12-2019

Campylobacter infection of young children in a Latin American middle-income country and its impact on the gastrointestinal environment

Jessica Tweedie
University of Tennessee, jtweedie@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Jessica Tweedie entitled "Campylobacter infection of young children in a Latin American middle-income country and its impact on the gastrointestinal environment." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Microbiology.

Jeremiah Johnson, Major Professor

We have read this thesis and recommend its acceptance:

Heidi Goodrich-Blair, Sarah Lebeis

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Campylobacter infection of young children in a Latin American middle-income country and its impact on the gastrointestinal environment

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Jessica Logan Tweedie
December 2019

DEDICATION

To my babies

Ozy, Clark, and Okie

my mother

Paula Denise Couch

my best friend

Taylor Webster Gunes

ABSTRACT

Campylobacter infections are the leading bacterial cause of diarrhea worldwide with potentially profound impacts on pediatric patients in the developing world. It is still unclear how *Campylobacter* impacts human hosts, though it is becoming increasingly evident that *Campylobacter*s infection's impact is a multifactorial process that depends on the host immune response, the gastrointestinal microbiota, various bacterial derived metabolites, and possibly the nutritional status of the host. Since these factors likely differ between human-host and bacteria in the developed and developing worlds, it is important that studies comprehensively define these attributes in well characterized clinical cohorts in both settings. In this study, we analyzed the microbiota, the metabolome, and the micronutrient profile of fecal samples from *Campylobacter*-infected pediatric subjects in Colombia during a case-controlled study of acute diarrheal disease. Here, we report that *Campylobacter*-infected children exhibited significant changes to the gastrointestinal environment when compared to uninfected cohorts, including shifts in *Proteobacteria* abundance and concentrations of gastrointestinal metabolites, and decreases in almost all essential micronutrients. These observations have led to several hypotheses on how *Campylobacter* infection may affect host biology, and ultimately negatively impact clinical outcome in young children throughout the developing world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE – LITERATURE REVIEW

Statement of Contribution

The research presented in this thesis is a represents a significant portion of a collaborative study. The fecal samples and patient data were provided by Dr. Oscar Gomez-Duarte of University at Buffalo (Buffalo, New York, USA) and AE Farfán-García of Universidad de Santander (Bucaramanga, Colombia). The UHPLC-MS/MS analysis of fecal metabolites and inductively-coupled plasma mass spectrometry (ICP-MS) analysis of fecal micronutrients were both outsourced to other laboratories. All other data collection, processing, and analysis were performed by the author.

Campylobacter

Campylobacter is a Gram-negative, spiral or 'S'-shaped bacterium, and a member of the epsilon group of the phylum Proteobacteria [1]. *Campylobacter* is a motile organism propelled in a corkscrew fashion by either a single polar flagellum or bipolar flagella [1]. While *Campylobacters* are the most common cause of bacterial derived-gastroenteritis in the world [2], there is significantly less knowledge of the pathogenic mechanisms of its primary human pathogens *C. jejuni* and *C. coli* compared to other enteric pathogens such as *Escherichia coli, Salmonella,* and *Vibrio* species [2]. This is attributed to the overall microaerophilic nature of the *Campylobacter* genus, which along with the inability to utilize mice as a pertinent animal model for human disease, increases the difficulty of studying *Campylobacter* in a laboratory setting [3]. There are currently 17 species that make up the genus *Campylobacter* with *C. jejuni*

and *C. coli* most commonly associated with human disease [1]. *Campylobacter* that cause disease are found in food animals such as poultry, cattle, pigs, and sheep, as well as in pets, including cats and dogs [1, 4]. The most common route of transmission in developed countries is through the consumption of undercooked meats, especially chicken [4, 5]. The sources of infection in low- and middle –income countries (LMIC) are less understood, though it is thought to be primarily contaminated drinking water [6].

Onset of *Campylobacter*-induced disease usually occurs 2-5 days after ingestion, and symptoms include diarrhea (frequently bloody), abdominal pain, fever, headache, nausea, and/or vomiting [5]. The disease is considered self-limiting with symptoms lasting 3-6 days. *Campylobacter*-infection is associated with post-infection complications. One association is an intestinal disorder causing stomach pain, gas, diarrhea, and constipation known as irritable bowel syndrome [1, 7]. Reactive arthritis, a painful inflammation of the joints which can last for several months, is another associated possible complication [1]. *Campylobacter*-infection is recognized as the trigger for up to 50% of Guillain-Barré cases [8], which is a rare neurological autoimmune disorder causing the immune system to attack the peripheral nervous system leading to weakness, numbness, and tingling, and can eventually cause paralysis [9].

Campylobacter jejuni

Campylobacter jejuni is considered to be responsible for ~80-90% [10] of the 95 million annual cases of *Campylobacter*-induced gastroenteritis worldwide [2].This is exemplified by a surveillance study in the United Kingdom found that *C. jejuni*

accounted for 93% of the *Campylobacter* infections out of 3,764 cases [11]. *C. jejuni*'s infectious dose of fewer than 500 organisms [9] is generally lower than those of other enteric bacteria including *Salmonella typhmurium* or pathogenic *E. coli* strains [12].

Antibiotic resistance in *C. jejuni* has increased over the last couple of decades [13], leading the Center for Disease Control (CDC) to classify antibiotic resistant *C. jejuni* as a "Serious Threat" to public health [14]. The U.S. Food and Drug Administration (FDA) withdrew the final remaining fluoroquinolone approval for poultry flocks in 2005, but the National Antimicrobial Resistance Monitoring System (NARMS) surveillance has observed an increase of ciprofloxacin (a fluoroquinolone) resistance in *C. jejuni* clinical isolates from 19.6% in 2006 to 25.3% in 2015 [15]. Since the rise of fluoroquinolone resistance, when treatment against *C. jejuni* is necessary, macrolides like azithromycin are now the drugs of choice since only 2.7% of *C. jejuni* clinical isolates were resistant to azithromycin in 2015 [15]. While this demonstrates the issue of rising antibiotic-resistance in *C. jejuni* in the developed world, it is also an increasing issue in the undeveloped world where there are less regulations on antibiotic use.

Campylobacter coli

Campylobacter coli is the second most common *Campylobacter* species associated with human infection [1]. The UK surveillance study previously discussed found that *C. coli* was responsible for 7% of cases and it is considered to be responsible for ~10% of all *Campylobacter* human infections worldwide [11]. *C. coli* also shares very similar clinical manifestation of human infection as *C. jejuni* [11]. Although *C. coli* is responsible for fewer human infections, NARMS reported the species has a higher rate

of antibiotic resistance [15]. The 2015 Human Isolates Surveillance Report states that 39.8% of *C. coli* isolates were resistant to ciprofloxacin and 12.7% were resistant to azithromycin [15]. Although *C. coli* is responsible for fewer clinical cases, the increase of antibiotic resistance may suggest an association with lower rates of successful antibiotic treatment than observed in *C. jejuni.*

Other *Campylobacter* Species

While significantly less frequent than *C. coli* or *C. jejuni*, other species of *Campylobacter* have been reported to cause human disease [16]. NARMS reported testing 46 (4% of study population) non-*C. jejuni/C. coli Campylobacter* species in 2015 [15]. Studies suggesting these "other species of *Campylobacter"* are increasing in burden and importance in developing countries [8, 17]. For example, *Campylobacter upsaliensis* was found to be present in 23% of *Campylobacter* stool isolates in South Africa using filtration culture [18, 19]. In contrast, Irish and Canadian studies using filtration culture and/or PCR detected *C. upsaliensis* in only 0.7 - 2.1% of stool specimens [20-22]. Although *Campylobacter* is more often associated with livestock, there are reports of pet-to-human transmission of *C. upsaliensis* from domestic dogs and cats, a reservoir of *C. upsaliensis* [23-25].

There are reports of human campylobacteriosis caused by other *Campylobacter* species, including *C. concisus* [26]*, C. sputorum* [27], *C. ureolyticus* [28], and *C. hyointestinalis* [29]. *C. concisus,* a colonizer of the human oral cavity, is suspected in causing a primary barrier function defect leading to the onset of inflammatory bowel disease [30, 31], and is associated with Crohn's Disease in children [32].

In light of new diagnostic testing techniques, *Campylobacter* epidemiological studies are now able to detect *Campylobacter* incidences more readily with a wider net for species detection [33]. Previous methods were highly selective for *C. jejuni/C. coli* in comparison to other *Campylobacters*, suggesting these other *Campylobacters* may have a higher incidence rate than previously considered [17].

Diagnostic Testing of *Campylobacter*

Current Clinical Diagnostics in Developed World

Laboratory diagnostics to identify *Campylobacter* infections are beginning to expand to using culture-independent technologies, but culture-dependent methodologies still dominate the field [1, 8, 34]. The clinical field primarily uses techniques to target *C. jejuni/C. coli* species. The routine technique for culturing *C. jejuni* or *C. coli* from feces utilizes selective media infused with multiple antibiotics and sheep or horse blood [34]. Incubation at the human body temperature of 37°C or the chicken cecum at 42°C in combination with decreased oxygen levels, and increased carbon dioxide and nitrogen levels promote the growth of *C. jejuni* [34].

A major drawback to the culture-based methods is that *Campylobacter* detection is sensitive to the fecal transport media and viability is affected between bowel movement to actual fecal culturing [1, 35, 36]. A recent study provided a quantitative estimate of *C. jejuni/C. coli* detection in fecal-derived cultures and viability losses of *Campylobacter*-positive specimens in the commonly used Cary-Blair transport medium [35]. The study results suggest poor stability of *Campylobacter* in Cary-Blair transport media [36]. They observed a 94% loss of culturable *C. jejuni* when fecal samples spiked

with *C. jejuni* and diluted in Cary-Blair medium were stored at 2-8°C for 24 hours, and *C. jejuni* was unable to be cultured at 48 or 96 hours [35]. This demonstrates how difficult accurately culturing *Campylobacter* in a clinical setting can be.

In an attempt to more accurately detect *Campylobacter*, culture-independent methods have been developed over the last two decades [37]. The use of immune and nucleic acid-based methodologies are becoming increasingly routine for the fast detection and identification of foodborne pathogens [38]. Enzyme immunoassay (EIA)-based stool antigen tests are commercially available assays for *Campylobacter* detection [39]. The EIA-based methodologies exploit antibody affinity for specific target antigens found on the surface of *Campylobacter* cells [38]. The EIA is desirable due to the relative simplicity of the protocol and speed of results [38], but there are conflicting reports on its sensitivity although it is highly specific [39]. The nucleic acid-based technologies, such as Polymerase-Chain Reaction (PCR), utilize the highly specific and distinct nature of DNA or RNA sequences [38]. PCR primers for detecting *Campylobacter* have been created to amplify specific regions of DNA allowing specific detection of *Campylobacter* [33]. While these methodologies have their advantages, there are still drawbacks.

Both EIA and PCR-based methodologies were primarily designed for *C. jejuni* or *C. jejuni* and *C. coli* detection, thus excluding detection of all other *Campylobacter* species [17, 33]. There have been conflicting reports on how sensitive and specific enzyme immunoassay (EIA)-based stool antigen tests and nucleic acid-based tests are in comparison to culture in children and adults [8, 37, 39]. Clinical labs have reported a

high false positivity rate among the culture independent techniques, which has led to the current gold standard for detection to be either culture-positive or both PCR and EIA positive [8, 37]. This addition of culture-independent techniques is especially crucial since a study that found culture-based methods only had a 37% sensitivity rate in over 500 tested samples [37].

Current Clinical Diagnostics for Studies in Developing Countries

A recently published *Campylobacter* study performed by the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development study (MAL-ED), suggested that *C. jejuni* and *C. coli* may not be the primary *Campylobacter* pathogens as previously considered in low-income countries [17]. This MAL-ED case-control study revealed that of the 439 stool samples tested 76.4% of the 216 positive *Campylobacter* detections were non-*C. jejuni/C. coli* species [17]. Children experiencing dysentery were more likely to be infected with *C. jejuni/C. coli* species than "other *Campylobacters"* [17]. The other *Campylobacters* were more prevalent across all age groups than *C. jejuni/C. coli* with older children bearing a higher burden of other *Campylobacter,* and other *Campylobacter* are more likely to detected in an asymptomatic child [17]. The other *Campylobacters* were also statistically just as likely as *C. jejuni/C. coli* species to cause severe diarrheal episodes [17].

Researchers from the MAL-ED study also determined the current "standard" methods for the detection of *Campylobacter* in feces in developing countries [33]. In order to make this determination, the group compared commonly utilized culture,

enzyme immunoassay, and PCR techniques for *Campylobacter* detection sensitivity and accuracy [33]. They ultimately recommend the use of quantitative-PCR (qPCR) “as the preferred diagnostic modality for detection of *Campylobacter* infection for epidemiologic studies in the developing world” [33]. The advantage of qPCR over PCR is that the fluorescent labeling allows for the quantification of the amplified DNA region [33]. This recommendation is based on results suggesting that qPCR can both determine the presence and burden of infection, and can give species specific information [33]. Although these results are dependent on primer choice [33]. For example, the use of 16S rRNA gene will detect any *Campylobacter* species and potentially identify the specific species but does require DNA sequencing and analysis [33]. There are other *Campylobacter*-specific primers, but most are designed to detect *C. jejuni/C. coli* strains [38]. The genomic plasticity contributes to the difficulty of designing a single accurate and specific *Campylobacter* genus primer [38].

Overview of *Campylobacter* Metabolism

Campylobacter’s metabolism is still not fully understood but is known for some of its distinctive metabolic traits [40]. Most *Campylobacter* are thermophilic bacteria that grow preferentially between 42°C (chicken body temperature) and 37°C (human body temperature), does not proliferate below 30°C, but is able to survive and perform basic metabolic functions in temperatures as low as 4°C [41]. This ability to survive diverse temperatures, pH, oxygen concentrations, and osmotic environments is what allows *Campylobacter*s to colonize a diverse range of animal hosts [1, 9, 10, 40, 42].

The most notable metabolic property of *C. jejuni* is its non-glycolytic nature due to its inability to catabolize glucose [10, 43]. Some strains of *C. coli* have been found to perform glycolysis [40]. A few strains of *C. jejuni* have the genomic island that encodes the enzymes required for fucose utilization and the chemotaxis receptors for it [44] but is not widely possessed by most strains [40, 44]. The citric acid (TCA) cycle and gluconeogenesis are required for biosynthesis of precursor substrates for lipo-oligosaccharides (LOS) and capsular polysaccharides [40]. Catabolism of amino acids, keto acids, organic acids, and short chain fatty acids (SCFA) is the primary source of energy production for *Campylobacter*s growth [40]. *C. jejuni* prefers amino acids, specifically serine, aspartate, glutamate, and proline for its carbon source, but some strains can utilize asparagine and glutamine as well [40].

While *C. jejuni* is microaerophilic and does not use oxygen as a terminal electron acceptor, it still runs the TCA cycle in the oxidative direction by utilizing the bifunctional enzyme FrdABC complex [10]. Under strict microaerophilic conditions, FrdABC serves as succinate dehydrogenase and as fumarate reductase to convert fumarate to succinate [10]. *C. jejuni* also commonly utilizes formate, nitrate, and nitrite as alternative electron acceptors [10]. It has been suggested that *Campylobacters* have a mechanism for recognizing differences in the spatial distribution of specific SCFAs and other metabolites to discriminate between different regions of the chicken (and potentially other hosts) intestinal tract and exploits this to coordinate expression of genes necessary for optimal colonization [45]. Research has shown that lactate (abundant in the chick's upper intestine) is associated with repression of catabolic gene expression,

while acetate and butyrate (abundant in the chick's lower intestine) is associated with an increase in catabolic gene expression [45]. This suggests that there may be "cues" that *C. jejuni* uses to colonize the human gastrointestinal tracts (GI).

Phase Variation

While *Campylobacters* have relatively small genomes (~1.7 Mb), they display extensive genetic variations [5, 46]. *C. jejuni*'s genome contains a high prevalence of hypervariable sequences consisting of homopolymeric tracts (HT) [5, 46]. These HTs allow for *C. jejuni* to rapidly mutate and increase evolutionary adaptation by a process known as Phase Variation (PV) [5]. Phase variation is a form of ON/OFF switches in gene expression mediated by changes in the number of repeats in repetitive DNA sequences [5]. These repetitive DNA sequences, known as simple sequence repeats (SSRs), are repeats of the same DNA sequence directly adjacent to one another [47]. PV is the reversible change of these SSRs, while being characteristically stochastic, high-frequency, reversible, and causes heritable phenotypic changes [5]. There are two types of PV: 1) Translational PV occurs when the SSR is located in the open reading frame leading to frameshift mutations, which routinely cause truncation of the protein leading to the production of a non-functional protein product [46, 47]. 2) Transcriptional PV alters the activity of a regulatory element (i.e. promoter) due to changes in repeat number of an intergenic SSR [46, 47].

A recent study used an 'omic based approach to analyze the distribution and conservation of the "phasome" (putative and known SSR-driven PV genes) in *Campylobacter* [47]. The study utilized the program Phasome*It*, a python-based tool for

locating, classifying, and grouping phase variable SSRs in genomic collections [47], and 77 whole genome sequences from 14 *Campylobacter* species to identify new trends in *Campylobacter* phasome [47]. Each strain had between 5-81 PV genes, and *C. jejuni* (n=35) and *C. coli* (n=10) had means of 20 and 15 PV genes, respectively [47]. The majority of the identified homologous PV genes across the genus have putative roles in modification of the flagellin, or encode enzymes with roles in modification of the LOS, capsule polysaccharide, or flagellum [47]. The flagella, LOS, and capsule have been associated with roles in *Campylobacter* virulence [5]. Virulence is the organism's ability to infect a host and cause disease [48]. Virulence factors of an organism are the gene products that help to i) invade the host, ii) cause disease, and iii) evade host defenses [48]. The adaptative process of phase variation is able to genetically regulate virulence factors and is advantageous for *Campylobacter*'s colonization of diverse hosts and immune evasion [5, 46, 47].

A study on *C. jejuni* transcriptional and genetic adaption was able to identify influential variants selected for during adaptation of a new host environment utilizing whole genome sequencing and transcriptomics [46]. The study infected 14 human volunteers with *C. jejuni* strain CG-8421, tracking transcriptional and genetic changes between inoculum and human infection isolates [46]. The comparison of gene expression was accomplished using RNA transcriptomes from three of the human infected feces, the inoculum strain cultured on blood agar, Mueller-Hinton (MH) Agar, and MH broth, and previously published transcriptomes of *C. jejuni* strain 81-176 colonizing the chicken cecum [46]. The comparison identified 39 core genes that are

differentially regulated across host and colonization behavior. These genes are involved in iron acquisition, inorganic phosphate uptake, and protection from peroxide stress [46]. Interestingly, *chuA, chuD* (genes involved in hemin uptake), and *cfrA* (gene involved in uptake of siderophores) were all upregulated >100 fold in human infection verses chicken colonization [46]. *ChuA* is potentially a critical virulence factor for *C. jejuni* as it is not required for chick colonization [49] but was found to have >1000-fold expression increase in human infection [46].

The study identified variations in the genomic sequences of *Campylobacter* populations by calling variants, a single nucleotide polymorphism (SNP), phase variation (PV), or multiple nucleotide polymorphism (MVP) if it occurred in >25% of the genomic population [46]. There were 48 unique genomic variants were identified, and 19 of the 28 PV poly G/C tracts in CG8421 genome varied in at least one of the human infection populations [46]. Genetic variants of three genes with protein products with roles in metabolism, iron/heme acquisition, and respiration were selected in all population isolates [46].

Five of the volunteers experienced recrudescent infections allowing for the comparison of acute and persistent infection variants [46]. Since the number of variants did not increase over time during primary infection, but twice as many gene variants were observed in the recrudescent population suggests there is an immediate and consistent selection pressure during the primary infection [46]. Seven out of the eight (identified by this study) flagella modification genes were associated with either acute or persistent infection populations suggesting that flagellar modification plays a large role

in *Campylobacter* virulence [46]. This study's investigation of *C. jejuni*'s transcriptional and genetic adaption in the human infection environment gave new insights to the *Campylobacter* pathogenesis research field, including the significance of genetic adaption and population selection during infection [46].

Human Microbiome and Colonocyte Metabolism

Understanding the environment being invaded when investigating *Campylobacter*'s pathogenesis is critical for identifying links between *Campylobacter*'s metabolism, virulence, and genetic adaption and the host environment. It is essential for the mammal gut immune system to maintain the integrity of epithelial lining and other host tissues, while detecting and destroying potential pathogens without needlessly reacting to food, normal commensal microorganisms, and any other innocuous luminal content [50]. The immune regulation/response of the gastrointestinal tract is required to be highly sophisticated. The rapidly increasing research on the human microbiota has revealed the immune system may play a role in balancing microbial communities inhabiting human bodies [51].

The largest barrier delaying the understanding of specific immune functions or cell types required for maintaining microbiota homeostasis in the colon is the difficulty of defining what a "balanced" microbiota should be [51]. This is a challenging task because these microbial communities are i) highly diverse, ii) differ between individuals, and iii) shift with diet changes [51, 52]. Research suggests the immune system shapes the microbiota to be beneficial in order to maintain homeostasis [53]. A beneficial microbiota would aid in the digestion of nutrients that the host cannot breakdown such as complex

carbohydrates [54], promote immune system development [55], and provide niche protection against enteric pathogens [56].

Host mechanism research indicates colonic epithelial cells (colonocytes) have a central role in shaping the microbiota [53, 57]. The colonic stem cells, located at the base of the intestinal crypts, generate transit-amplifying cells by asymmetrical division [58]. These transit-amplifying cells will eventually terminally differentiate into various epithelial cell types including, colonocytes, enteroendocrine cells, or goblet cells [58]. Activation of epithelial differentiation requires a nuclear receptor known as peroxisome proliferator activated receptor - γ (PPAR-γ) [51]. PPAR-γ is primarily synthesized in the differentiated cells of the epithelium of humans and rodents [59, 60]. PPAR-γ also activates fatty acid metabolism resulting in mitochondrial β-oxidation of long-chain and SCFAs (Fig. 1a) [54, 61, 62]. The mature colonocytes primarily use mitochondrial β-oxidation, which consumes oxygen via oxidative phosphorylation [51]. Increasing colonocyte oxygen consumption leads to a decrease in the amount of oxygen freely diffusing from the epithelial surface into the lumen [51]. This decrease of oxygen partial pressure to less than 1% or 7.6mmHg results in a condition termed epithelial hypoxia [63]. Epithelial hypoxia is critical for maintaining anaerobiosis in the intestinal lumen [63].

