Soc Sci*. 2013; 9(2):78–84.

58. Micceri T. The unicorn, the normal curve, and other improbable creatures. *The unicorn, the normal curve, and other improbable creatures*. *Psychol Bull*. 1989; 105(1):156.
59. Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 1.8.12.—Google Search [Internet]. [cited 2019 Mar 24]. Available from: [https://www.google.com/search?q=Revelle%2C+W.++\(2018\)+psych%3A+Procedures+for+Personality+and+Psychological+Research%2C+Northwestern+University%2C+Evanston%2C+Illinois%2C+USA%2C+https%3A%2F%2FCRAN.R-project.org%2Fpackage%3Dpsych+Version+%3D+1.8.12.&dq=Revelle%2C+W.++\(2018\)+psych%3A+Procedures+for+Personality+and+Psychological+Research%2C+Northwestern+University%2C+Evanston%2C+Illinois%2C+USA%2C+https%3A%2F%2FCRAN.R-project.org%2Fpackage%3Dpsych+Version+%3D+1.8.12.&aq=chrome..69i57.400j0j4&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=Revelle%2C+W.++(2018)+psych%3A+Procedures+for+Personality+and+Psychological+Research%2C+Northwestern+University%2C+Evanston%2C+Illinois%2C+USA%2C+https%3A%2F%2FCRAN.R-project.org%2Fpackage%3Dpsych+Version+%3D+1.8.12.&dq=Revelle%2C+W.++(2018)+psych%3A+Procedures+for+Personality+and+Psychological+Research%2C+Northwestern+University%2C+Evanston%2C+Illinois%2C+USA%2C+https%3A%2F%2FCRAN.R-project.org%2Fpackage%3Dpsych+Version+%3D+1.8.12.&aq=chrome..69i57.400j0j4&sourceid=chrome&ie=UTF-8)
60. Abdi H. Factor rotations in factor analyses. *Encycl Res Methods Soc Sci Sage* Thousand Oaks CA. 2003;792–795.
61. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958 Sep 1; 23(3):187–200.
62. Mason NW, Mouillot D, Lee WG, Wilson JB. Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos*. 2005; 111(1):112–118.
63. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science*. 2016 Apr 29; 352(6285):560–4. <https://doi.org/10.1126/science.aad3503> PMID: 27126039
64. Hagerty SL, Ellingson JM, Hutchison KE. Biological Systems Are a Common Link Between Alcohol Use Disorder and Co-Occurring Psychiatric and Medical Conditions. *Alcohol Clin Exp Res*. 2018 Feb; 42(2):248–51. <https://doi.org/10.1111/acer.13570> PMID: 29197116
65. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin Psychol Rev*. 1988 Jan 1; 8(1):77–100
66. Mouillot D, Mason WHN, Dumay O, Wilson JB. Functional regularity: a neglected aspect of functional diversity. *Oecologia*. 2005 Jan; 142(3):353–9. <https://doi.org/10.1007/s00442-004-1744-7> PMID: 15655690
67. Mouillot D, Wilson JB. Can we tell how a community was constructed? A comparison of five evenness indices for their ability to identify theoretical models of community construction. *Theor Popul Biol*. 2002 Mar; 61(2):141–51. <https://doi.org/10.1006/tpbi.2001.1565> PMID: 11969386
68. Purvis A, Hector A. Getting the measure of biodiversity. *Nature*. 2000 May; 405(6783):212. <https://doi.org/10.1038/35012221> PMID: 10821281
69. Leclercq S, Cani PD, Neyrinck AM, Stärkel P, Jamar F, Mikolajczak M, et al. Role of intestinal permeability and inflammation in the biological and behavioral control of alcohol-dependent subjects. *Brain Behav Immun*. 2012 Aug 1; 26(6):911–8. <https://doi.org/10.1016/j.bbi.2012.04.001> PMID: 22521198
70. Leclercq S, Matamoros S, Cani PD, Neyrinck AM, Jamar F, Stärkel P, et al. Intestinal permeability, gut-bacterial dysbiosis, and behavioral markers of alcohol-dependence severity. *Proc Natl Acad Sci*. 2014 Oct 21; 111(42):E4485–93. <https://doi.org/10.1073/pnas.1415174111> PMID: 25288760