Anaerobiosis ensures the dominance of obligate anaerobic bacteria in the colonic microbiota [64]. The adult human colonic obligate anaerobic community is predominately comprised of members of the phyla Firmicutes and Bacteroidetes, which include members that benefit the host with their broad spectrum of enzymes for

complex carbohydrate hydrolysis [51]. In turn these obligate anaerobes assist with maintaining the epithelial hypoxia since many members of the class Clostridia (Firmicutes) encode enzymes to convert dietary fiber into butyrate [65], and butyrate activates PPAR-γ signaling in the epithelial cells (Fig. 1a) [54, 62, 66].

Since the colonic epithelium and its metabolism may contribute to the immune functions that maintain the shaping of the microbiota to a beneficial homeostasis [61], an underlying defect in the immune functions that maintain the homeostasis may cause a microbiota “imbalance” [53, 67]. This imbalance, known as dysbiosis, is commonly associated with an increased abundance of facultative anaerobic bacteria [59, 68]. This dysbiosis is detrimental to the host since facultative anaerobes, including members of the phylum Proteobacteria, cannot produce fermentation products from fiber and will consume other microbiota-derived fermentation products [69]. Dysbiosis is associated antibiotic therapy [70], consuming a high-fat western style diet [71, 72], colorectal cancer [73], inflammatory bowel disease [74], irritable bowel syndrome [75], and necrotizing enterocolitis [76].

The “oxygen hypothesis” suggests a disruption in anaerobiosis, mediated by the ability of facultative anaerobes to respire oxygen, would induce a shift in the microbial community from obligate to facultative dominance (Fig. 1b) [74]. Research on the impact of antibiotic therapies on the microbiome has provided valuable information on mechanisms that disrupt intestinal microbiota homeostasis [51]. Antibiotic treatments deplete fermentation product producers leading to decreased availability of butyrate, propionate, acetate, and other fermentation products [51]. In murine models, SCFAs are

absorbed into the colon where they will bind to G protein-coupled receptors, and inhibit intestinal inflammation by maintaining the population of regulatory T-cells in the mucosa [55, 77, 78]. The depletion of butyrate reduces PPAR-γ signaling in human and murine epithelial cells [55, 79]. Silencing of PPAR-γ signaling leads to the reduction of the number of regulatory T-cells in the murine mucosa [80]. Therefore, antibiotic treatment increases the inflammatory state of the colonic mucosa [61], and the inflammatory signals shifts the metabolism of terminally differentiated surface colonocytes toward anaerobic glycolysis (Fig. 1b) [81, 82].

Anaerobic glycolysis metabolism is characterized by its high consumption of glucose, abundant production and release of lactate, and low oxygen consumption [51]. The colonocyte metabolic shift to anaerobic glycolysis leads to the loss of epithelial hypoxia due to the lack of high oxygen consumption causing an increased concentration of oxygen to diffuse into the lumen [51, 59]. The reduction of PPAR-γ signaling by the depleted concentration of butyrate leads to an increase of inducible nitric oxide synthase (iNOS) synthesis causing nitric oxide productio (Fig. 1b) [61]. The nitric oxide is converted to nitrate in the intestinal lumen [83], and contributes to the expansion of Proteobacteria when used as a terminal electron acceptor for anaerobic respiration [68, 83].

Excessive epithelial repair drives the accumulation of dividing transit-amplifying cells that cause crypt elongation, known as colonic crypt hyperplasia, a common feature of ulcerative colitis [84]. Colonic crypt hyperplasia leads to the thinning of the mucus layer by reducing the number of terminally differentiated epithelial goblet cells, which is

the primary producer of the mucus layer [58]. The reduced number of terminally differentiated epithelial cells leads to reduced production of PPAR-γ and activation of mitochondrial β-oxidation of fatty acids, which decreases epithelial hypoxia [84]. Additionally, low PPAR-γ synthesis increases intestinal permeability ("leaky gut") in the rat colitis model [85]. This is caused by PPAR-γ regulation of tight junction molecules in the epithelial cells [86].

An important function of the colonic surface is utilizing niche protection against facultative anaerobic enteric pathogens by limiting the luminal availability of host-derived respiratory electron acceptors and carbon sources [51, 56]. Some species are able to overcome niche protection by the use of their virulence factors [87]. *Salmonella enterica* triggers severe acute intestinal inflammation with its virulence factors, which induces neutrophil migration into the intestinal lumen causing the depletion of both *Clostrida spp.* and their SCFA fermentation products [88, 89]. Additionally, the respiratory burst of phagocytes generates electron acceptors such as nitrate and tetrathionate [90]. *S. enterica* outcompetes obligates by consuming microbiota-derived fermentation products and host-released lactate as carbon sources [88]. Although *Campylobacter*'s invasion/colonization mechanism in human's is not well understood, it is known for inducing severe acute intestinal inflammation [5]. This process could potentially mirror *S. enterica's* method of depleting *Clostrida spp.* and their SCFA fermentation products, which are compatible with *Campylobacter's* catabolism [40].

Another enteric pathogen, *Citrobacter rodentium,* is a mouse pathogen frequently studied as a model for virulence mechanisms used by human attaching-and-

effacing pathogens such as enteropathogenic *E. coli* or enterohemorrhagic *E. coli* [91]. *C. rodentium* uses virulence factors to attach to the colonic surface and cause epithelial injury to induce colonic crypt hyperplasia by excessive epithelial repair [91]. *C. rodentium* will utilize the increased oxygen concentration and microbiota-derived products to compete with the commensal gut microbiota [91]. *Campylobacter* is known to attach to colonic surfaces and is associated with epithelial damage and villous blunting [7], which could indicate pathogenic mechanisms similar to *C. rodentium.* These examples demonstrate the importance of host-microbe and microbiota-microbe interactions when examining *Campylobacter* human pathogenesis.

CHAPTER TWO – INTRODUCTION

Malnutrition is a complex condition responsible for approximately 45% of deaths in children under five years old [92]. Chronic pediatric malnutrition has multiple causes, including food insecurity, recurrent enteric infections, gut dysbiosis, and impaired immunity [93]. The interrelated nature of these conditions makes it difficult to distinguish impaired immunity and increased susceptibility from persistent enteric infections and gastrointestinal dysbiosis [94]. Chronic pediatric malnutrition can then subsequently lead to growth faltering, cognitive impairment, and/or other long-term pathologies [95].

In the example of recurrent enteric infections, continuous exposure to gastrointestinal pathogens can lead to prolonged intestinal inflammation, which can further lead to the development of Environmental Enteric Dysfunction (EED) [93]. This disorder is characterized by villous blunting and decreased intestinal integrity, which drives reduced nutrient absorption and is associated with stunting (low height-for-age) [93].

Campylobacter was overlooked as a serious public health issue for many years primarily because the difficulty of culturing and previous lack of PCR techniques for diagnosis in combination with the frequently asymptomatic nature of *Campylobacter* in the developing world [96]. In 2013, a controlled etiological study of moderate-to-severe diarrhea in children in Asia and Africa designed to comprehensively identify the pathogen-specific etiology and burden of disease associated with each pathogen the Global Enteric Multicenter Study (GEMS) was published and highlighted the prevalence

of *Campylobacter* infections in these environments [97]. GEMS occurred between December 2007 and March 2011, in seven different sites across Africa and Asia, and enrolled 9439 children with moderate-to-severe diarrhea and 13129 children as asymptomatic controls all between the ages of 0-59 months [97]. The cases and controls were all matched for age, sex, and village, and the detection of *Campylobacter* by culture using selective blood agar plates incubated at 42°C [97]. They found that *Campylobacter* was in the top five pathogens associated with burden of diarrhea in three out of the seven sites [97].

To identify the specific enteric pathogens that may cause EED, the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development study (MAL-ED) worked to identify which infections most often correlated with impaired physical growth, cognitive development, and immunity in pediatric populations in the developing world [98]. To accomplish this, the investigators monitored children until two years of age from eight geographically distinct sites known to have high rates of malnutrition to determine which exposures correlated with gastrointestinal disease and ultimately which may lead to decreased growth and development [99]. Not surprisingly, several common enteric pathogens significantly correlated with diarrhea in the first year of life, including (in order of prevalence) norovirus GII, rotavirus, *Campylobacter* spp., Astrovirus, and *Cryptosporidium* spp. Similarly, *Campylobacter spp.*, norovirus GII, rotavirus, Astrovirus, and *Shigella spp*. were associated with diarrhea in the second year of life [98]. While *Campylobacter* was the third pathogen most often associated with diarrhea in the first year of life, it

increased to the most often associated in year two. The predominance with which *Campylobacter* associated with diarrhea occurred across all sites [100] and another study reported asymptomatic *Campylobacter* infection can result in increased intestinal permeability and stunted growth [101].

To begin characterizing these events, groups have examined both the host and bacterial responses to infection using human clinical samples, including those obtained from human feeding studies [46], and have leveraged various animal models of disease [6, 102]. These studies have demonstrated that *Campylobacter*-induced disease is likely a multifactorial process dependent on the host immune response, the host microbiota, host nutritional status, and bacterial factors.

As host and bacterial characteristics likely differ between populations in industrialized nations and LMICs, it is going to be important to conduct multiple studies using well characterized clinical groups within each population, including specifically examining samples from pediatric subjects. To that end, we obtained pediatric fecal samples collected during a case-controlled study of acute diarrheal disease in Colombia, a middle-income country, and performed untargeted, omics-based analyses on the samples [103]. These methods allowed us to begin examining the impact *Campylobacter* infection has on the gastrointestinal environment of pediatric populations in middle-income countries.

CHAPTER THREE – MATERIALS AND METHODS

Subject Recruitment and Data Collection

Study Site

As previously described [103], the study was conducted in the Bucaramanga, Colombia metropolitan area. Bucaramanga is the capital city of the Santander department of Colombia. This city includes four municipalities, a population of 525,000, and 90% availability of basic utilities including those for water, electricity, gas, telephone, and garbage collection frequency of three times a week [103]. The coverage of pediatric hospital beds in this region is 0.48 per 1000 for children under 12 years of age [103]. The under-5 mortality rate in Colombia was 5.26 per 100,000 children in 2010, with acute diarrheal disease (ADD) being the third leading cause of morbidity among children less than five years of age in Santander, Colombia [103].

Subject Recruitment

Samples were collected during a study designed as a prospective, matched-for-age, case-control study to determine the etiology of ADD in children from two weeks to 59 months of age in the Bucaramanga metropolitan area, Colombia [103]. The ADD study was approved by Institutional Review Board, Vanderbilt University School of Medicine (IRB number 130327). From August 2013 to December 2014 subjects were recruited from emergency, inpatient, and outpatient facilities of major medical institutions including Unidad Intermedia Materno Infantil Santa Teresita (UIMIST), Centro de Salud el Rosario, Fundación Oftalmológica de Santander Carlos Ardila Lulle

(FOSCAL), Hospital Local del Norte, Clínica Materno Infantil San Luis, and Hospital San Juan de Dios de Floridablanca in the Bucaramanga metropolitan area, Colombia.

After informed written consent was obtained, an interview questionnaire was administered to the subject's parents or guardians and recorded in Spanish by trained clinical researchers at enrollment and two and six weeks after. Data collected included demographics, medical history, epidemiological factors, socioeconomic factors, nutrition, education, immunization history, water sources, and housing. In some cases, information was obtained about clinical manifestations specifically about characteristics of diarrhea, vomiting, abdominal pain, and dehydration. Access to the metadata and fecal samples was approved by and conducted under the University of Tennessee Institutional Review Board protocol #IRB-17-03795-XM.

Sample Collection

Stool samples were collected from children age two weeks to 59 months on either the day of enrollment or up to one-week post-enrollment from August to December 2014. Samples were collected in sterile disposable plastic containers and transported to the Laboratorio de Investigaciones Biomédicas y Biotecnológicas (LIBB) at the Universidad de Santander, Bucaramanga, Colombia within four hours. Once samples were examined for color, consistency, presence of blood and/or mucus, aliquots were made and divided into those to be immediately tested and those for storage at -20°C and -80°C [103].

Detection of Gastrointestinal Pathogens

The samples were tested for viral, parasitic, and bacterial infections at the Laboratorio de Investigaciones Biomédicas y Biotecnológicas (LIBB) at the Universidad de Santander, Bucaramanga, Colombia. Trained clinical technicians tested the fecal samples for Adenovirus, Astrovirus, Norovirus, Rotavirus, Sapovirus, *Entamoeba histolytica, Giardia lamblia, Cryptosporidium spp., Campylobacter spp., Salmonella spp., Shigella spp., Yersinia spp.*, enteroaggregative *Escherichia coli* (EAEC), diffuse-adhering *E. coli* (DAEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), enterotoxigenic *E. coli* (ETEC), *and* shigatoxigenic *E. coli* (STEC) [103].

Detection of Virial and Parasitic Pathogens

Described previously [103], RNA extracted from human stool samples were used for detection of Norovirus, astrovirus, and Sapovirus, while extracted DNA was used for Adenovirus detection. Detection of Norovirus GI and GII was conducted according to the US CDC protocol: GI/GII Norovirus Multiplex (TaqMan Life Technologies, Carlsbad, USA) real-time PCR (RT-PCR) Assay. Astrovirus and Sapovirus detection were conducted from isolated RNA according to the United States Center for Disease Control (CDC) protocol: Astrovirus/Sapovirus Duplex Real-Time (TaqMan) RT-QPCR Assay. RNA preparation from astrovirus and Sapovirus negative and positive human stool samples were negative and positive controls in qRT-PCR, and water was non-template control. Adenovirus detection from DNA isolated from stool samples was performed by qPCR. DNA samples obtained from adenovirus negative and positive human stool samples were used as negative and positive controls, respectively. Water was used as

a non-template control. Identification of Rotavirus was done by the Rotavirus Stool Antigen Detection ELISA assay according to manufacturer's instructions (Diagnostic Automation, Calabasas, USA).

All stool samples were processed for detection of *Giardia lamblia, Entamoeba histolytica, Entamoeba dispar,* and *Cryptosporidium* antigens by direct ELISA (Diagnostic Automation, Calabasas, USA) following manufacturer's instructions.

Clinical Detection and Identification of *Escherichia coli*

E.coli screening and identification protocol is previously described [103]. Stool samples were plated onto MacConkey (Merck, Darmstadt, Germany) agar. Isolates that were positive for lactose fermentation were used to inoculate Eosin Methylene Blue (EMB) agar (Merck, Darmstadt, Germany). If isolates had metallic shine on EMB agar plates, they would then inoculate SIM medium (Merck, Darmstadt, Germany). Isolates positive for motility and indole production, but negative for sulfide production would be confirmed as *E. coli* by *uidA* DNA amplification by PCR. A random subset was selected for API20E testing for further confirmation. Confirmed *E. coli* isolates were stored at -80°C in Luria broth (LB) (Scharlau, Sentmenat, Spain) supplemented with 15% glycerol (Merck, Darmstadt, Germany).

The bacterial stocks of the isolates were inoculated on LB plate (Scharlau, Sentmenat, Spain) and cultured at 37°C overnight. Harvested cultures were suspended in one-milliliter sterile water, vortexed for 10 seconds, and centrifuged at maximum speed for two minutes. The supernatants that contained bacterial DNA were incubated for five minutes at 95°C on a thermocycler and used as DNA template in multiplex

polymerase-chain reaction (mPCR). The mPCR was designed to detect six well-established pathogenic *E. coli* pathotypes including diffusely adherent *E. coli* (DAEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), and Shiga-toxin producing *E. coli* (STEC). The mPCR was divided into four reaction mixes with primers targeting specific virulence genes of each pathotype: reaction mix 1 targets lngA, bfpA, and stx2 genes; reaction mix 2 targets eae, daaE, ipaD, and aaiC genes; reaction mix 3 targets ipaH, escN, aggR, and estA genes; reaction mix 4 targets elt and stx1 genes. The strains used as controls in the mPCR included the following: negative control DH5α E. coli, negative control TOP10F' E. coli, C1845 DAEC, JM221 EAEC, EC-12 EIEC, E2348/69 EPEC, E9034A ETEC, and 2060-004 EHEC [104]. In each 25 μl mPCR reaction, we add 21 μl Platinum Blue PCR SuperMix (Life Technologies, Carlsbad, USA), 2 μl DNA template, and 2 μl primer mixes specifically indicated above. The final concentration of each primer in mPCR reaction was 0.4 μM. The thermocycling was set as 94°C for 5 minutes, 40 cycles of denaturing at 94°C for 30 seconds, annealing at 58°C for 30 seconds, and extension at 72°C for 60 seconds, and final extension at 72°C for 5 minutes. The annealing temperature for reaction 4 was set at 60°C to increase PCR specificity. The PCR products were loaded onto a 2.0% agarose gel containing SYBR safe DNA gel stain (Life Technologies, Carlsbad, USA) with DNA ladder, used electrophoresis to separate DNA bands, and visualized under UV exposure.

Clinical Detection of Shigella, Salmonella, and Yersinia spp.

Screening protocol of *Shigella, Salmonella, and Yersinia* spp. is previously described [103]. Briefly, stool samples were used to inoculate Salmosyst broth (Merck, Darmstadt, Germany) with supplements to select for *Shigella* and *Salmonella* species. The bacteria able to grow in the selective media were subsequently cultured on Salmonella-Shigella (SS) agar (Scharlau, Sentmenat, Spain) for isolation. Alternatively, CESY Broth (Merck, Darmstadt, Germany) was used for recovery of *Yersinia* species from stool samples, and subsequently cultured on CIN-Agar (Merck, Darmstadt, Germany) for further selection of *Yersinia* species. All incubation temperatures and times were 37°C for 24 hours.

Clinical Detection of *Campylobacter*

Campylobacter was detected by bacterial isolation on Campylosel agar (Biomerieux, Marcy-l'Étoile, France) and confirmed through biochemical testing using API CAMPY (SKU#20800, Biomerieux, Marcy-l'Étoile, France) [103] and quantitative PCR (qPCR) testing as previously described [33]. Briefly, human stool samples were diluted 1:10 in DEPC-treated water (Life Technologies, Carlsbad, USA), vortexed, and centrifuged. DNA was extracted using a QIAamp® DNA stool mini kit (Qiagen, Valencia, CA) and stored at −20°C until use. A qPCR assay was developed using a previously described assay for *C. jejuni/C. coli* on the CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules, California, USA) [105]. Each well included a 25 µl reaction mixture with 1 µl of DNA sample, 12.5 µl of TaqMan® environmental master mix, 9 µl of nuclease free water, 1 µl of each primer of *cad*F-Forward

"CTGCTAAACCATAGAAATAAAATTTCTCAC" and *cad*F-Reverse "CTTTGAAGGTAATTTAGATATGGATAATCG" at a final concentration of 0.4 μM, and 0.5 μl of *cad*F-Probe "[HEX]-CATTTTGACGATTTTTGGCTTGA-[BHQ2]" at a final concentration of 0.2 μM. The cycling conditions were as follows: 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and 55°C for 1 min.

Preparation of Fecal Samples for Study

Classification of Cohorts for this Study

The inclusion and exclusion criteria for this study's subjects is shown in table 1. The subjects for this study were originally divided into 2 groups:

1) "*Campylobacter* infected" – these fecal samples were positive for *Campylobacter* by qPCR identification (n=20).
2) "*Campylobacter* uninfected"– these fecal samples were negative for *Campylobacter* by qPCR identification (n=20).

In order to control for presence of acute diarrheal disease in some of our *"Campylobacter* uninfected" subjects, the cohorts were divided again:

1) "*Campylobacter* infected - symptomatic" or "CIS" – these fecal samples were positive for *Campylobacter* by qPCR identification and presented with diarrhea (n=14).
2) "*Campylobacter* infected - asymptomatic" or "CIA" – these fecal samples were positive for *Campylobacter* by qPCR identification but presented without diarrhea (n=6).

3) "*Campylobacter* uninfected - symptomatic" or "CUS" – these fecal samples were negative for *Campylobacter* by qPCR identification but presented with diarrhea (n=8).
4) "*Campylobacter* uninfected - asymptomatic" or "CUA" – these fecal samples were negative for *Campylobacter* by qPCR identification and presented without diarrhea (n=12).

DNA Extraction from Fecal Samples

Total microbial DNA was isolated from feces using the DNeasy PowerSoil Kit (QIAGEN Sciences Inc. Germantown, MD) following the manufacturer's protocol by thesis author. DNA was stored at -20°C until 16S rRNA gene amplification.

Generation and Sequencing of 16S rRNA Gene Amplicon Libraries

Primers targeting the V3-4 region of the 16S rRNA gene were used to create a single amplicon of approximately 460 bp (16S Amplicon PCR Forward Primer: 5'-CCTACGGGNGGCWGCAG-3' and 16S Amplicon PCR Reverse Primer: 5'-GACTACHVGGGTATCTAATCC-3') according to the Illumina 16S Metagenomic Sequencing Library Preparation Guide. Once the library was quantified and normalized, it was pooled and sequenced on an Illumina MiSeq using v3 600 cycle kit. The generation and sequencing of the 16S rRNA gene amplicon libraries were performed by Veronica Brown in the University of Tennessee Genomics Core facility with assistance from thesis author.

Extraction of Fecal Metabolites

Extraction of metabolites from fecal samples was performed under a biological safety hood by thesis author. A 0.1 gram sample was prepared by adding 0.650 mL of Metabolic Extraction Solvent (MES) (20:40:40 H_20:ACN:MeOH + 0.1 M Formic acid) and mixing thoroughly before adding another 0.650 mL of MES and chilling at -20°C for 20 minutes. Samples were subsequently centrifuged (4°C at 13,500 rpm) and the supernatant was transferred to a fresh microcentrifuge tube and stored at 4°C. The resulting pellet was resuspended in 0.2 mL of MES and chilled at -20°C for 20 minutes before centrifugation and transfer of the supernatant to a fresh microcentrifuge tube. The sample tubes were immediately put on ice and transported to the Biological and Small Molecule Mass Spectrometry Core (BSMMSC) of the University of Tennessee – Knoxville. Once at the BSMMSC, the supernatant was then dried under nitrogen and the resulting pellet was resuspended in 0.300 mL of milliQ H_2O before being transferred to autosampler vials.

Ultra-High-Performance Liquid Chromatography Tandem Mass Spectrometry (UHPLC–MS/MS) Analysis of Fecal Metabolites

The Generation of fecal metabolomes was performed by the Biological and Small Molecule Mass Spectrometry Core of the University of Tennessee – Knoxville. Samples were separated on a Phenomonex Synergi Hydro RP, 2.5 um, 100 mm x 2.0 mm column at 25˚C. Mobile phase elution of metabolites were (A) 97:3 Methanol to Water with 11 mM Tributylamine and 15 mM Acetic acid and (B) 100% Methanol. The gradient used for mobile phase A during the 25-minute method with a flow rate of 0.2 mL/minute

was 100% at minute zero, 80% at minute 5, 45% at minute 13, 5% at minute 15.5, 100% at minute 19 and 25, while the mobile phase B gradient equalized the percentage. The Exactive Plus Orbitrap used an electrospray ionization (ESI) probe operating in negative mode with a scan range of 72-1000 m/z.

Inductively-Coupled Plasma Mass Spectrometry (ICP-MS) Analysis of Fecal Micronutrients

The ICP-MS analysis was performed by the Spectroscopy and Biophysics Core of the University of Nebraska, Lincoln. The fecal samples were weighed and suspended in 1 mL of nitric acid (70% w/v). After overnight digestion at 65°C, the tubes were cooled, loaded in triplicate into 96-well plates, and diluted 20-fold with a solution of 50 ppb Ga in 2% nitric acid. Counts for cobalt, chromium, copper, iron, manganese, molybdenum, selenium, and zinc were normalized using this internal standard and were converted to concentrations using an external calibration curve. The instrument setup and acquisition method for ICP-MS were conducted as previously described [106].

Statistical Analyses

Epidemiological Analysis of Subject Data

Descriptive statistics were generated for all clinical and demographic variables. For continuous variables, the association with being a *Campylobacter* case was determined by t-test or the non-parametric equivalent, the Mann-Whitney U-test. Categorical variables were analyzed with chi-square tests, with the exception of Fisher's exact tests were used when expected cell counts were less than five. Wilcox two-

sample t-tests, Odds ratios, 95% confidence intervals, and p-values were determined with significance inferred at a p-value < 0.05.

Analysis of Fecal Microbiomes

The amplicon sequencing reads were analyzed using mothur, an open source and expandable software [107]. Sequencing and PCR errors were first reduced by combining forward and reverse reads and scoring sequencing alignment, followed by alignment of unique sequences to the SILVA 16S rRNA gene reference database (version 132), which were then clustered into the corresponding operational taxonomic units (OTUs) in the SILVA database.

To improve analytic results, data filtering was used to identify and remove features (taxa in this case) that are unlikely to be of use when modeling the data [108]. These taxa are determined by either filtering based on abundance levels (minimum counts) across samples (prevalence) or using minimum count cutoff based on their mean and median values by the program [108]. Taxa with constant or less variable abundances are crucial for comparative analysis and are filtered based on their inter-quantile ranges, standard deviations or coefficient of variations. By default, taxa having zero counts across all the samples, or that only appear in one sample will be removed from further analysis [108].

In order to ensure proper filtration, many settings were tested. The results shown in the following chapter were produced using a low count filter with the minimum count set to 4 and a prevalence in samples set to 20%. While in appendix results for settings using a low count filter with the minimum count set to 2 and 6 and a prevalence in

samples set to 10% and 25% are available. The low variance filter settings shown in the results section data/figures were to remove 10% based on the inter-quantile range. After filtration, 138 features were still remaining. In appendix, results from other low variance filter settings can be viewed. These settings include removal of 15% based on inter-quantile range.

To normalize, the data were rarefied to the minimum library size (Fig. 2). No dating scaling or transformation was performed for any results shown in the results section. In the appendix, results from other normalization settings can be viewed. These settings include total sum scaling and relative log expression.

R or MicrobiomeAnalyst [108] were used for OTU and population-level analyses, which determined whether any OTUs were differentially represented between the samples. ANOSIM or PERMANOVA were used to determine whether statistically significant differences existed between the microbial communities in infected and uninfected samples. Briefly, a Mann-Whitney U-test was performed to compare the bacterial taxa present in the two groups and the p-values were adjusted for multiple comparisons using the Benjamini and Hochberg False Discovery Rate (FDR) and significance inferred at $p < 0.05$. Additionally, visualizations of the different taxa between the groups were created using the linear discriminant analysis effect size (LEfSe) [109].

Analysis of Fecal Metabolomes

Samples were normalized by weight prior to inputting data into MetaboAnalyst, a statistical, functional and integrative analysis of metabolomics data [110]. Since data containing zeros or missing variables will cause difficulties in downstream analyses,

these were were replaced with the small value of 4.431321 in the metabolomic data. This value is half of the minimum positive values in the original data, which is assumed to be the detection limit. The assumption of this approach is that most missing values are caused by low abundance metabolites (i.e. below the detection limit) [110].

For multi-group analysis, MetaboAnalyst provides one-way Analysis of Variance (ANOVA). While ANOVA only tells whether the overall comparison is significant or not, it can be followed by post-hoc analyses in order to identify which two levels are different [110]. MetaboAnalyst provides two most commonly used methods for this purpose - Fisher's least significant difference method (Fisher's LSD) and Tukey's Honestly Significant Difference (Tukey's HSD). ANOVA and post-hoc analyses were performed with significance inferred at a p-value < 0.05.

Principal Component Analysis (PCA), an unsupervised method aiming to find the directions that best explain the variance in a data set (X) without referring to class labels (Y) [111], was performed using the prcomp package and requiring the R script 'chemometrics.R' in MetaboAnalyst [110]. The data in the PCA are summarized into fewer variables called scores which are weighted average of the original variables, and the weighting profiles are called loadings [111].

Another analysis utilized from MetaboAnalyst [110] is the Partial Least Squares - Discriminant Analysis (PLS-DA). PLS is a supervised method using multivariate regression techniques to extract the information, via linear combination of original variables (X), that can predict the class membership (Y) [111]. A permutation test was performed to assess the significance of class discrimination, and in each permutation, a

PLS-DA model was built between the data (X) and the permuted class labels (Y) using the optimal number of components determined by cross validation for the model based on the original class assignment [112].

MetaboAnalyst supports two types of test statistics for measuring the class discrimination based on: 1) prediction accuracy during training, and 2) separation distance based on the ratio of the between group sum of the squares and the within group sum of squares (B/W- ratio). The class discrimination cannot be considered significant from a statistical point of view, if the test statistic is part of the distribution based on the permuted class assignments [110]. Two variable importance measures in PLS-DA are: 1) Variable Importance in Projection (VIP) is a weighted sum of squares of the PLS loadings taking into account the amount of explained Y-variation in each dimension [111], and 2) based on the weighted sum of PLS-regression. The weights are a function of the reduction of the sums of squares across the number of PLS components [111]. VIP scores are calculated for each of the components and when more than components are used to calculate the feature importance, the average of the VIP scores are used [110].

The sparse PLS-DA (sPLS-DA) algorithm can be used to produce robust and easy-to-interpret models by reducing the number of metabolites in high-dimensional metabolomic data [113]. Also controlling the number of components in the model and the number of variables in each component, the sparseness of the model can be controlled [113].

Random Forest is a supervised learning algorithm suitable for high dimensional data analysis by using an ensemble of classification trees, each of which is grown by random feature selection from a bootstrap sample at each branch [110]. Random Forest analysis was used to detect the impact of different metabolites between groups [114]. Also, Random Forest provides other useful information such as OOB (out-of-bag) error, variable importance measure, and outlier measures [110]. Significant features identified by Random Forest, and are ranked by the mean decrease in classification accuracy when they are permuted [110]. Additionally, all t-tests were performed with significance inferred at a p-value < 0.05.

Analysis of Micronutrient Profiles

Differences between infected and uninfected groups were detected using the ROUT method to identify outliers with a Q = 0.1, then differences between infected and uninfected cohorts were identified using an unpaired t-test (GraphPad Prism Software, La Jolla California USA) and significance inferred at a *p*-value < 0.05.

CHAPTER FOUR – RESULTS

Characteristics of Subjects

This study was conducted using fecal samples from 40 children under the age of five years old from Bucaramanga, Colombia metropolitan area. The inclusion and exclusion criteria for this study's subjects is shown in Table 1. The subjects for this study were originally divided into 2 broader cohorts: i) "*Campylobacter* infected" – these fecal samples were positive for *Campylobacter* by qPCR identification (n=20), and ii) "*Campylobacter* uninfected" – these fecal samples were negative for *Campylobacter* by qPCR identification (n=20). These cohorts will be referred to as "infected" and "uninfected" from this point on.

To control for presence of acute diarrheal disease in some of our *"Campylobacter* uninfected" subjects, the cohorts were divided again into two. These cohorts include the symptomatic/asymptomatic status of the subject**: CIS:** "*Campylobacter* infected - symptomatic" (n=14**), CIA:** "*Campylobacter* infected - asymptomatic" (n=6**), CUS:** "*Campylobacter* uninfected - symptomatic" (n=8), and **CUA:** "*Campylobacter* uninfected - asymptomatic" (n=12). These groups will be referred to as individually CIS, CIA, CUS, and CUA or as "the four cohorts" from this point on, except in appendix 2 where they are referred to as 1, 2, 3, and 4, respectively.

The sociodemographic information is included in Table 2 for the "infected" and "uninfected" cohorts. The age group 12 to 24 months was the only variable that significantly differed between the two cohorts (p-value = 0.048). The mean and median

age for all subjects were 21.2 and 16.28 months, respectively, and contained 65% female and 70% mestizo participants. The sociodemographic information for the four cohorts is shown in Table 3. None of the variables were found to be significantly different between the four cohorts by the Kruskal-Wallis test.

Clinical Symptoms of Subjects

The Wilcoxon two-sample t-test revealed that the *Campylobacter* infected cohort had a significantly longer number of days experiencing diarrhea than the *Campylobacter* uninfected cohort ($p = 0.030$), and also experienced more diarrheal episodes in a 24-hour period ($p = 0.043$). Although no significant differences were detected for the presence of fever, vomiting, diarrhea, bloody stool, mucus, or abdominal pain, the clinical symptoms for each patient were recorded and are shown in Table 4. Due to the low subject numbers in the four cohorts, clinical symptom analysis was not performed.

Environmental Exposures and *Campylobacter* Infection

To determine whether certain environments correlate with *Campylobacter* infection, we examined for associations of each group with various exposures. The only significant correlation that was found was that children in the uninfected cohort were more likely to have had exposure to birds in the week prior to enrollment in the study (Table 5). This is somewhat surprising since birds are a natural host for *Campylobacter*. In contrast, children exposed to cats trended toward infection with *Campylobacter* than children who were not exposed (Table 5).

Other Bacterial and Viral Pathogens Detected in Subject Samples

In terms of pathogen manifestations, 100% of the *Campylobacter* infected samples were positive for *Campylobacter* infection. The *Campylobacter* uninfected subjects were positive for other pathogens 15% of the time, while 85% of the *Campylobacter* uninfected samples were negative for the tested bacterial or viral infections (Table 6). Although co-infections were observed in both cohorts, no correlations between any specific pathogen and cohort, aside from *Campylobacter*, was identified. Children with *Campylobacter* infection were approximately five times more likely to be colonized by non-pathogenic (commensal) *E. coli* than the uninfected children, but pathogenic enteroaggregative *E. coli* colonization occurred in 10% of the samples in each cohort. Table 7 shows the bacterial and viral pathogens detected in all 4 cohorts. The most interesting observation being that CIS has the highest number of co-infections (5), while cohorts CIA and CUA only have 1 each (Table 7).

Microbiome Analysis

Analysis of alpha-diversity (the diversity within a sample based on richness and evenness) showed no significant differences in Chao1, Shannon diversity index, or species richness between infected and uninfected cohorts (p-value=0.18744) (Fig. 3A). When analyzing the four cohorts for alpha-diversity, no significant differences were detected (p-value=0.327) (Fig. 4). Similarly, the beta-diversity (the difference in taxonomic abundance profiles between the cohorts) were analyzed by principal coordinate analysis (PCoA) and no significant differences in the microbiome were

detected when comparing infected and uninfected groups (p-value=0.074) (Fig. 3B) or among all four cohorts (p-value=0.227) (Fig. 5).

Linear discriminant analysis (LDA) effect size (LEfSe), an analysis that characterizes differences between two or more sample groups with an emphasis on statistical significance and biological relevance using an algorithm for High-Dimensional biomarker discovery and explanation that identifies genomic features, was used to identify bacterial taxa that differed between infected and uninfected groups. First, LEfSe utilizes the non-parametric factorial Kruskal-Wallis (KW) sum-rank test for differential abundance analysis, followed by determining biological relevance by using the (unpaired) Wilcoxon rank-sum test, and lastly applying the Linear Discriminant Analysis to estimate the effect size of each differentially abundant taxon.

As expected, the only statistically significant family in infected children was Campylobacteraceae (LDA= 2.76; p=0.004). Although Enterobacteriaceae exhibited the largest effect size in uninfected children, it was not statistically significant (LDA= -3.27; p=0.148). Using a univariate statistical comparison of the groups, only the Epsilonproteobacteria, of which *Campylobacter* is a member, in the infected children was statistically significant ($p < 0.001$; FDR < 0.0268) (Fig. 6).

When using LEfSe to analyze all four cohorts, the class Epsilonproteobacteria was the only one with significant differences (p-value= 0.0011647; FDR = 0.013976; LDA = 2.98). At the order level, Campylobacterales was the only significantly different taxon by FDR (p-value= 0.0011647; FDR = 0.01747; LDA = 2.98). Pasteurellales did have a significant p-value of 0.0202, but the FDR was insignificant at 0.15162. The last

taxonomic level with a significant LEfSe difference was Family. Campylobacteraceae was the only significantly different taxon (p-value= 0.0011647; FDR = 0.037269; LDA = 2.98).

While not statistically significant, direct quantitative comparison of taxonomic abundances from the infected and uninfected cohorts revealed that, at the phylum level, Firmicutes were similar between both cohorts, while Bacteroidetes had a higher relative abundance in *Campylobacter* infected children (19% infected; 13% uninfected). In contrast, Proteobacteria comprised 32% of the overall relative abundance in uninfected children and 24% in infected children. At the family level of Proteobacteria, the percentage of Enterobacteriaceae comprised 98% of the total Proteobacteria abundance in the uninfected samples yet only 70% in the infected, likely due to Campylobacteraceae abundance being 2% in uninfected children, but 21% in the infected cohort (Fig. 7).

When analyzing all four cohorts, more differences noticeable differences among the cohorts are observed (Fig. 8). Interestingly, the percentage of the phylum Proteobacteria abundance is 46.7% in CUS compared to the average percentage of 24.5 of the three other cohorts. CUS also has the lowest percentage of Firmicutes (Fig. 8). Although, CUS was culture and qPCR negative for *Campylobacter*, the percent abundance at the family level for Campylobacteraceae is higher in this cohort (1.73%) than in the qPCR positive CIA (0.11%). When looking specifically at the Proteobacteria phylum, Enterobacteriaceae (*E. coli*'s family) was lowest in CIS (14.2%) and highest in CUS (44.9%) (Fig. 9).

Metabolomic Analysis

Since microbiome analysis by 16S rRNA gene characterization lacks the ability to give detailed information about the functional microbial activity, metabolomic analysis can give us a glimpse of the intermediate phenotype mediating host–microbiome interactions [115]. Using metabolomic analysis of the fecal samples by ultra-high-performance liquid chromatography tandem mass spectrometry (UHPLC–MS/MS), we observed 533 metabolites (peaks) of which 62 were assigned as specific metabolites in the database. Differences in metabolite abundance were detected and visualized using a combination of partial least squares - discriminant analysis (PLS-DA), Random Forest analysis, and univariate statistics.

PLS-DA demonstrated clustering of most samples to the center with some separation due to a few samples. This visualization is provided for *Campylobacter* infected vs. uninfected in Fig. 10A, and for the separation among all four cohorts in Fig. 11. More specific comparisons of metabolite abundance identified several that were significantly affected and drove much of the separation observed in the PLS-DA plot of figure 10A, including three metabolites that were shown to be significantly higher in uninfected children: ascorbate, cholate, and phenylpyruvate and three in the infected cohort: hydroxyisocaproic acid, malate, and 2-hydroxy-2-methylsuccinate. It is interesting that when the cohorts are divided by presence of diarrhea, some trends are visible. Whereas some of the major metabolites still associate specifically with *Campylobacter* infection status, such as cholate and malate, some of the metabolites seem to possibly associate with the presence of diarrhea such as hydroxyisocaproic

acid (Fig. 11). The sPLS-DA shows a more significant division between CIS, CUS, CUA (Fig. 12).

Using an ensemble learning method for classification trees (Random Forest analysis), we found that folate, riboflavin, lysine, and glutathione disulfide were all lower in infected children, whereas glucose-1-phosphate, glucose-6-phosphate, orotate, creatinine, succinate, tryptophan, and others were increased (Fig. 10B). Fig. 13 shows similar results from the random forest analysis of the four cohorts.

When cross referencing the top 15 metabolites responsible for the separation in the infected vs. uninfected PLS-DA plot with the top 15 identified by the infected vs. uninfected Random Forest analysis, we found that 4 metabolites were identified by both. These metabolites were succinate, tryptophan, leucine, and malate. Once these metabolites were cross referenced with the results of the four cohort analysis, we found that tyrosine and malate were the only metabolites present in the top 15 of every single analysis.

Fecal Micronutrient Comparisons

In addition to examining for changes to metabolites during pediatric *Campylobacter* infection, we also quantified using inductively-coupled plasma mass spectrometry (ICP-MS) the essential micronutrients cobalt, chromium, copper, iron, manganese, molybdenum, selenium, and zinc. ICP-MS is capable of detecting metals and several non-metals at concentrations as low as one part per quadrillion on non-interfered low-background isotopes by ionizing the sample with inductively coupled plasma and then

using a mass spectrometer to separate and quantify those ions [37]. From these fecal micronutrient profiles, we observed significant decreases for the infected cohort in all, but chromium (Fig. 14). Statistical analysis of Co, Cr, Cu, Fe, Mn, Mo, Se, and Zn levels gave p-values = 0.0258, 0.2711, 0.0003, 0.0050, 0.0131, 0.0367, 0.0017, and 0.0009 respectively.

CHAPTER FIVE – DISCUSSION

Campylobacter jejuni is responsible for 96 million annual cases of bacterial-derived gastroenteritis worldwide [2], exceeding the number of cases produced by other well-known enteric pathogens including strains of pathogenic *E. coli* and *Salmonella* [2]. To date, the impacts and mechanisms of *Campylobacter* infection in human hosts remain mostly unknown though it appears to affect the developed and developing worlds in unique ways. One such area is the impact of *Campylobacter* infection on pediatric subjects in the developing world, which includes the development of EED, growth faltering, and cognitive impairment.

It was these negative effects on pediatric populations in the developing world that initially led us to examine changes to the microbiota, metabolome, and micronutrient abundances of Colombian children infected with *Campylobacter*, though it is important to note that these cohorts were not evaluated for the above conditions (e.g. EED). It is important to note that in our study, the isolation and testing of *Campylobacter* was specifically geared towards *C. jejuni* and *C. coli*, which could potentially explain why *Campylobacter* 16S rRNA genes were identified in CUS. These could have been non-*C. jejuni/C. coli* species that weren't identified due to targeting *C. jejuni/C. coli* with our methods.

Additionally, we sought to determine which sociodemographic variables and disease manifestations are most often associated with infection. Surprisingly, bird exposure was more closely associated with the uninfected cohort while children exposed to domestic cats, while not statistically significant, were approximately five

times more likely to be in the infected cohort. While cats are seldomly considered a reservoir of *Campylobacter*, several studies have reported that a small percentage of cats are asymptomatically infected with *Campylobacter* [116-118].

Children with *Campylobacter* infection were more likely to experience increased diarrheal episodes in a 24-hour period and more days with diarrhea, which are typical manifestations of acute *Campylobacter* infection [119]. Clinical microbial testing found that *Campylobacter*-infected children were more likely to be colonized by commensal *E. coli* than uninfected children leading us to further examine the impacts of *Campylobacter* infection on the gastrointestinal microbiota. These impacts were an important consideration since: i) the mucosal immune system and intestinal microbiota are interrelated and develop during the first years of life, ii) diet and enteric infections shape these systems, and iii) the gastrointestinal immune status and microbiota broadly impact gut health and nutrient absorption [94, 120-122].

The healthy human gut microbiota is dominated by anaerobes with an abundance of 2-3 orders of magnitude higher than facultative anaerobes [123]. Dysbiosis occurs when the population of facultative anaerobic bacteria expands due to a reduction in epithelial hypoxia, which is induced by colonic inflammation – a hallmark of EED [51]. In an inflammatory environment, terminally differentiated surface colonocyte metabolism shifts from mitochondrial β-oxidation of fatty acids toward anaerobic glycolysis [61], which results in high glucose consumption, low oxygen, and increased release of lactate [61]. This shift in colonocyte metabolism ultimately leads to

elevated amounts of oxygen diffusing to the mucosal surface, which favors the expansion of facultative anaerobic bacteria [61].

This facultative anaerobic bloom, which is dominated by members of the Enterobacteriaceae, results in an inflammatory response that alters the nutritional environment of the gut through the generation of reactive oxygen species (ROS) and reactive nitrogen species (RNS) [124]. While ROS and RNS have antimicrobial activity, they are also quickly converted to non-toxic compounds such as nitrate, the most potent electron acceptor under anaerobic conditions [125]. Since the ability to utilize nitrate respiration is highly conserved among the Enterobacteriaceae, this nitrate conversion further supports the expansion of the commensal Enterobacteriaceae present [124]. Another nutrient source provided by host inflammation is migration of neutrophils to the intestinal lumen, which can produce ROS and convert endogenous sulfur compounds into electron acceptors such as conversion of thiosulfate into tetrathionate [125]. These environmental modifications would benefit *Campylobacter* since its able to use both nitrate and tetrathionate for respiration [126].

The impact of *Campylobacter* infection on the pediatric microbiota is understudied, particularly as it relates to EED development. A recent longitudinal study of the gut microbiota in young children from southern India [127] found these children were enriched for Campylobacterales, though the impact of *Campylobacter* colonization on microbial community composition was not investigated. In our study, which examined the relationship of *Campylobacter* infection and the microbiota, we found that total relative abundance of Proteobacteria was higher in in uninfected children (32%),

yet less relatively abundant in the infected subjects (24%). When examining for changes at the Proteobacteria family level, we observed a shift in the relative abundance of Enterobacteriaceae from 98% in the uninfected cohort to comprising only 70% in the infected. We observed a change in Campylobacteraceae abundance of 2% in the uninfected cohort and 21% in the infected cohort within the phylum, it is possible that Enterobacteriaceae were displaced by *Campylobacter* or by Pasteurellaceae, which was in infected children. These results were interesting in that we did not observe a change in obligate anaerobes like Bacteroidetes or Firmicutes despite relatively high numbers of facultative anaerobes in both cohorts. This indicates that the uninfected children are already abundant for Enterobacteriaceae and that these organisms have already limited the dominance of anaerobic members of the community.

It is likely that when *Campylobacter* infection occurs in these children, since the anaerobic members have already been limited, no further decrease occurs. Since *Campylobacters* are predominantly microaerophilic, with an oxygen preference of 3-12%, the loss of epithelial hypoxia in children enriched for Enterobacteriaceae may ultimately benefit *Campylobacter* growth and allow for the persistent colonization that can occur in children from these communities [51, 126]. This concept is highlight in Fig. 1 with (a) showing a “healthy” epithelial lining [51] and (b) showing what may possibly be going on to allow *Campylobacter* asymptomatic colonization.

Analysis of the metabolomes from *Campylobacter*-uninfected and *Campylobacter*-infected children identified several metabolites that were significantly

different between the two groups. In uninfected children, three metabolites were found to be higher, including ascorbate, cholate, and phenylpyruvate.

Ascorbate (vitamin C) is of interest as children in the developing world are prone to low ascorbate and may develop scurvy [128]. Ascorbate boosts intestinal absorption of non-heme iron and is an important antioxidant that reacts with hydrochlorous acid, a product of an anti-microbial agent produced by neutrophils via myeloperoxidase (MPO) [129]. MPO is a commonly used biomarker for inflammation [130] and oxidizes ascorbate directly leading to depleted ascorbate levels [129]. Although we were unable to quantify MPO levels due to the limited quantity of sample available, the presence of fever, diarrhea, and the number of diarrheal episodes in the infected cohort suggests MPO levels may have been elevated. A plausible explanation for higher ascorbate levels in the uninfected cohort is that there is less MPO to oxidize the ascorbate in the intestine allowing for more absorption and excretion.

Cholate was not only higher in the fecal samples of uninfected children relative to those of infected children, but these results were mirrored in the analysis of the four cohorts. Cholate is a bile acid that is often converted to the secondary bile acid, deoxycholate, through microbial dehydroxylation in the colon [131]. Deoxycholate stimulates the immune system, leading to increased reactive oxygen species (ROS) generation and negatively impacting the health of intestinal tissue [131]. Phenylpyruvate, higher in the uninfected cohort, is generated from phenylalanine by oxidative deamination, leading to the replacement of alanine with pyruvate [132]. Conversely, we observed higher levels of phenylalanine in the fecal samples of infected

children relative to the uninfected, but the four-cohort analysis revealed higher relative concentrations in symptomatic subjects than *Campylobacter*-infected subjects. This suggests reduced catabolism or absorption of the amino acid in the gut of symptomatic subjects. Phenylalanine is converted to the essential amino acid, tyrosine, by the enzyme phenylalanine hydroxylase (PAH) [133]. Tyrosine was one of the two metabolites to be identified in all analyses, and had higher relative concentrations in symptomatic subjects. Significantly, when a systemic decrease in phenylalanine occurs, it may lead to decreased systemic levels of tyrosine, as observed in patients with phenylketonuria (PKU) [133]. This could lead to decreased cognitive function and physical development, which result from low levels of tyrosine [133]. Since we do not have dietary records for each subject, we cannot eliminate the possibility that these differences were caused by variations in diet. The changes in relative concentrations of essential nutrients detected in the uninfected cohort's feces are likely explained by an increase of excretion indicating: 1) poor absorption leading to lower systemic levels, 2) sufficient systemic levels leading to lack of absorption, or 3) differences in the microbiota that metabolize these nutrients [125, 134-137].

Additionally, in the symptomatic subjects, we observed higher relative concentrations in hydroxyisocaproic acid, which is an end-product of leucine metabolism; malate, which is an intermediate of central metabolism, and methylsuccinate, which is a methyl-branched fatty acid [138]. We employed a Random Forest analysis to identify additional molecules that are significantly different between the groups. From this analysis, we also identified that glucose-1-phosphate, glucose-6-

phosphate, orotate, creatinine, succinate, and tryptophan were higher in infected children. Glucose-1-phosphate and glucose-6-phosphate are both intermediates of many metabolic pathways such as glycolysis, gluconeogenesis, glycogenolysis, and the pentose phosphate pathway. Orotate is involved in pyrimidine biosynthesis, with orotidine-5'-monophosphate serving as the final intermediate before uridine monophosphate synthesis [139]. Creatinine is a byproduct of muscle metabolism and serum creatinine is often excreted from the body through the urine [140]. Succinate is a key intermediate in the aerobic TCA cycle, providing electrons to the electron transport chain for energy production [40]. These changes in relative concentrations suggest that there was either a large surplus of these metabolites, changes in the host colonocyte metabolism, or a shift in the microbiota that would reduce the uptake/consumption of these exceptionally beneficial energy sources.

Tryptophan (Trp) is known for being the least abundant and most conserved among the 20-common canonical amino acids [141]. Humans primarily rely on dietary Trp intake, but also on Trp derived from the gut microbiota [141]. Interestingly, Trp metabolism plays a critical role in mediating immune system-intestinal microbiota interactions through the detection of Trp-derived metabolites like indole. *Campylobacter* could be disturbing this important host pathway by either directly driving Trp accumulation by synthesizing Trp from indole, or indirectly by altering the microbiota causing a decrease of Trp metabolism. Recently, a study of a novel mouse model of *Campylobacter*-induced enteropathy and diarrhea observed an increase of Trp

concentrations during *Campylobacter* infection [142], which was supported by our observation that Trp was higher in the infected cohort relative to the uninfected.

Micronutrient analyses were performed using ICP-MS for essential metals on all fecal samples. This technique allowed for the simultaneous detection of cobalt, chromium, copper, iron, manganese, molybdenum, selenium, and zinc with high sensitivity [106]. We found that all essential metals except chromium were significantly reduced by *Campylobacter* infection. While micronutrient depletion by diarrhea is likely, the presence of diarrhea was insignificant between cohorts (p-value > 0.057). However, the number of events and the duration of diarrhea did significantly associate with infection status (p-value < 0.03). This potentially indicates that having more persistent diarrhea, which would likely occur during the development of EED, may lead to micronutrient deficiencies that would impact host development.

We were particularly interested in the reductions in iron and zinc concentrations because they are associated with immune suppression and decreased development, two manifestations of EED that were found to correlate with pediatric *Campylobacter* infection [143]. On average, adults require a daily intake of iron sufficient to replace that lost through the GI tract [137]. Children, due to their constant growth and increase in muscle mass, require the intake of significantly more iron to meet demands [137, 144]. Not only did we observe significantly lower concentrations in iron and zinc, but copper as well. The lower concentration in copper is supportive of a concomitant diminishment in iron as anemia may be associated with a deficiency in metals [145]. These results suggest micronutrient levels have a significant role in *Campylobacter* infections.

CHAPTER SIX – RECOMMENDATIONS AND FUTURE WORK

Campylobacter is estimated by the CDC to be the cause of approximately 9% of all foodborne illnesses in the United States, and 15% of all campylobacteriosis cases leading to hospitalization [146]. While multiple antibiotic resistant *C. jejuni* is considered a “Serious Threat” to public health in the United States [14], *Campylobacter* is also implicated in growth faltering among children <2 years of age in low to middle-income countries [100]. With the mounting evidence regarding the burden of *Campylobacter*-attributed diarrhea as well as growth-faltering in these populations, it is imperative for research to be conducted to understand *Campylobacter*’s mechanisms of pathogenesis and the both the short- and long-term impacts it has on the human host.

This study demonstrated that *Campylobacter* infection correlated with significant alterations to the gastrointestinal environment in pediatric populations. In the future, It would be interesting to analyze the 16S *Campylobacter* sequences from this study to identify which species are included in each cohort. This may give insights on species that may be more associated with asymptomatic colonization or which species were not identified by our qPCR primers. It would be beneficial to use the weaning-age ferret model of Campylobacteriosis to replicate the results from this study, and possibly give more mechanistic data on *Campylobacters* pathogenesis. In order to more accurately represent these middle-income populations, a humane induction of malnutrition would be the best option by decreasing the animals’ food intake by 15%.

In the future, it would also be interesting to conduct a similar, but longitudinal study to determine whether any of these children become persistently colonized and what the impact of that colonization is on the gastrointestinal environment.

Using knowledge gained from the MAL-ED and other similar studies, a longitudinal study designed specifically for *Campylobacter* would gain critical knowledge for the field. *Campylobacter* is strongly associated with negative outcomes during colonization of the second year of life [100], so enrolling subjects prior to 6 months of age and following them until 24 months would be a necessary minimum requirement. To be collected at the time of enrollment, every month until 1 year, every 3 months until 24 months: 1) a complete clinical and epidemiological questionnaire to be given to the parents/guardians of the subject, 2) fecal samples from the subject and all members and animals of the household, and 3) blood and urine sample from subject. If the subject has a diarrheal episode in between visits, a fresh diarrheal sample, height and weight measurements would be immediately needed.

The fecal samples would be utilized to characterize the microbial populations by 16S rRNA gene analysis at each time point and diarrheal episode, which would allow the observation of changes/shifts in the microbiome throughout time or course of infection. These samples would also be analyzed for biomarkers of intestinal inflammation and injury, such as methylene blue, fecal occult test, and myeloperoxidase. These biomarkers would allow insight on the inflammatory status of the subjects' gastrointestinal tract [130, 147]. The urine and blood samples would allow us to determine the systemic micronutrient levels more accurately. This is crucial for

determining the nutritional status of the subject and how that status may associate with colonization or infection.

Due to the high genomic plasticity of *Campylobacter* species, it would be interesting to utilize whole genome sequencing of *Campylobacters* isolated from the feces to see identify potential virulence factors in the genome that may be associated with a specific disease outcome. This study design would be incredibly useful for determining what characteristics (host or microbial) that lead to asymptomatic *Campylobacter* infections. It would also be informative to perform metagenomics on the entire fecal population, which could potentially link functional pathways to identified metabolites from the UHPLC-MS/MS analysis of the samples.

Such a longitudinal study would allow for the identification of children that exhibit decreased development and whether that outcome is potentially due to alteration in the gastrointestinal environment. These insights would contribute to EED prevention and/or treatment leading to enormously positive impacts on pediatric health, but also a substantially positive economic impact on the developing world.

REFERENCES

1. Facciolà, A., et al., *Campylobacter: from microbiology to prevention.* Journal of Preventive Medicine and Hygiene, 2017. **58**(2): p. E79-E92.
2. Havelaar, A.H., et al., *World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010.* PLOS Medicine, 2015. **12**(12): p. e1001923.
3. Newell, D.G., *Animal models of Campylobacter jejuni colonization and disease and the lessons to be learned from similar Helicobacter pylori models.* Journal of Applied Microbiology, 2001. **90**(S6): p. 57S-67S.
4. Johnson, T.J., J.M. Shank, and J.G. Johnson, *Current and Potential Treatments for Reducing Campylobacter Colonization in Animal Hosts and Disease in Humans.* Frontiers in microbiology, 2017. **8**: p. 487-487.
5. Young, K.T., L.M. Davis, and V.J. DiRita, *Campylobacter jejuni: molecular biology and pathogenesis.* Nature Reviews Microbiology, 2007. **5**: p. 665.
6. Shank, J.M., et al., *The Host Antimicrobial Protein Calgranulin C Participates in the Control of Campylobacter jejuni Growth via Zinc Sequestration.* Infection and Immunity, 2018. **86**(6).
7. Epps, S.V.R., et al., *Foodborne Campylobacter: infections, metabolism, pathogenesis and reservoirs.* International journal of environmental research and public health, 2013. **10**(12): p. 6292-6304.
8. Platts-Mills, J.A. and M. Kosek, *Update on the burden of Campylobacter in developing countries.* Current opinion in infectious diseases, 2014. **27**(5): p. 444-450.
9. Kaakoush, N.O., et al., *Global Epidemiology of Campylobacter Infection.* Clinical Microbiology Reviews, 2015. **28**(3): p. 687-720.
10. Hofreuter, D., *Defining the metabolic requirements for the growth and colonization capacity of Campylobacter jejuni.* Frontiers in cellular and infection microbiology, 2014. **4**: p. 137-137.
11. Gillespie, I.A., et al., *A case-case comparison of Campylobacter coli and Campylobacter jejuni infection: a tool for generating hypotheses.* Emerging infectious diseases, 2002. **8**(9): p. 937-942.
12. Kothary, M.H. and U.S. Babu, *INFECTIVE DOSE OF FOODBORNE PATHOGENS IN VOLUNTEERS: A REVIEW.* Journal of Food Safety, 2001. **21**(1): p. 49-68.
13. CDC, *National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Surveillance Report for 2015*

(Final Report). Atlanta, Georgia: U.S. Department of Health and Human Services, 2018.
14. Hampton, T., *Report reveals scope of us antibiotic resistance threat.* JAMA, 2013. **310**(16): p. 1661-1663.
15. CDC, *National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human Isolates Surveillance Report for 2015*

(Final Report). Atlanta, Georgia: U.S. Department of Health and Human Services, 2018.

16. Liu, F., et al., *The Clinical Importance of Campylobacter concisus and Other Human Hosted Campylobacter Species.* Frontiers in cellular and infection microbiology, 2018. **8**: p. 243-243.
17. François, R., et al., *The other Campylobacters: Not innocent bystanders in endemic diarrhea and dysentery in children in low-income settings.* PLoS Neglected Tropical Diseases, 2018. **12**(2): p. e0006200.
18. Lastovica, A.J., *Emerging Campylobacter spp.: The tip of the iceberg.* Clinical Microbiology Newsletter, 2006. **28**(7): p. 49-56.
19. Couturier, B.A., D.C. Hale, and M.R. Couturier, *Association of Campylobacter upsaliensis with Persistent Bloody Diarrhea.* Journal of Clinical Microbiology, 2012. **50**(11): p. 3792.
20. Bullman, S., et al., *Molecular-based detection of non-culturable and emerging campylobacteria in patients presenting with gastroenteritis.* Epidemiology and Infection, 2012. **140**(4): p. 684-688.
21. Bullman, S., et al., *Emerging dynamics of human campylobacteriosis in Southern Ireland.* FEMS Immunology & Medical Microbiology, 2011. **63**(2): p. 248-253.
22. Inglis, G.D., V.F. Boras, and A. Houde, *Enteric Campylobacteria and RNA Viruses Associated with Healthy and Diarrheic Humans in the Chinook Health Region of Southwestern Alberta, Canada.* Journal of Clinical Microbiology, 2011. **49**(1): p. 209.
23. Chaban, B., M. Ngeleka, and J.E. Hill, *Detection and quantification of 14 Campylobacter species in pet dogs reveals an increase in species richness in feces of diarrheic animals.* BMC microbiology, 2010. **10**: p. 73-73.
24. Goossens, H., et al., *Campylobacter upsaliensis enteritis associated with canine infections.* The Lancet, 1991. **337**(8755): p. 1486-1487.
25. Gurgan, T. and K.S. Diker, *Abortion associated with Campylobacter upsaliensis.* Journal of Clinical Microbiology, 1994. **32**(12): p. 3093.
26. Kalischuk, L.D. and G.D. Inglis, *Comparative genotypic and pathogenic examination of Campylobacter concisus isolates from diarrheic and non-diarrheic humans.* BMC microbiology, 2011. **11**: p. 53-53.
27. Akar, M., et al., *The possible relationship between Campylobacter spp./Arcobacter spp. and patients with ulcerative colitis.* European Journal of Gastroenterology & Hepatology, 2018. **30**(5).
28. O'Donovan, D., et al., *Campylobacter ureolyticus: a portrait of the pathogen.* Virulence, 2014. **5**(4): p. 498-506.
29. Wilkinson, D.A., et al., *Updating the genomic taxonomy and epidemiology of Campylobacter hyointestinalis.* Scientific reports, 2018. **8**(1): p. 2393-2393.
30. Zhang, L., et al., *Campylobacter concisus and inflammatory bowel disease.* World journal of gastroenterology, 2014. **20**(5): p. 1259-1267.
31. Ismail, Y., et al., *Investigation of the enteric pathogenic potential of oral Campylobacter concisus strains isolated from patients with inflammatory bowel disease.* PloS one, 2012. **7**(5): p. e38217-e38217.
32. Kaakoush, N.O. and H.M. Mitchell, *Campylobacter concisus - A new player in intestinal disease.* Frontiers in cellular and infection microbiology, 2012. **2**: p. 4-4.

33. Platts-Mills, J.A., et al., *Detection of Campylobacter in Stool and Determination of Significance by Culture, Enzyme Immunoassay, and PCR in Developing Countries.* Journal of Clinical Microbiology, 2014. **52**(4): p. 1074-1080.
34. M'Ikanatha, N.M., et al., *Culturing stool specimens for Campylobacter spp., Pennsylvania, USA.* Emerging infectious diseases, 2012. **18**(3): p. 484-487.
35. Buss, J.E., et al., *Campylobacter culture fails to correctly detect Campylobacter in 30% of positive patient stool specimens compared to non-cultural methods.* European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2019. **38**(6): p. 1087-1093.
36. Wang, W.L., et al., *Evaluation of transport media for Campylobacter jejuni in human fecal specimens.* Journal of clinical microbiology, 1983. **18**(4): p. 803-807.
37. Ghosh, R., et al., *A Comparative Study of Conventional and Molecular Techniques in Diagnosis of Campylobacter Gastroenteritis in Children.* Annals of Clinical & Laboratory Science, 2014. **44**(1): p. 42-48.
38. Ricke, S.C., et al., *Developments in Rapid Detection Methods for the Detection of Foodborne Campylobacter in the United States.* Frontiers in microbiology, 2019. **9**: p. 3280-3280.
39. Granato, P.A., et al., *Comparison of premier CAMPY enzyme immunoassay (EIA), ProSpecT Campylobacter EIA, and ImmunoCard STAT! CAMPY tests with culture for laboratory diagnosis of Campylobacter enteric infections.* Journal of clinical microbiology, 2010. **48**(11): p. 4022-4027.
40. Stahl, M., J. Butcher, and A. Stintzi, *Nutrient Acquisition and Metabolism by Campylobacter jejuni.* Frontiers in Cellular and Infection Microbiology, 2012. **2**: p. 5.
41. Hazeleger, W.C., et al., *Physiological activity of Campylobacter jejuni far below the minimal growth temperature.* Applied and environmental microbiology, 1998. **64**(10): p. 3917-3922.
42. Trudy M. Wassenaara, M.J.B., *Pathophysiology of Campylobacter jejuni infections of humans.* Microbes and Infection, 1999.
43. Velayudhan, J. and D.J. Kelly, *Analysis of gluconeogenic and anaplerotic enzymes in Campylobacter jejuni: an essential role for phosphoenolpyruvate carboxykinase.* Microbiology, 2002. **148**(3): p. 685-694.
44. Stahl, M., et al., *L-fucose utilization provides Campylobacter jejuni with a competitive advantage.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(17): p. 7194-7199.
45. Luethy, P.M., et al., *Microbiota-Derived Short-Chain Fatty Acids Modulate Expression of Campylobacter jejuni Determinants Required for Commensalism and Virulence.* mBio, 2017. **8**(3): p. e00407-17.
46. Crofts, A.A., et al., *Campylobacter jejuni transcriptional and genetic adaptation during human infection.* Nature microbiology, 2018. **3**(4): p. 494-502.
47. Aidley, J., et al., *PhasomeIt: an 'omics' approach to cataloguing the potential breadth of phase variation in the genus Campylobacter.* Microbial genomics, 2018. **4**(11): p. e000228.

48. Brenda Wilson, A.S., Dixie Whitt, Malcolm Winkler, *Bacterial Strategies for Evading or Surviving Host Defense Systems*, in *Bacterial Pathogenesis: A Molecular Approach*. 2011, ASM Press: Washington, DCC.
49. Johnson, J.G., J.A. Gaddy, and V.J. DiRita, *The PAS Domain-Containing Protein HeuR Regulates Heme Uptake in Campylobacter jejuni.* mBio, 2016. **7**(6): p. e01691-16.
50. Belkaid, Y. and O.J. Harrison, *Homeostatic Immunity and the Microbiota.* Immunity, 2017. **46**(4): p. 562-576.
51. Litvak, Y., M.X. Byndloss, and A.J. Bäumler, *Colonocyte metabolism shapes the gut microbiota.* Science (New York, N.Y.), 2018. **362**(6418): p. eaat9076.
52. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* Science (New York, N.Y.), 2005. **308**(5728): p. 1635-1638.
53. Foster, K.R., et al., *The evolution of the host microbiome as an ecosystem on a leash.* Nature, 2017. **548**(7665): p. 43-51.
54. Donohoe, Dallas R., et al., *The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon.* Cell Metabolism, 2011. **13**(5): p. 517-526.
55. Furusawa, Y., et al., *Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells.* Nature, 2013. **504**: p. 446.
56. Olsan, E.E., et al., *Colonization resistance: The deconvolution of a complex trait.* The Journal of biological chemistry, 2017. **292**(21): p. 8577-8581.
57. Kelly, C.J., et al., *Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function.* Cell host & microbe, 2015. **17**(5): p. 662-671.
58. Barker, N., M. van de Wetering, and H. Clevers, *The intestinal stem cell.* Genes & development, 2008. **22**(14): p. 1856-1864.
59. Litvak, Y., et al., *Dysbiotic Proteobacteria expansion: a microbial signature of epithelial dysfunction.* Current Opinion in Microbiology, 2017. **39**: p. 1-6.
60. Lefebvre, M., et al., *Peroxisome proliferator-activated receptor gamma is induced during differentiation of colon epithelium cells.* Journal of Endocrinology, 1999. **162**(3): p. 331-340.
61. Byndloss, M.X., et al., *Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion.* Science (New York, N.Y.), 2017. **357**(6351): p. 570-575.
62. Duszka, K.O., Matej; Le May, Cedric; König, Jürgen; Wahli, Walter. , *PPARγ Modulates Long Chain Fatty Acid Processing in the Intestinal Epithelium.* International Journal of Molecular Sciences, 2017. **18**(12).
63. Furuta, G.T., et al., *Hypoxia-inducible factor 1-dependent induction of intestinal trefoil factor protects barrier function during hypoxia.* The Journal of experimental medicine, 2001. **193**(9): p. 1027-1034.
64. Roediger, W.E., *Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man.* Gut, 1980. **21**(9): p. 793-798.
65. Liu, H., et al., *Butyrate: A Double-Edged Sword for Health?* Advances in nutrition (Bethesda, Md.), 2018. **9**(1): p. 21-29.

66. Tylichová, Z., et al., *Activation of autophagy and PPARγ protect colon cancer cells against apoptosis induced by interactive effects of butyrate and DHA in a cell type-dependent manner: The role of cell differentiation.* The Journal of Nutritional Biochemistry, 2017. **39**: p. 145-155.
67. Byndloss, M.X. and A.J. Bäumler, *The germ-organ theory of non-communicable diseases.* Nature Reviews Microbiology, 2018. **16**: p. 103.
68. Shin, N.-R., T.W. Whon, and J.-W. Bae, *Proteobacteria: microbial signature of dysbiosis in gut microbiota.* Trends in Biotechnology, 2015. **33**(9): p. 496-503.
69. Spiga, L., et al., *An Oxidative Central Metabolism Enables Salmonella to Utilize Microbiota-Derived Succinate.* Cell host & microbe, 2017. **22**(3): p. 291-301.e6.
70. Vollaard, E.J., H.A.L. Clasener, and A.J.H.M. Janssen, *Co-trimoxazole impairs colonization resistance in healthy volunteers.* Journal of Antimicrobial Chemotherapy, 1992. **30**(5): p. 685-691.
71. Martinez-Medina, M., et al., *Western diet induces dysbiosis with increased E coli in CEABAC10 mice, alters host barrier function favouring AIEC colonisation.* Gut, 2014. **63**(1): p. 116.
72. Devkota, S., et al., *Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice.* Nature, 2012. **487**(7405): p. 104-108.
73. Arthur, J.C., et al., *Intestinal inflammation targets cancer-inducing activity of the microbiota.* Science (New York, N.Y.), 2012. **338**(6103): p. 120-123.
74. Rigottier-Gois, L., *Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis.* The ISME journal, 2013. **7**(7): p. 1256-1261.
75. Carroll, I.M., et al., *Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome.* Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society, 2012. **24**(6): p. 521-e248.
76. Normann, E., et al., *Intestinal microbial profiles in extremely preterm infants with and without necrotizing enterocolitis.* Acta Paediatrica, 2013. **102**(2): p. 129-136.
77. Atarashi, K., et al., *Induction of colonic regulatory T cells by indigenous Clostridium species.* Science (New York, N.Y.), 2011. **331**(6015): p. 337-341.
78. Arpaia, N., et al., *Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation.* Nature, 2013. **504**(7480): p. 451-455.
79. Alex, S., et al., *Short-chain fatty acids stimulate angiopoietin-like 4 synthesis in human colon adenocarcinoma cells by activating peroxisome proliferator-activated receptor γ.* Molecular and cellular biology, 2013. **33**(7): p. 1303-1316.
80. Smith, P.M., et al., *The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis.* Science (New York, N.Y.), 2013. **341**(6145): p. 569-573.
81. Gillis, C.C., et al., *Dysbiosis-Associated Change in Host Metabolism Generates Lactate to Support Salmonella Growth.* Cell host & microbe, 2018. **23**(1): p. 54-64.e6.
82. Reese, A.T., et al., *Antibiotic-induced changes in the microbiota disrupt redox dynamics in the gut.* eLife, 2018. **7**: p. e35987.

83. Spees, A.M., et al., *Streptomycin-induced inflammation enhances Escherichia coli gut colonization through nitrate respiration.* mBio, 2013. **4**(4): p. e00430-13.
84. Dubuquoy, L., et al., *Impaired expression of peroxisome proliferator-activated receptor γ in ulcerative colitis.* Gastroenterology, 2003. **124**(5): p. 1265-1276.
85. Ponferrada, Á., et al., *The Role of PPARγ on Restoration of Colonic Homeostasis After Experimental Stress-Induced Inflammation and Dysfunction.* Gastroenterology, 2007. **132**(5): p. 1791-1803.
86. Ogasawara, N., et al., *PPARγ agonists upregulate the barrier function of tight junctions via a PKC pathway in human nasal epithelial cells.* Pharmacological Research, 2010. **61**(6): p. 489-498.
87. Kamada, N., et al., *Regulated virulence controls the ability of a pathogen to compete with the gut microbiota.* Science (New York, N.Y.), 2012. **336**(6086): p. 1325-1329.
88. Rivera-Chávez, F., et al., *Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella.* Cell host & microbe, 2016. **19**(4): p. 443-454.
89. Stecher, B., et al., *Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota.* PLoS biology, 2007. **5**(10): p. 2177-2189.
90. Winter, S.E., et al., *Gut inflammation provides a respiratory electron acceptor for Salmonella.* Nature, 2010. **467**(7314): p. 426-429.
91. Collins, J.W., et al., *Citrobacter rodentium: infection, inflammation and the microbiota.* Nature Reviews Microbiology, 2014. **12**: p. 612.
92. Agbessi Amouzou, L.C.V., *One is too many: Ending child deaths from pneumonia and diarrhea*. 2016, UNICEF: New York.
93. Syed, S., A. Ali, and C. Duggan, *Environmental Enteric Dysfunction in Children: A Review.* Journal of pediatric gastroenterology and nutrition, 2016. **63**(1): p. 6-14.
94. Oriá, R.B., et al., *Early-life enteric infections: relation between chronic systemic inflammation and poor cognition in children.* Nutrition reviews, 2016. **74**(6): p. 374-386.
95. Dewey, K.G. and K. Begum, *Long-term consequences of stunting in early life.* Maternal & Child Nutrition, 2011. **7**(s3): p. 5-18.
96. Lee, G., et al., *Symptomatic and Asymptomatic Campylobacter Infections Associated with Reduced Growth in Peruvian Children.* PLoS Neglected Tropical Diseases, 2013. **7**(1): p. e2036.
97. Kotloff, K.L., et al., *Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study.* The Lancet, 2013. **382**(9888): p. 209-222.
98. ASeidmanJessica, M.-E.N.I.A.M.b.R.B.T.P.R.S.R.O.A.K.J.T.S.P.V.L.S.J.W.K.D.R.J.A., *Relationship between growth and illness, enteropathogens and dietary intakes in the first 2*

years of life: findings from the MAL-ED birth cohort study. BMJ Global Health, 2017. **2**(4): p. e000370.
99. Jones, P.C., et al., *Measuring home environments across cultures: Invariance of the HOME scale across eight international sites from the MAL-ED study().* Journal of School Psychology, 2017. **64**: p. 109-127.
100. Amour, C., et al., *Epidemiology and Impact of Campylobacter Infection in Children in 8 Low-Resource Settings: Results From the MAL-ED Study.* Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America, 2016. **63**(9): p. 1171-1179.
101. Round, J.L. and S.K. Mazmanian, *The gut microbiome shapes intestinal immune responses during health and disease.* Nature reviews. Immunology, 2009. **9**(5): p. 313-323.
102. Bell, J.A. and D.D. Manning, *A domestic ferret model of immunity to Campylobacter jejuni-induced enteric disease.* Infection and Immunity, 1990. **58**(6): p. 1848.
103. Farfán-García AE, Z.C., Imdad A, , *Case-Control Pilot Study on Acute Diarrheal Disease in a Geographically Defined Pediatric Population in a Middle Income Country.* International Journal of Pediatrics, 2017. **2017**: p. 10.
104. Gómez-Duarte, O.G., et al., *Detection of Escherichia coli enteropathogens by multiplex polymerase chain reaction from children's diarrheal stools in two Caribbean-Colombian cities.* Foodborne pathogens and disease, 2010. **7**(2): p. 199-206.
105. Checkley, W., et al., *Multi-country analysis of the effects of diarrhoea on childhood stunting.* International Journal of Epidemiology, 2008. **37**(4): p. 816-830.
106. Ma, S., et al., *Organization of the mammalian ionome according to organ origin, lineage specialization and longevity.* Cell reports, 2015. **13**(7): p. 1319-1326.
107. Kozich, J.J., et al., *Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.* Applied and Environmental Microbiology, 2013. **79**(17): p. 5112.
108. Dhariwal, A., et al., *MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data.* Nucleic Acids Research, 2017. **45**(W1): p. W180-W188.
109. Segata, N., et al., *Metagenomic biomarker discovery and explanation.* Genome Biology, 2011. **12**(6): p. R60.
110. Chong, J., et al., *MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis.* Nucleic Acids Research, 2018. **46**(W1): p. W486-W494.
111. Bruce, B.a., *Practical Statistics for Data Scientists*. 2017, Sebastopol, CA: O'Reilly Media Inc.
112. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.

113. Lê Cao, K.-A., S. Boitard, and P. Besse, *Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems.* BMC bioinformatics, 2011. **12**: p. 253-253.
114. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.
115. Zierer, J., et al., *The fecal metabolome as a functional readout of the gut microbiome.* Nature Genetics, 2018. **50**(6): p. 790-795.
116. Sandberg, M., et al., *Risk factors for Campylobacter infection in Norwegian cats and dogs.* Preventive Veterinary Medicine, 2002. **55**(4): p. 241-253.
117. Acke, E., *Campylobacteriosis in dogs and cats: a review.* New Zealand Veterinary Journal, 2018. **66**(5): p. 221-228.
118. Torkan, S., et al., *Prevalence of thermotolerant Campylobacter species in dogs and cats in Iran.* Veterinary medicine and science, 2018. **4**(4): p. 296-303.
119. Bank, W., *Poverty and Shared Prosperity 2018 : Piecing Together the Poverty Puzzle*. 2018, Washington, DC: World Bank.
120. !!! INVALID CITATION !!! {Subramanian, 2015 #117;Stanislawski, 2018 #345;Robertson, 2019 #334}.
121. Ihekweazu, F.D. and J. Versalovic, *Development of the Pediatric Gut Microbiome: Impact on Health and Disease.* The American Journal of the Medical Sciences, 2018. **356**(5): p. 413-423.
122. Robertson, R.C., et al., *The Human Microbiome and Child Growth – First 1000 Days and Beyond.* Trends in Microbiology, 2019. **27**(2): p. 131-147.
123. Winter, S.E. and A.J. Bäumler, *Dysbiosis in the inflamed intestine: chance favors the prepared microbe.* Gut microbes, 2014. **5**(1): p. 71-73.
124. Bäumler, A.J. and V. Sperandio, *Interactions between the microbiota and pathogenic bacteria in the gut.* Nature, 2016. **535**(7610): p. 85-93.
125. Faber, F. and A.J. Bäumler, *The impact of intestinal inflammation on the nutritional environment of the gut microbiota.* Immunology letters, 2014. **162**(2 Pt A): p. 48-53.
126. Guccione, E.J., et al., *Transcriptome and proteome dynamics in chemostat culture reveal how Campylobacter jejuni modulates metabolism, stress responses and virulence factors upon changes in oxygen availability.* Environmental microbiology, 2017. **19**(10): p. 4326-4348.
127. Dinh, D.M., et al., *Longitudinal Analysis of the Intestinal Microbiota in Persistently Stunted Young Children in South India.* PLoS ONE, 2016. **11**(5): p. e0155405.
128. Agarwal, A., et al., *Scurvy in pediatric age group – A disease often forgotten?* Journal of Clinical Orthopaedics and Trauma, 2015. **6**(2): p. 101-107.
129. Traber, M.G., G.R. Buettner, and R.S. Bruno, *The relationship between vitamin C status, the gut-liver axis, and metabolic syndrome.* Redox biology, 2018. **21**: p. 101091-101091.
130. McCormick, B.J.J., et al., *Dynamics and Trends in Fecal Biomarkers of Gut Function in Children from 1–24 Months in the MAL-ED Study.* The American Journal of Tropical Medicine and Hygiene, 2017. **96**(2): p. 465-472.
131. Casellas, F., et al., *Bile acid induced colonic irritation stimulates intracolonic nitric oxide release in humans.* Gut, 1996. **38**(5): p. 719-723.

132. Liu, S.P., et al., *Metabolic engineering of Escherichia coli for the production of phenylpyruvate derivatives.* Metabolic Engineering, 2015. **32**: p. 55-65.
133. van Vliet, D., et al., *Can untreated PKU patients escape from intellectual disability? A systematic review.* Orphanet journal of rare diseases, 2018. **13**(1): p. 149-149.
134. Carr, A.C. and S. Maggini, *Vitamin C and Immune Function.* Nutrients, 2017. **9**(11): p. 1211.
135. Lopez, C.A., et al., *Collateral damage: microbiota-derived metabolites and immune function in the antibiotic era.* Cell host & microbe, 2014. **16**(2): p. 156-163.
136. Hibberd, M.C., et al., *The effects of micronutrient deficiencies on bacterial species from the human gut microbiota.* Science translational medicine, 2017. **9**(390): p. eaal4069.
137. Nishito, Y. and T. Kambe, *Absorption Mechanisms of Iron, Copper, and Zinc: An Overview.* Journal of Nutritional Science and Vitaminology, 2018. **64**(1): p. 1-7.
138. Rosenthal, J., A. Angel, and J. Farkas, *Metabolic fate of leucine: a significant sterol precursor in adipose tissue and muscle.* American Journal of Physiology-Legacy Content, 1974. **226**(2): p. 411-418.
139. Berg JM, T.J., Stryer L., *Biochemistry 5th edition*. Section 25.1, In de Novo Synthesis, the Pyrimidine Ring Is Assembled from Bicarbonate, Aspartate, and Glutamine. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22447/. 2002, New York: W H Freeman.
140. McDonald, T., et al., *Creatinine inhibits bacterial replication.* The Journal Of Antibiotics, 2012. **65**: p. 153.
141. Agus, A., J. Planchais, and H. Sokol, *Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease.* Cell Host & Microbe, 2018. **23**(6): p. 716-724.
142. Giallourou, N., et al., *A novel mouse model of Campylobacter jejuni enteropathy and diarrhea.* PLOS Pathogens, 2018. **14**(3): p. e1007083.
143. Crane, R.J., K.D.J. Jones, and J.A. Berkley, *Environmental enteric dysfunction: an overview.* Food and nutrition bulletin, 2015. **36**(1 Suppl): p. S76-S87.
144. Oski, F.A., *Iron Deficiency in Infancy and Childhood.* New England Journal of Medicine, 1993. **329**(3): p. 190-193.
145. Matak, P., et al., *Copper deficiency leads to anemia, duodenal hypoxia, upregulation of HIF-2α and altered expression of iron absorption genes in mice.* PloS one, 2013. **8**(3): p. e59538-e59538.
146. Scallan, E., et al., *Foodborne illness acquired in the United States--major pathogens.* Emerging infectious diseases, 2011. **17**(1): p. 7-15.
147. Lehmann, F.S., E. Burri, and C. Beglinger, *The role and utility of faecal markers in inflammatory bowel disease.* Therapeutic Advances in Gastroenterology, 2015. **8**(1): p. 23-36.

APPENDICES

Appendix 1

Table 1. Inclusion and exclusion criteria for cohorts

Inclusion	Children less than 60 months of age	
	Child who resides within the metropolitan area of Bucaramanga, Colombia	
	Presence or absence of acute, moderate-to-severe *****diarrhea and/or vomiting within the past 10 days.	
	Diarrhea is moderate to severe and must meet at least one of the following criteria:	(i) Sunken eyes, confirmed by parent/caretaker
		(ii) Loss of skin turgor by skin pinch (≤2 s slow or >2 very slow)
		(iii) Intravenous rehydration prescribed or administered
		(iv) Dysentery (1 or more bloody stools)
		(v) Evaluated in the emergency department or admitted to the hospital for diarrhea
Exclusion	Children older than 60 months of age	
	Children who reside outside of the metropolitan area of Bucaramanga, Colombia	
	Presence of chronic diarrhea (>10 days) or other comorbid conditions such as Crohn's disease or ulcerative colitis	

(*) The World Health Organization defines diarrhea as 3 or more episodes of loose or liquid stools within 24 hours

Table 2. Sociodemographic of subjects

Variable	Infected (%)	Uninfected (%)	*p*-value[1]
Gender			
Male	8 (40)	6 (30)	0.204
Female	12 (60)	14 (70)	
Race			
Mestizo	14 (70)	14 (70)	1
White	6 (30)	6 (30)	
Age Group (months)			
0 to 12	5 (25)	9 (45)	0.320
12 to 24	11 (55)	4 (20)	0.048
24 to 60	4 (20)	7 (55)	0.480

[1]Two-sided *p*-values were generated using Fisher's Exact Test.

Table 3. Sociodemographic of subjects

Variable	**1 (%)**	**2 (%)**	**3 (%)**	**4 (%)**	***p*-value[1]**
Gender					0.47
Male	5 (35)	3 (50)	2 (25)	4 (33%)	
Female	9 (65)	3 (50)	6 (75)	8 (67%)	
Race					0.87
Mestizo	9 (64)	5 (83)	2 (25)	12 (100)	
White	5 (36)	1 (17)	6 (75)	0 (0)	
Age Group (months)					0.71
0 to 12	3 (21)	2 (33.3)	3 (37.5)	6 (50)	
12 to 24	9 (64)	2 (33.3)	2 (25)	2 (16)	
24 to 60	2 (14)	2 (33.3)	3 (37.5)	4 (34)	
Total Subjects	**14**	**6**	**8**	**12**	

[1]Two-sided *p*-values were generated using Kruskal-Wallis Test.

Table 4. Clinical Symptoms of cohorts

	Fever	Vomiting	Diarrhea	Antidiarrheal in last 14 days	Blood in stool	Mucus in stool	Abdominal pain
Infected (%)	14 (70)	10 (50)	14 (70)	9 (45)	4 (20)	15 (75)	6 (30)
Uninfected (%)	8 (40)	7 (35)	8 (40)	4 (20)	4 (20)	10 (50)	5 (25)
***p*-value[1]**	0.056	0.337	0.056	0.091	1	0.102	0.723

[1]Two-sided *p*-values were generated using Chi Squared Test.

Table 5. Animal and illness exposure

	Any Animal Exposure	Bird	Cat	Dog	Chicken	International Travel	Exposure to diarrhea (outside of home)	Exposure to diarrhea (inside of home)	Attends Daycare
Infected (%)	10 (50)	0 (0)	7 (35)	8 (40)	3 (15)	2 (10)	1 (5)	1 (5)	4 (20)
Uninfected (%)	8 (40)	7 (35)	2 (10)	8 (40)	2 (20)	2 (10)	0 (0)	1 (5)	5 (25)
p-value[1]	0.525	0.003	0.058	1	0.632	0.465	0.311	1	0.705

[1]Two-sided *p*-values were generated using Chi Squared Test.

Table 6. Proportion of infected subjects by either single or multiple pathogens (co-infections)

Pathogens	**Total (%)**	**Infected (%)**	**Uninfected (%)**	***p*-value[1]**
Giardia lamblia	2 (5)	2 (10)	0 (0)	0.4872
Blastocytis	2 (5)	0 (0)	2 (10)	0.4872
Campylobacter	20 (50)	20 (100)	0 (0)	<0.0001
Commensal E. coli	29 (72.5)	16 (80)	11 (55)	0.1760
Pathogenic E. coli	4 (10)	2 (10)	2 (10)	1
Norovirus	3 (7.5)	3 (15)	0 (0)	0.230
Sapovirus	1 (2.5)	1 (5)	0 (0)	1
Co-infections	6 (15)	3 (15)	1 (5)	0.6050
Negative for any pathogen	17 (42.5)	0 (0)	17 (85)	0.0001

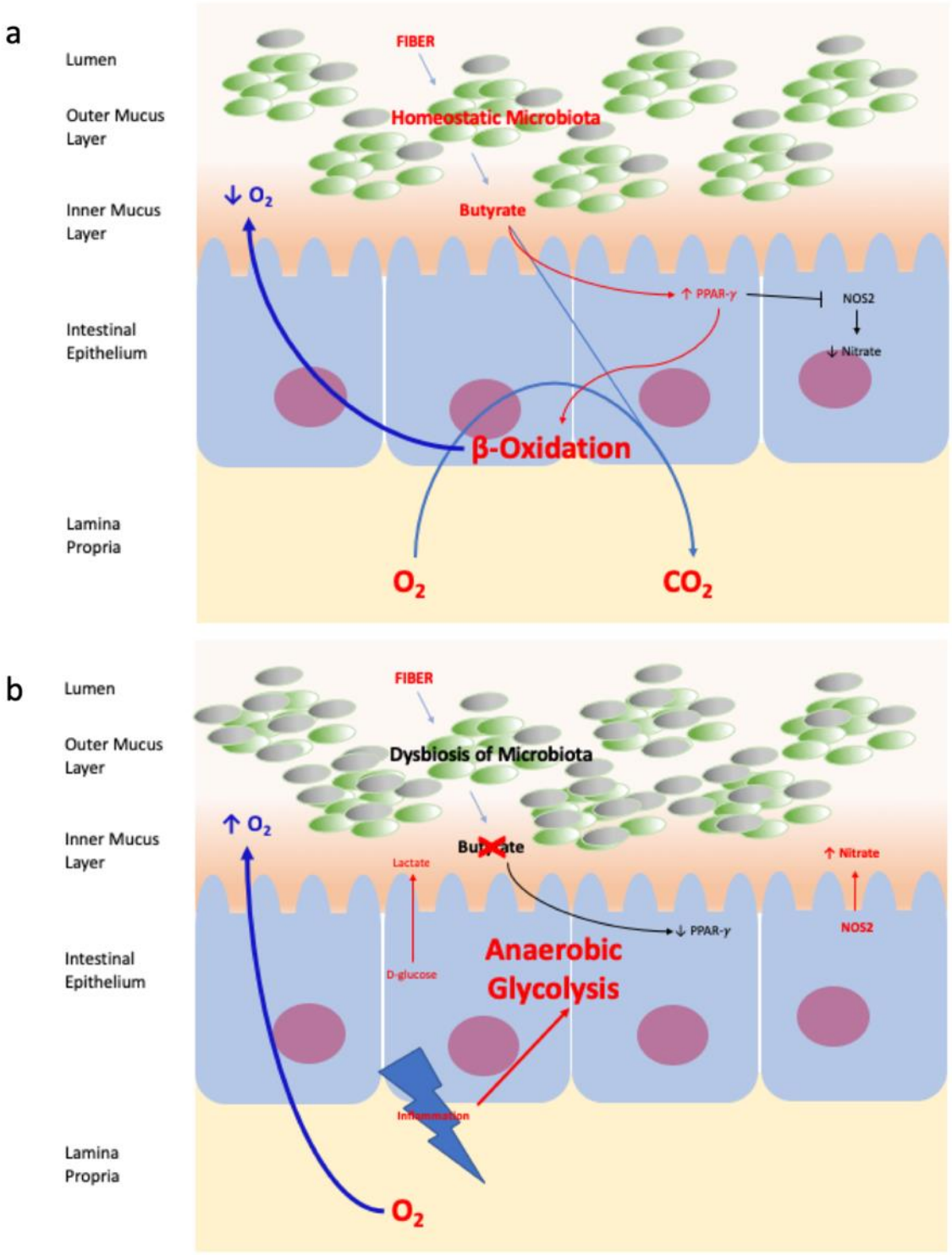
[1]Two-sided *p*-values were generated using Fisher's Exact Test

Table 7. Proportion of infected subjects by either single or multiple pathogens (co-infections)

Pathogens	**Total (%)**	**1 (%)**	**2 (%)**	**3 (%)**	**4 (%)**
Giardia lamblia	2 (5)	1	1	0	0
Blastocytis	2 (5)	0	0	1	1
Campylobacter	20 (50)	14	6	0	0
Commensal *E. coli*	29 (72.5)	11	5	5	6
Pathogenic E. coli	4 (10)	2	0	0	2
Norovirus	3 (7.5)	2	1	0	0
Sapovirus	1 (2.5)	1	0	0	0
Co-infections	7 (17.5)	5	1	0	1
Negative for any pathogen	17 (42.5)	0 (0)	0 (0)	7 (87.5)	10 (83)

Figure 1. Epithelial metabolism and the colonic microbiota.

(a) When the gut microbiota is in homeostasis, the obligate anaerobic microbes will metabolize fiber from the diet to produce fermentation products such as butyrate that will help maintain the β-oxidation metabolic state of the colonocyte. This metabolic state is characterized by high oxygen consumption which allows for the maintenance of the epithelial hypoxia (<1% oxygen) to limit the amount of oxygen that diffuses into the lumen. **(b)** A dysbiosis of the microbiota would reduce the production of fermentation products such as butyrate, which leads to the colonocyte to switch from β-oxidation to anaerobic glycolysis. This metabolic shift would allow for oxygen, lactate, and nitrate concentrations in the lumen to significantly increase.



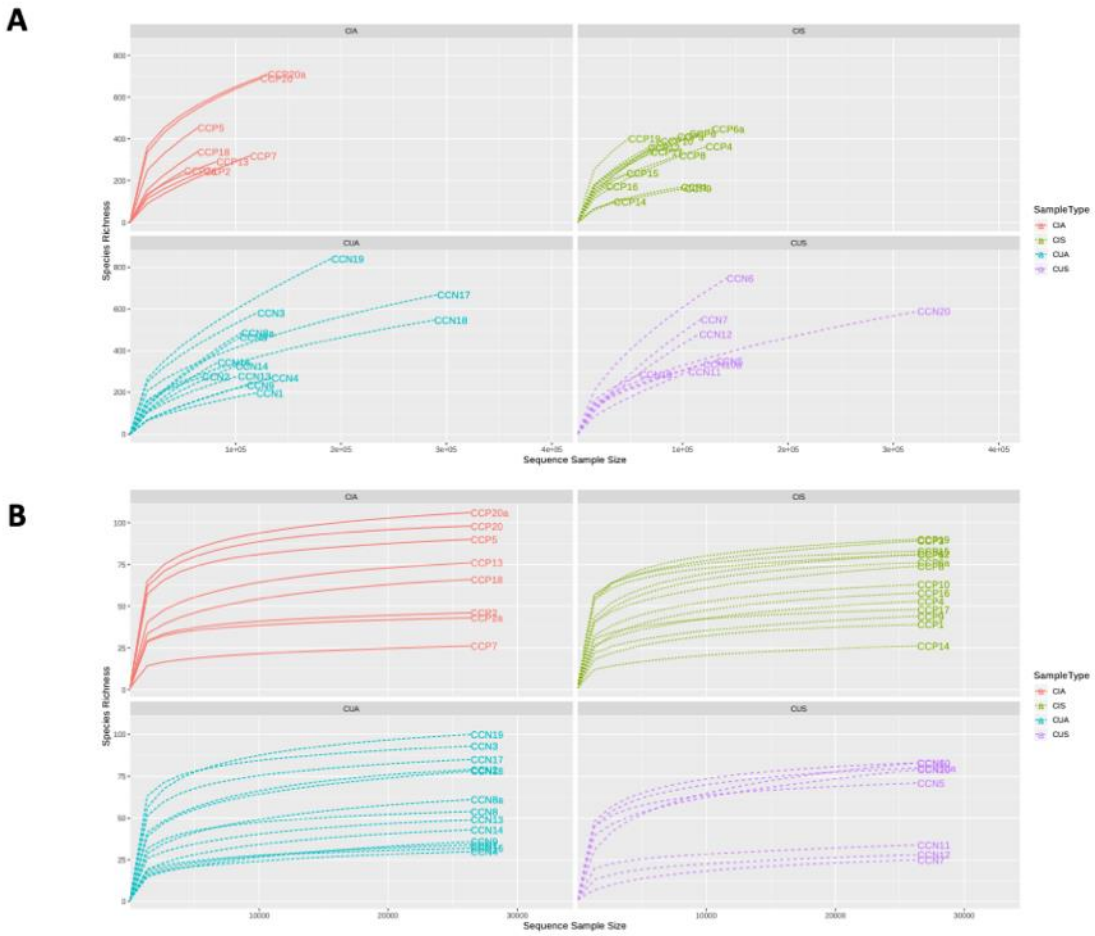


Figure 2. Rarefaction Curve of 16S rRNA Amplicons

(a) Rarefaction curve of samples before normalization. **(b)** Rarefaction curve of samples after normalization by rarefying data.

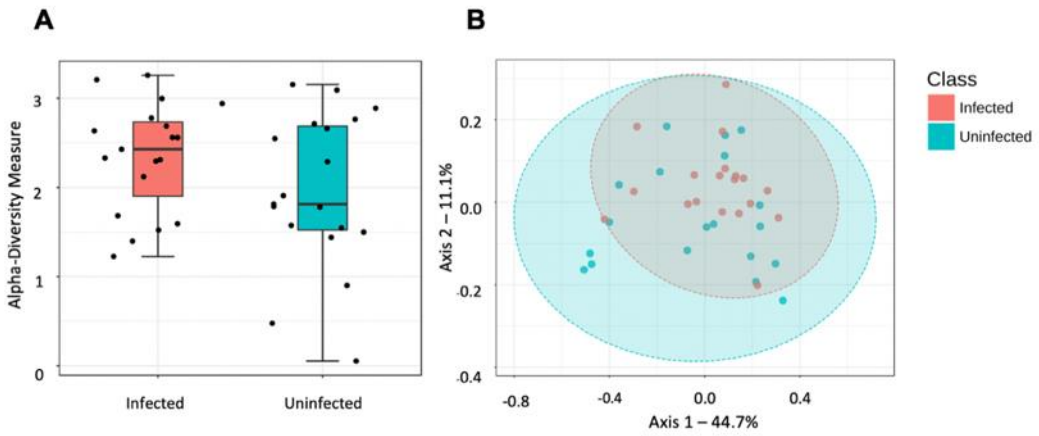


Figure 3. Bacterial Diversity Measures Between Infected and Uninfected Cohorts

(A) Alpha Diversity was calculated by Shannon index using richness and evenness to describe the actual diversity of the communities, which is represented as a box plot at the genus level. Each boxplot represents the diversity distribution present within each group (p-value = 0.18744). **(B)** Beta Diversity was calculated using the phylogenetic-based weighted unifrac distances and visualized using Principle Coordinate Analysis (PCoA) plots. The X-axis represents the percent of the variation between the samples with the X-axis representing the highest dimension of variation and the Y-axis representing the second highest dimension of variation (p-value = 0.074).

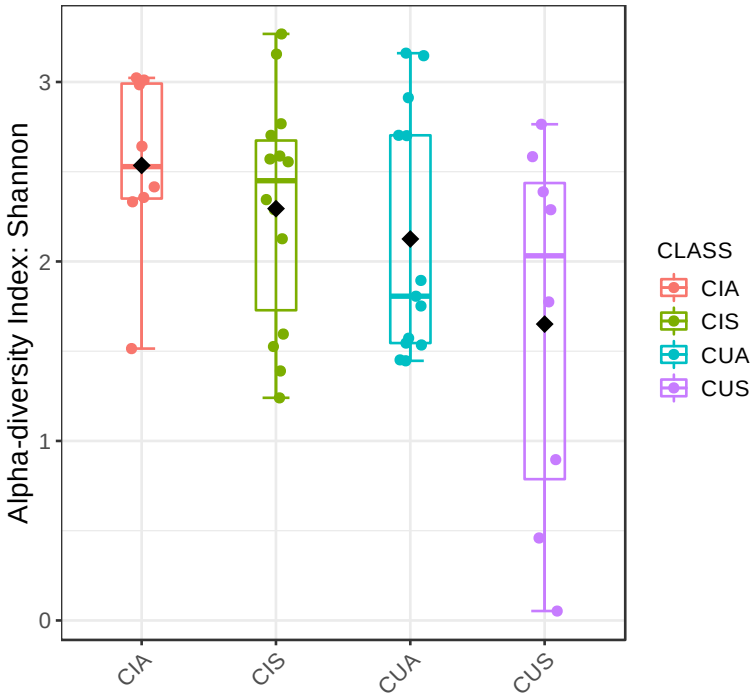


Figure 4. Alpha Diversity of the Four Cohorts

Alpha Diversity was calculated by Shannon index using richness and evenness to describe the actual diversity of the communities, which is represented as a box plot at the OTU level. Each boxplot represents the diversity distribution present within each group (p-value = 0.36257). n's of cohorts CIS, CIA, CUS, CUA are 14, 6, 8, and 12, respectively.

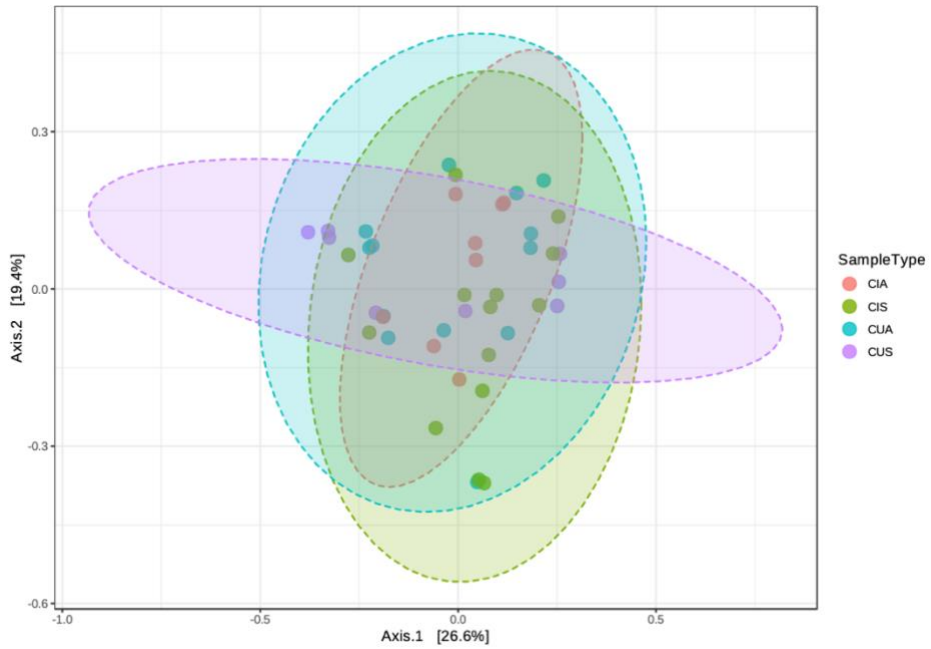


Figure 5. Beta Diversity of the Four Cohorts

Beta Diversity was calculated using the Jensen-Shannon Divergence Index and visualized using Principle Coordinate Analysis (PCoA) plots. The X-axis represents the percent of the variation between the samples with the X-axis representing the highest dimension of variation and the Y-axis representing the second highest dimension of variation (p-value = 0.121). n's of cohorts CIS, CIA, CUS, CUA are 13, 6, 8, and 12, respectively.

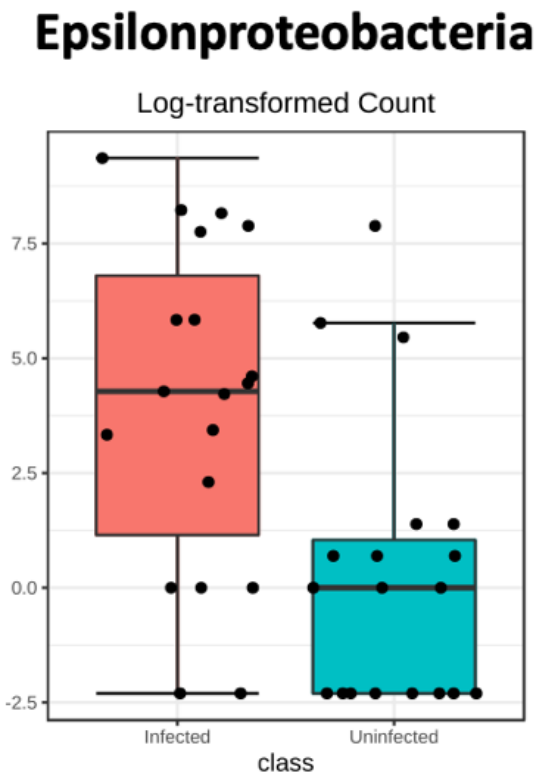


Figure 6. Univariate Analysis Between Campylobacter-Infected and Uninfected Microbiomes

The univariate analysis of *Epsilonproteobacteria* using Mann-Whitney/Kruskal-Wallis t-test resulted in a p-value < 0.001; FDR < 0.0268 (n=19 infected; n=20 uninfected).

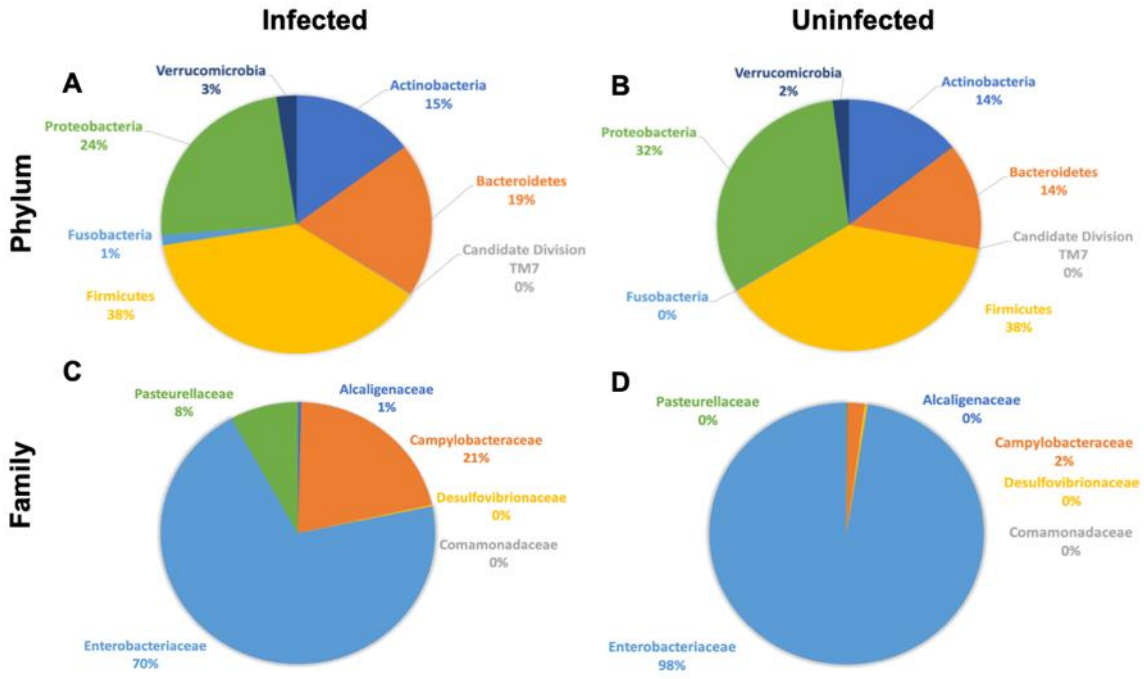


Figure 7. Microbiome Differences Between Campylobacter-infected and uninfected Samples

The combined infected **(A)** and uninfected **(B)** relative abundances found in the microbiome at the phylum level, along with the combined infected **(C)** and uninfected **(D)** relative abundances found in the microbiome at the family level are shown as a pie chart to demonstrate changes that occurred (n=19 infected; n=20 uninfected).

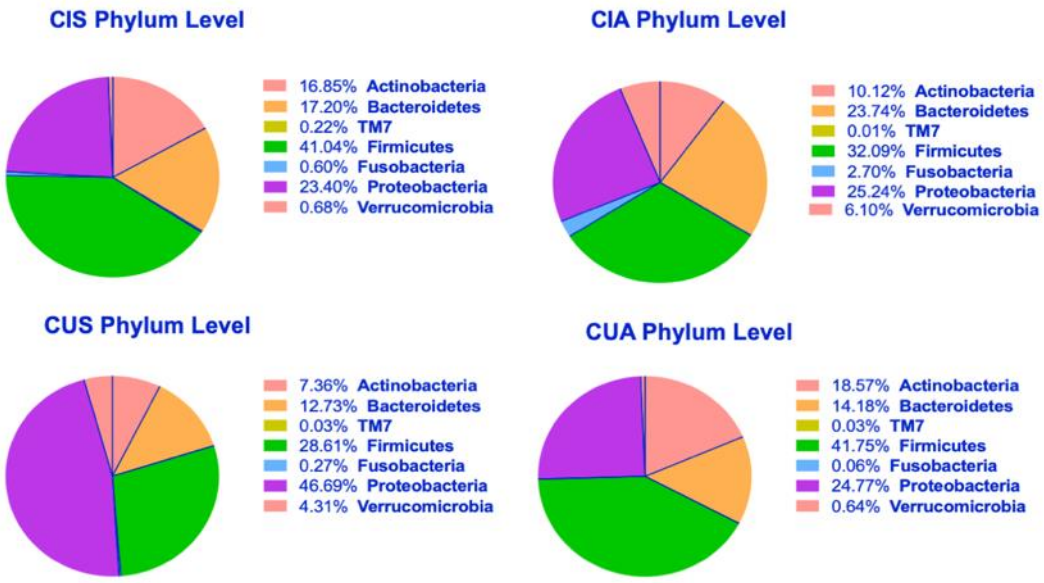


Figure 8. Phylum Level Microbiome Differences between the Four Cohorts

Relative abundances found in the microbiome at the phylum level between the four cohorts. n's of cohorts CIS, CIA, CUS, CUA are 13, 6, 8, and 12, respectively.

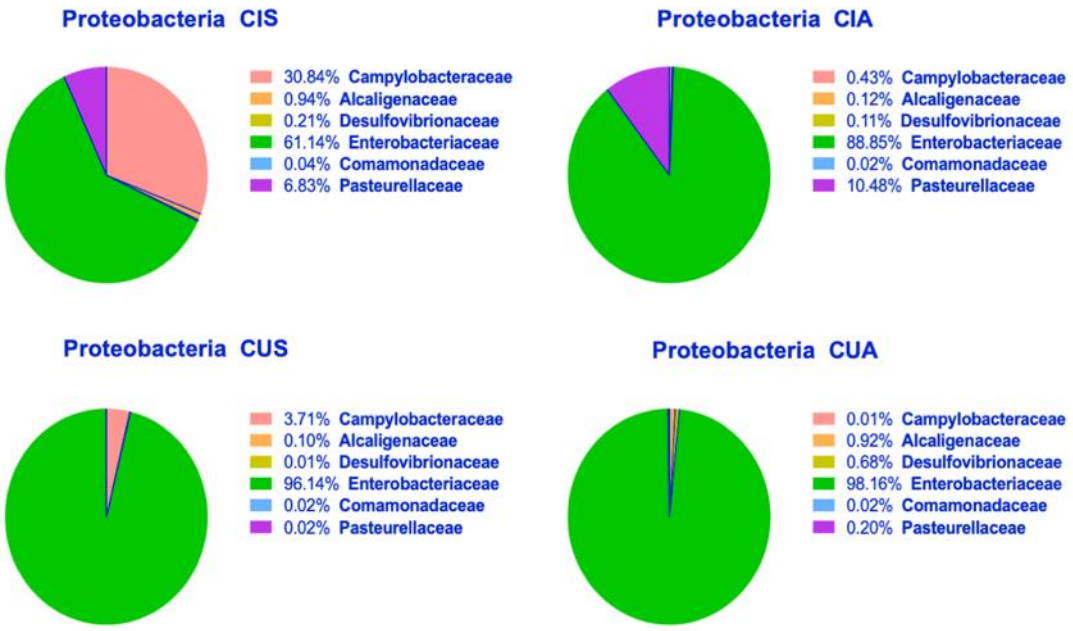


Figure 9. Proteobacteria's Family Level Microbiome Differences between the Four Cohorts

Relative abundances found in the microbiome at the Proteobacteria's family level between the four cohorts. n's of cohorts CIS, CIA, CUS, CUA are 13, 6, 8, and 12, respectively.

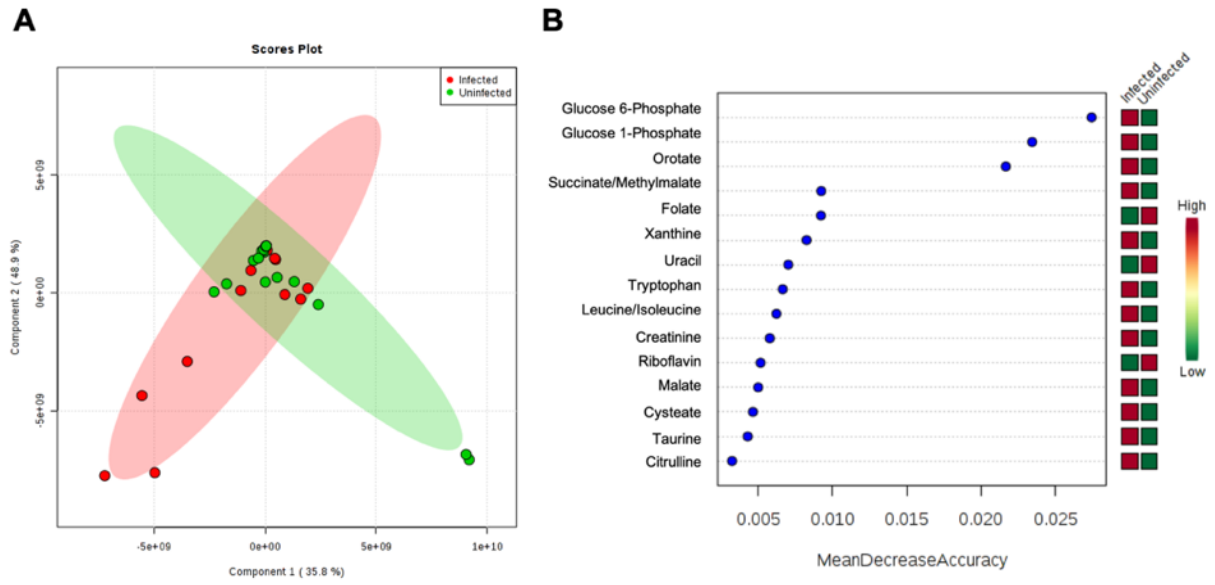


Figure 10. Metabolomic differences between infected and uninfected samples

(A) The differences between the metabolomes of the *Campylobacter* infected and uninfected samples were determined by using the Least Partial Squares – Discriminant Analysis (PLS-DA). The visualization of these differences is plotted 2-dimensionally with each point representing the metabolomic profile of a single sample (N=40, 20 per group). **(B)** The top 15 metabolites determined by Random Forest analysis to have the highest discriminatory power between the two groups with red fields representing high abundance and green fields representing low abundance for each metabolite per infection status.

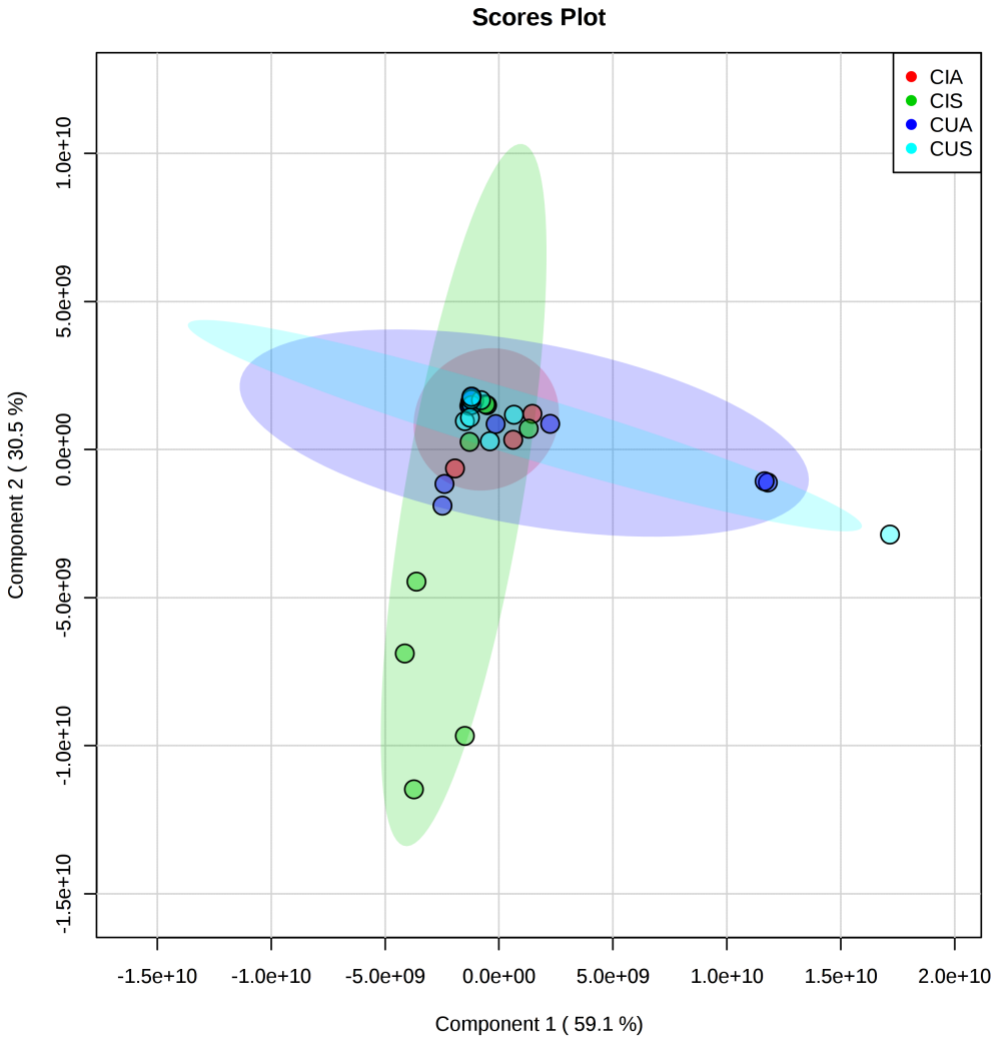


Figure 11. PLS-DA Plot of the Four Cohorts

The differences between the metabolomes of the four cohorts' fecal samples were determined by using the Least Partial Squares – Discriminant Analysis (PLS-DA). The visualization of these differences is plotted 2-dimensionally with each point representing the metabolomic profile of a single sample. n's of cohorts CIS, CIA, CUS, CUA are 14, 6, 8, and 12, respectively.

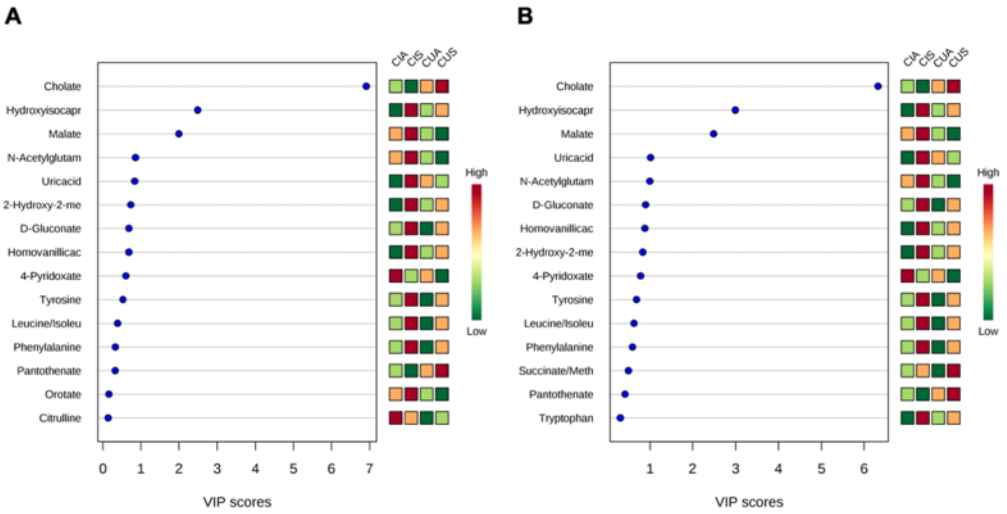


Figure 12. PLS-DA VIP Scores

The differences between the metabolomes of the four cohorts' fecal samples were determined by using the Least Partial Squares – Discriminant Analysis (PLS-DA). **(a)** Metabolite VIP scores from Component 1. **(b)** Metabolite VIP scores from component 2. n's of cohorts CIS, CIA, CUS, CUA are 14, 6, 8, and 12, respectively.

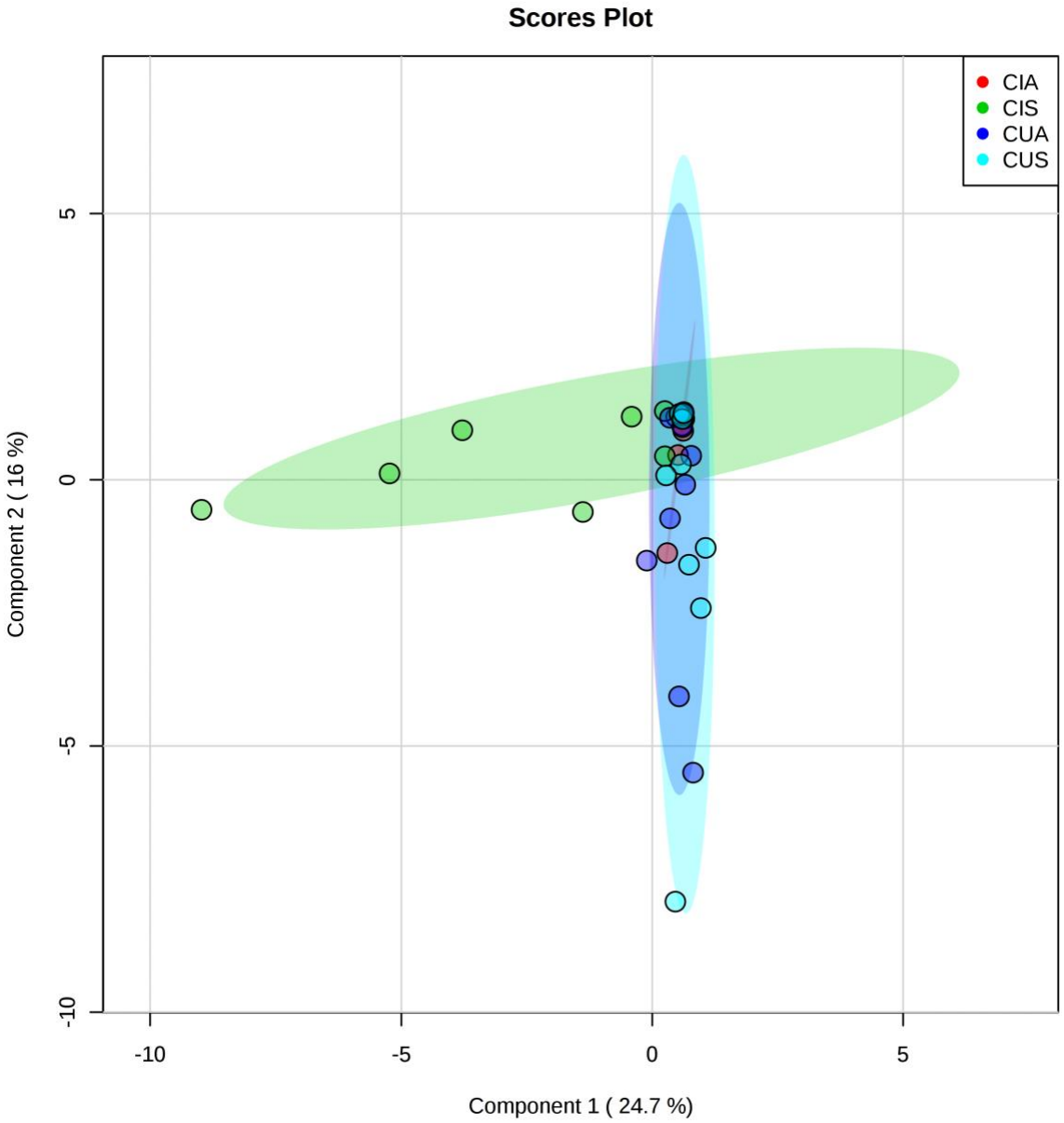


Figure 13. sPLS-DA Plot of Four Cohorts

The differences between the metabolomes of the four cohorts' fecal samples were determined by using the Sparse Least Partial Squares – Discriminant Analysis (sPLS-DA).

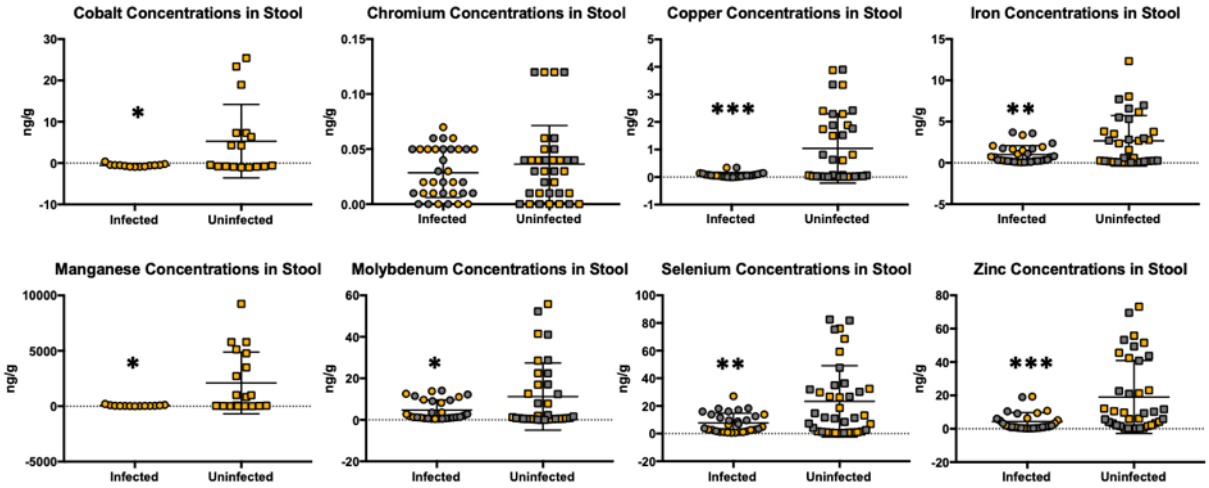


Figure 14. Micronutrient differences between infected and uninfected samples

Different colored data points were used to represent the different isotopes measured by the ICP-MS (orange represents ^{59}Co, ^{52}Cr, ^{63}Cu, ^{56}Fe, ^{55}Mn, ^{95}Mo, ^{78}Se, and ^{66}Zn; grey represents ^{53}Cr, ^{65}Cu, ^{57}Fe, ^{98}Mo, ^{82}Se, and ^{64}Zn). The asterisks represent significance calculated by unpaired t-tests (*$p < 0.05$, **$p < 0.005$, **$p < 0.001$; n=17 infected, n=19 uninfected).

Appendix 2

% Abundances at Phylum Level

	1	2	3	4
Actinobacteria	16.8	10.1	7.4	18.6
Bacteroidetes	17.2	23.7	12.7	14.2
Candidate_division_TM7	0.2	6.3	0.0	0.0
Firmicutes	41.0	32.1	28.6	41.7
Fusobacteria	0.6	2.7	0.3	0.0
Proteobacteria	23.4	25.2	46.7	24.8
Verrucomicrobia	0.7	6.1	4.3	0.6

Appendix 2- 1. Percent Abundances at Phylum Level

% Abundance at Class Level

	1	2	3	4
Actinobacteria	16.84395871	10.13251475	7.380798987	18.36889377
Bacilli	13.44280838	3.660893622	3.775668755	11.52366696
Bacteroidia	17.1748326	23.78064956	12.72975329	14.17206705
Betaproteobacteria	0.228259254	0.032844456	0.056304782	0.214752214
Candidate_division_TM7_unclassified	0.221845839	0.00757949	0.02544543	0.023054282
Clostridia	26.13670485	26.65201299	23.95606061	30.40007074
Deltaproteobacteria	0.051890354	0.026528215	0.005955313	0.157590228
Epsilonproteobacteria	7.207220339	0.109270979	1.734620403	0.004105557
Erysipelotrichi	1.557585175	1.719912583	0.850527045	1.980773361
Fusobacteria	0.608399824	2.703983022	0.268530499	0.066952161
Gammaproteobacteria	15.84783883	25.06474148	44.90089817	22.44508028
Verrucomicrobiae	0.678655865	6.10906886	4.315436714	0.642993393

Appendix 2- 2. Percent Abundances at Class Level

% abundance at Order Level

	1	2	3	4
Actinobacteridae	16.10874819	8.780207425	5.930409455	13.37085181
Bacillales	0.093577548	0.023370094	0.049808076	0.34928816
Bacteroidales	17.1748326	23.78064956	12.72975329	14.17206705
Burkholderiales	0.228259254	0.032844456	0.056304782	0.214752214
Campylobacterales	7.207220339	0.109270979	1.734620403	0.004105557
Candidate_division_TM7_unclassified	0.221845839	0.00757949	0.02544543	0.023054282
Clostridiales	26.13670485	26.65201299	23.95606061	30.40007074
Coriobacteridae	0.73521052	1.352307323	1.450389532	4.998041965
Desulfovibrionales	0.051890354	0.026528215	0.005955313	0.157590228
Enterobacteriales	14.24885797	22.53066535	44.89277729	22.39897172
Erysipelotrichales	1.557585175	1.719912583	0.850527045	1.980773361
Fusobacteriales	0.608399824	2.703983022	0.268530499	0.066952161
Lactobacillales	13.34923083	3.637523528	3.725860678	11.1743788
Pasteurellales	1.59898085	2.534076123	0.008120882	0.046108564
Verrucomicrobiales	0.678655865	6.10906886	4.315436714	0.642993393

Appendix 2- 3. Percent Abundance at Order Level

% Abundance at Family Level

	1	2	3	4
Actinomycetales	0.00261492401561375	0.00062530791677201	0.00153755366549546	0.000947436237541213
Aerococcaceae	9.62012179657232E-05	0	0	1.57906039590202E-05
Alcaligenaceae	0.002189306505826	0.000309495837596796	0.000454769394019782	0.00209699220575789
Bacillaceae	9.03708411193157E-05	2.52649663344324E-05	9.74505844328105E-05	6.31624158360809E-05
Bacteroidaceae	0.165617684699051	0.184926921084878	0.0665370934821801	0.115868293730499
Bifidobacteriales	0.158472557873778	0.0871767663369589	0.0577665408832271	0.132761081845858
Campylobacteraceae	0.0720722033868659	0.0010927097939642	0.0173462040290403	4.10555702934526E-05
Candidate_division_TM7_unclassified	0.00221845839005804	7.57948990032971E-05	0.000254454303796783	0.000230542817801695
Carnobacteriaceae	0.000696730033145692	6.9478657419689E-05	0.00110985387826256	0.000394765098975506
Clostridiaceae	0.00712763569473313	0.00113692348504946	0.000281523910583675	0.0194698146814719
Clostridiales_unclassified	0.00136430818205935	0.000776897714783795	0.000227384697009891	0.00126009019592981
Comamonadaceae	9.32860295425195E-05	1.89487247508243E-05	0.000108278427147567	5.05299326688647E-05
Coriobacteriales	0.00735210520331982	0.0135230732305049	0.0145038953164166	0.0499804196510908
Desulfovibrionaceae	0.000518903539330265	0.00026528214651154	5.9553134931162E-05	0.00157590227511022
Enterobacteriaceae	0.142488579749352	0.225306653528884	0.448927772875171	0.223989717158702
Enterococcaceae	0.00432030924318793	0.0191571607230833	0.00314548830863683	0.0218257727921578
Erysipelotrichaceae	0.0155758517451775	0.0171991258321648	0.00850527045244141	0.0198077336061 95
Eubacteriaceae	6.99645221568896E-05	5.05299326688647E-05	1.62417640721351E-05	0.000656889124695241
Family_XI_Incertae_Sedis	0.000653002206797636	0.000126324831672162	0.000221970775652513	0.000214752213842675
Fusobacteriaceae	0.00608399823922619	0.0270398302194262	0.00268530499325967	0.000669521607862457
Lachnospiraceae	0.0960467129792934	0.13012089286391	0.126615378785008	0.201870239132906
Lactobacillaceae	0.0308864213438436	3.15812079180405E-05	3.24835281442702E-05	0.00501193769659302
Leuconostocaceae	0.00235838743437182	0.000101059865337729	0.000535978214380458	0.00752933894215 59
Pasteurellaceae	0.0159898085012725	0.0253407612334357	8.12088203606754E-05	0.000461085635603391
Peptostreptococcaceae	0.00244584308706793	0.00600042950442769	0.00350822103958118	0.0235722135900254
Porphyromonadaceae	0.00535228594500205	0.0181086646202044	0.0606792305734967	0.0250186329126716
Rikenellaceae	0.000778355308995397	0.0347709099177625	8.12088203606754E-05	0.000833743889036268
Ruminococcaceae	0.0548346942404622	0.0679122295069542	0.069211570632725	0.0266861206907442
Staphylococcaceae	0.000239045450702706	0.000132641073255 77	0.000200315090222999	0.00323391569080734
Streptococcaceae	0.0951342590028307	0.0170159548262402	0.0324348028520538	0.0569661828425614
Veillonellaceae	0.0994312467386329	0.0604716969214639	0.039678629628226	0.0304663912785336
Verrucomicrobiaceae	0.00678655864921829	0.0610906885966574	0.0431543671396629	0.00642993393211304

Appendix 2- 4. Percent Abundance at Family Level

LEfSe Output at Phylum Level

	Pvalues	FDR	1	2	3	4	LDAscore
Fusobacteria	0.25052	0.74535	161.69	713.83	72.857	17.167	2.54
Actinobacteria	0.30065	0.74535	4450.2	2674	1948.7	4840.1	3.16
Candidate_division_TM7	0.42307	0.74535	58.769	1.8333	8	6.4167	1.47
Proteobacteria	0.55694	0.74535	6183.8	6392	12319	6020.9	3.5
Firmicutes	0.60285	0.74535	10754	8729.2	7533.7	11594	3.31
Bacteroidetes	0.63887	0.74535	4595.3	6261.2	3368.7	3739.6	3.16
Verrucomicrobia	0.94195	0.94195	183.46	1615	1136	169.25	2.86

Appendix 2- 5.LEfSe Output at Phylum Level

LEfSe Output at Class Level

	Pvalues	FDR	1	2	3	4	LDAscore
Epsilonproteobacteria	0.0011647	0.013976	1904.9	26.5	459.57	0.75	2.98
Bacilli	0.14849	0.6346	3526.2	917.83	992.43	3039.8	3.12
Fusobacteria	0.25052	0.6346	161.69	713.83	72.857	17.167	2.54
Actinobacteria	0.30065	0.6346	4450.2	2674	1948.7	4840.1	3.16
Betaproteobacteria	0.32189	0.6346	61.462	8.5	14.714	57.083	1.44
Gammaproteobacteria	0.37364	0.6346	4203.8	6350.2	11843	5923.7	3.58
Erysipelotrichi	0.3761	0.6346	411.92	450.83	224	520.92	2.17
Candidate_division_TM7_unclassified	0.42307	0.6346	58.769	1.8333	8	6.4167	1.47
Bacteroidia	0.63887	0.85183	4595.3	6261.2	3368.7	3739.6	3.16
Clostridia	0.76362	0.91634	6815.8	7360.5	6317.3	8032.8	2.93
Deltaproteobacteria	0.88539	0.94195	13.615	6.8333	1.8571	39.417	1.3
Verrucomicrobiae	0.94195	0.94195	183.46	1615	1136	169.25	2.86

Appendix 2- 6. LEfSe Output at Class Level

LEfSe Output at Order Level

	Pvalues	FDR	4	3	1	2	LDAscore
Campylobacterales	0.0011647	0.01747	0.75	459.57	1904.9	26.5	2.98
Pasteurellales	0.020216	0.15162	11.25	2	428	668.67	2.52
Lactobacillales	0.1436	0.53684	2947.2	979	3500.5	912.5	3.11
Bacillales	0.16752	0.53684	92.583	13.429	25.692	5.3333	1.65
Enterobacteriales	0.24242	0.53684	5912.4	11841	3775.8	5681.5	3.61
Coriobacteridae	0.24273	0.53684	1317.8	385	194.62	358.83	2.75
Fusobacteriales	0.25052	0.53684	17.167	72.857	161.69	713.83	2.54
Burkholderiales	0.32189	0.60355	57.083	14.714	61.462	8.5	1.44
Erysipelotrichales	0.3761	0.62683	520.92	224	411.92	450.83	2.17
Candidate_division_TM7_unclassified	0.42307	0.6346	6.4167	8	58.769	1.8333	1.47
Bacteroidales	0.63887	0.86285	3739.6	3368.7	4595.3	6261.2	3.16
Actinobacteridae	0.69028	0.86285	3522.3	1563.7	4255.5	2315.2	3.13
Clostridiales	0.76362	0.8811	8032.8	6317.3	6815.8	7360.5	2.93
Desulfovibrionales	0.88539	0.94195	39.417	1.8571	13.615	6.8333	1.3
Verrucomicrobiales	0.94195	0.94195	169.25	1136	183.46	1615	2.86

Appendix 2- 7. LEfSe Output at Order Level

LEfSe Output at Family Level

	Pvalues	FDR	1	2	3	4	LDAscore
Campylobacteraceae	0.0011647	0.037269	1904.9	26.5	459.57	0.75	2.98
Clostridiaceae	0.015808	0.21563	189	485.33	7.4286	533.5	2.42
Pasteurellaceae	0.020216	0.21563	428	668.67	2	11.25	2.52
Streptococcaceae	0.068489	0.54791	2512.5	450	863.43	1507.6	3.01
Veillonellaceae	0.11021	0.70537	2634.8	1472.5	1046.7	802.25	2.96
Aerococcaceae	0.19062	0.80167	2.3077	0	0	0.5	0.333
Staphylococcaceae	0.19225	0.80167	6.3846	3.5	5.4286	85.417	1.62
Enterobacteriaceae	0.24242	0.80167	3775.8	5681.5	11841	5912.4	3.61
Coriobacteriales	0.24273	0.80167	194.62	358.83	385	1317.8	2.75
Fusobacteriaceae	0.25052	0.80167	161.69	713.83	72.857	17.167	2.54
Enterococcaceae	0.2785	0.80235	113.85	459.67	85.143	575.75	2.39
Carnobacteriaceae	0.32732	0.80235	18.154	1.8333	26.143	9.9167	1.12
Alcaligenaceae	0.35686	0.80235	58.923	7.3333	12	55.833	1.43
Lactobacillaceae	0.36702	0.80235	814.38	1	0.85714	132	2.61
Erysipelotrichaceae	0.3761	0.80235	411.92	450.83	224	520.92	2.17
Candidate_division_TM7_unclassified	0.42307	0.8335	58.769	1.8333	8	6.4167	1.47
Leuconostocaceae	0.45374	0.8335	39.308	0	3.4286	721.5	2.56
Bacteroidaceae	0.50862	0.8335	4424.5	4868	1763.7	3056.1	3.19
Lachnospiraceae	0.54217	0.8335	2568.8	3445.5	3346.9	5336.1	3.14
Peptostreptococcaceae	0.54449	0.8335	63.846	160.83	93.429	623.83	2.45
Family_XI_Incertae_Sedis	0.55599	0.8335	17.846	2.5	6.1429	5.9167	0.938
Bacillaceae	0.60399	0.8335	2.5385	0.16667	2.5714	1.8333	0.343
Actinomycetales	0.61417	0.8335	68.308	17	41.143	25.583	1.43
Rikenellaceae	0.62825	0.8335	21.769	916.17	2.2857	21.833	2.66
Ruminococcaceae	0.65117	0.8335	1319.8	1772.3	1815.9	686.08	2.75
Bifidobacteriales	0.68002	0.83695	4187.2	2298.2	1522.6	3496.8	3.12
Porphyromonadaceae	0.73568	0.85485	149	477	1602.7	661.67	2.86
Clostridiales_unclassified	0.75971	0.85485	36.538	22	5.8571	33.583	1.21
Comamonadaceae	0.77471	0.85485	2.5385	1.1667	2.7143	1.25	0.249
Eubacteriaceae	0.80591	0.85964	1.8462	1.1667	0.42857	16.917	0.966
Desulfovibrionaceae	0.88539	0.91395	13.615	6.8333	1.8571	39.417	1.3
Verrucomicrobiaceae	0.94195	0.94195	183.46	1615	1136	169.25	2.86

Appendix 2- 8. LEfSe Output at Family Level

LEfSe Output at Genus Level

	Pvalues	FDR	1	2	3	4	LDAscore
Campylobacter	0.0011647	0.060562	1904.9	26.5	459.57	0.75	2.98
Haemophilus	0.020216	0.52561	428	668.67	2	11.25	2.52
Johnsonella	0.038956	0.60279	20.385	0.33333	19.429	1.25	1.04
Clostridium	0.046369	0.60279	183.92	480.5	4.8571	529.33	2.42
Streptococcus	0.064036	0.66597	2423.4	443.67	860.86	1358.9	3
Veillonella	0.10287	0.89152	2449.2	1468.8	960.14	783	2.92
Abiotrophia	0.19062	0.92825	2.3077	0	0	0.5	0.333
Staphylococcus	0.19225	0.92825	6.3846	3.5	5.4286	85.417	1.62
Parasutterella	0.22414	0.92825	36.538	3.8333	0.14286	10.333	1.28
Enteric_Bacteria_cluster	0.24242	0.92825	3770.9	5680.8	11839	5911.7	3.61
Coriobacterineae	0.24273	0.92825	194.62	358.83	385	1317.8	2.75
Incertae_Sedis	0.24939	0.92825	1735.8	2546	1369.4	3245.8	2.97
Fusobacterium	0.25052	0.92825	161.69	713.83	72.857	17.167	2.54
Odoribacter	0.25917	0.92825	6.0769	71.333	6.2857	9.6667	1.53
Enterococcus	0.2785	0.92825	113.85	459.67	85.143	575.75	2.39
Subdoligranulum	0.28562	0.92825	70.077	335	359.43	204.42	2.16
Gemella	0.32003	0.9456	16.769	1.6667	5.4286	5.3333	0.932
Granulicatella	0.32732	0.9456	18.154	1.8333	26.143	9.9167	1.12
Lactobacillus	0.36702	0.9604	814.38	1	0.85714	132	2.61
Megasphaera	0.3907	0.9604	10.615	0.83333	10.143	8.0833	0.77
Candidate_division_TM7_unclassified	0.42307	0.9604	58.769	1.8333	8	6.4167	1.47
Finegoldia	0.44563	0.9604	0.76923	0.33333	0.57143	0.25	0.1
Peptostreptococcus	0.4471	0.9604	11.769	2.8333	1.7143	1.25	0.797
Weissella	0.45374	0.9604	39.308	0	3.4286	721.5	2.56
Clostridiaceae_unclassified	0.46173	0.9604	5.0769	4.8333	2.5714	4.1667	0.353
Bacteroides	0.50862	0.96042	4424.5	4868	1763.7	3056.1	3.19
Micrococcineae	0.52526	0.96042	45.769	4.8333	18.429	12.083	1.33
Megamonas	0.57381	0.96042	175	2.8333	76.429	11.167	1.94
Bacillus	0.60399	0.96042	2.5385	0.16667	2.5714	1.8333	0.343
Blautia	0.62183	0.96042	473.46	606	1748.6	2446.4	2.99
Actinomycineae	0.6253	0.96042	22.538	12.167	22.714	13.5	0.798
Alistipes	0.62825	0.96042	21.769	916.17	2.2857	21.833	2.66
Dorea	0.65981	0.96042	22.538	77.167	252.86	109	2.07
Bifidobacteriaceae	0.68002	0.96042	4187.2	2298.2	1522.6	3496.8	3.12

Appendix 2- 9. LEfSe Output at Genus Level

Actual Abundances at Family Level

	1	2	3	4
Actinomycetales	872	98	294	311
Aerococcaceae	36	0	0	6
Alcaligenaceae	758	47	83	666
Bacillaceae	32	3	19	22
Bacteroidaceae	57223	29279	12324	36858
Bifidobacteriales	54354	13870	10658	41995
Campylobacteraceae	24737	164	3205	10
Candidate_division_TM7_unclassified	775	11	48	76
Carnobacteriaceae	268	11	190	118
Clostridiaceae	2453	2903	51	6417
Clostridiales_unclassified	464	121	42	406
Comamonadaceae	29	6	17	17
Coriobacteriales	2519	2159	2703	15833
Desulfovibrionaceae	167	42	12	495
Enterobacteriaceae	49047	34055	82955	71023
Enterococcaceae	1481	2737	579	6912
Erysipelotrichaceae	5352	2740	1584	6278
Eubacteriaceae	24	8	3	207
Family_XI_Incertae_Sedis	240	19	42	72
Fusobacteriaceae	2077	4286	499	211
Lachnospiraceae	32794	20679	23387	64160
Lactobacillaceae	10600	5	3	1584
Leuconostocaceae	823	17	92	8717
Pasteurellaceae	5480	4015	17	142
Peptostreptococcaceae	839	957	651	7469
Porphyromonadaceae	1199	2518	11026	7134
Rikenellaceae	289	5521	16	268
Ruminococcaceae	18902	10776	12863	8478
Staphylococcaceae	86	22	40	1024

Appendix 2- 10. Actual Abundances

Unfiltered Goods Coverage

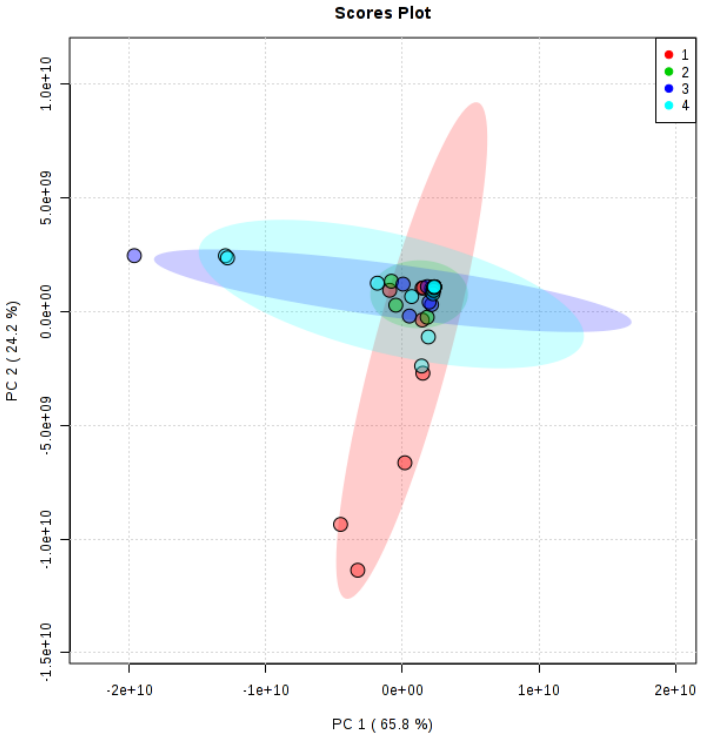
sample	no.singleton	no.seqs	goods
CCN1	116	119544	99.9029645988088
CCN10	131	59187	99.7786676128204
CCN11	210	105536	99.8010157671316
CCN12	349	115838	99.6987171739843
CCN13	149	102349	99.8544196816774
CCN14	201	99825	99.7986476333584
CCN16	240	82808	99.7101729301517
CCN17	297	291634	99.8981600224939
CCN18	241	288903	99.9165809977743
CCN19	456	190852	99.7610714061157
CCN2	109	68599	99.8411055554746
CCN20	289	323091	99.9105515164458
CCN3	295	120630	99.7554505512725
CCN4	190	134011	99.8582205938318
CCN5	188	131499	99.8570331333318
CCN6	435	141714	99.6930437359753
CCN7	390	117273	99.6674426338543
CCN8	336	104158	99.677413160775
CCN9	157	110665	99.85813039353
CCP1	74	98700	99.9250253292806
CCP10	257	79534	99.676867754671
CCP12	177	67173	99.7365012728328
CCP13	150	81755	99.8165249831815
CCP14	65	35098	99.8148042623511
CCP15	91	46797	99.805543090369
CCP16	88	27357	99.6783273019702
CCP17	194	69702	99.721672261915
CCP18	197	63774	99.6910966851695
CCP19	178	48801	99.6352533759554

Appendix 2- 11. Unfiltered Goods Coverage

Filtered Goods Coverage

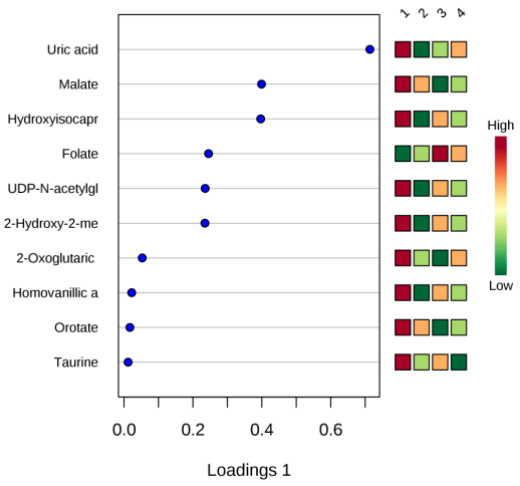
sample	no.singleton	no.seqs	goods
CCN1	6	26386	99.9772606685363
CCN10	9	26386	99.9658910028045
CCN11	5	26386	99.9810505571136
CCN12	5	26386	99.9810505571136
CCN13	8	26386	99.9696808913818
CCN14	6	26386	99.9772606685363
CCN16	6	26386	99.9772606685363
CCN17	13	26386	99.9507314484954
CCN18	16	26386	99.9393617827636
CCN19	8	26386	99.9696808913818
CCN2	10	26386	99.9621011142272
CCN20	16	26386	99.9393617827636
CCN3	7	26386	99.9734707799591
CCN4	10	26386	99.9621011142272
CCN5	7	26386	99.9734707799591
CCN6	12	26386	99.9545213370727
CCN7	12	26386	99.9545213370727
CCN8	6	26386	99.9772606685363
CCN9	6	26386	99.9772606685363
CCP1	10	26386	99.9621011142272
CCP10	13	26386	99.9507314484954
CCP12	7	26386	99.9734707799591
CCP13	6	26386	99.9772606685363
CCP14	5	26386	99.9810505571136
CCP15	6	26386	99.9772606685363
CCP16	8	26386	99.9696808913818
CCP17	6	26386	99.9772606685363
CCP18	11	26386	99.95831122565
CCP19	12	26386	99.9545213370727

Appendix 2- 12. Filtered Goods Coverage

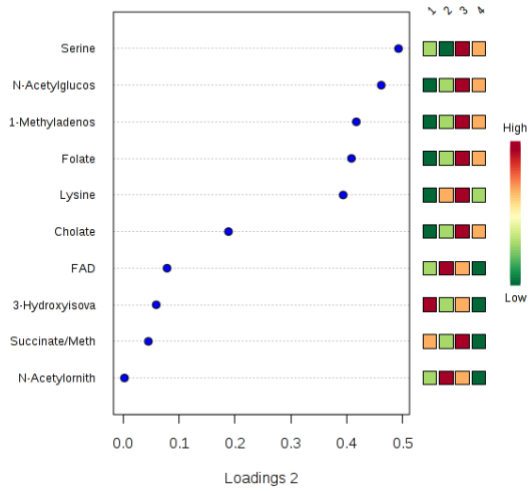


Appendix 2- 13. Metabolome PCA of Four Cohorts

a



b



Appendix 2- 14. sPLS-DA Loadings

univariate-analysis class MW

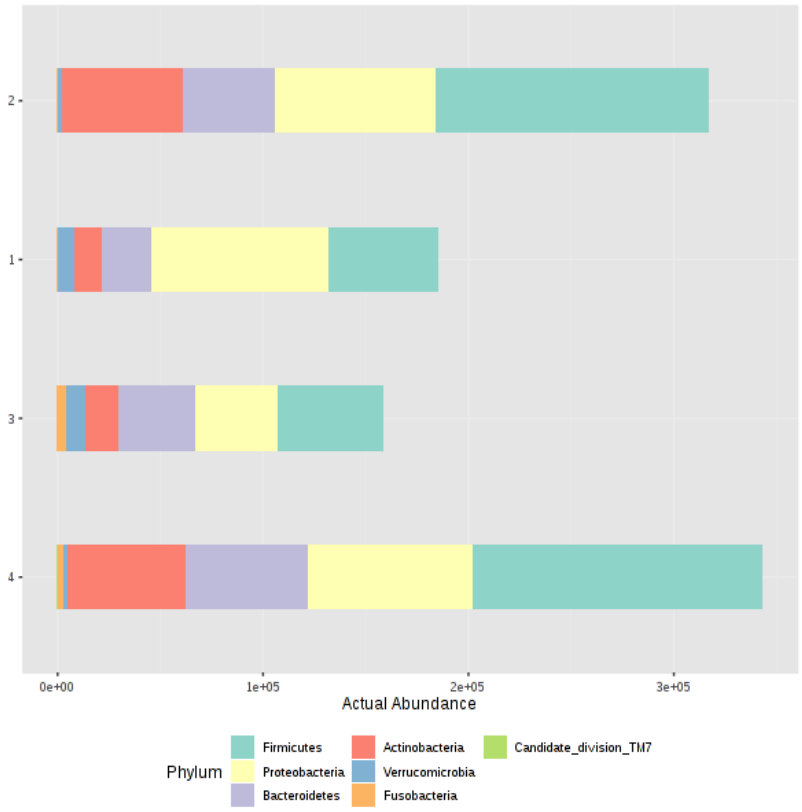
	Pvalues	FDR	Statistics
Epsilonproteobacteria	0.0011647	0.013976	15.943
Bacilli	0.14849	0.6346	5.3407
Fusobacteria	0.25052	0.6346	4.1033
Actinobacteria	0.30065	0.6346	3.6595
Betaproteobacteria	0.32189	0.6346	3.4913
Gammaproteobacteria	0.37364	0.6346	3.119
Erysipelotrichi	0.3761	0.6346	3.1024
Candidate_division_TM7_unclassified	0.42307	0.6346	2.8026
Bacteroidia	0.63887	0.85183	1.6913
Clostridia	0.76362	0.91634	1.1558
Deltaproteobacteria	0.88539	0.94195	0.6479
Verrucomicrobiae	0.94195	0.94195	0.39172

Appendix 2- 15. Microbiome Univariate Analysis of Four Cohorts

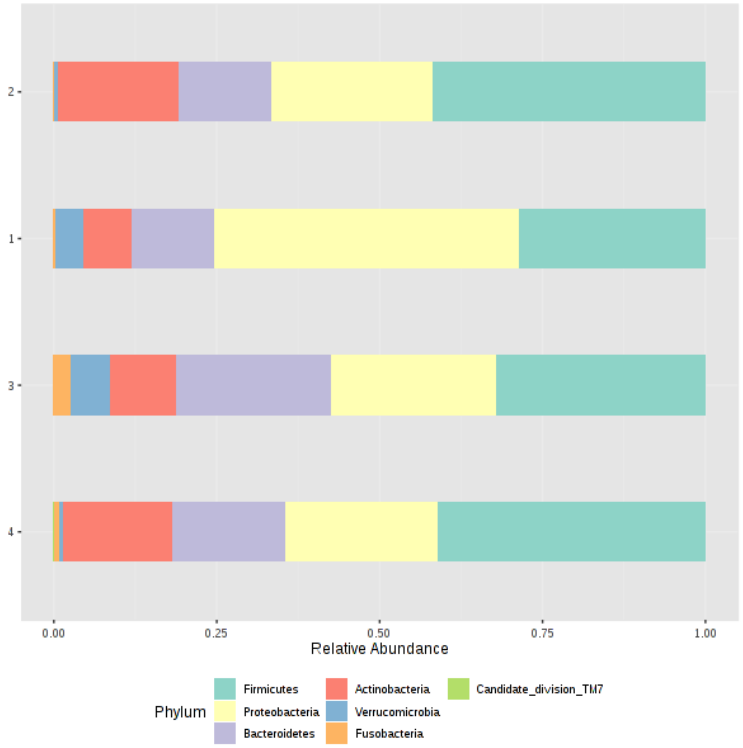
plsda_loadings_metabolome

	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8
1-Methyladenosine	0.0044658	-0.0032271	-0.0042346	0.0052202	-0.0038046	0.015528	0.0049193	-0.0037928
2-Hydroxy-2-methylsuccinate	-0.11395	-0.21234	0.2194	-0.033237	-0.29709	0.45918	0.48179	-0.40698
2-Oxo-4-methylthiobutanoate	-0.00077521	-0.00045196	-0.0042797	0.017117	0.0073939	0.011319	0.0016141	0.0027879
2-Oxoglutaric acid	-0.010376	-0.01347	-0.028439	0.054839	-0.035432	0.039106	-0.0010502	-0.0045116
3-Hydroxyisovaleric acid	-0.006031	-0.019942	0.034126	0.015217	-0.016785	0.025119	-0.023576	-0.0012967
3-Methylthiopropionate	-0.00014414	-9.8948E-05	-0.0037913	0.0029034	0.00067812	-0.0013526	-0.0018903	-0.01009
4-Pyridoxate	-0.022946	-0.025554	-0.39341	0.46254	0.13341	-0.018678	-0.032825	-0.4278
Allantoate	-3.3721E-05	-0.00012297	-0.00024169	-0.00058402	0.00049165	0.0010621	-0.00066249	0.00065313
Ascorbate	-0.0016368	-0.0013554	-0.030535	0.008077	-0.013485	0.065567	0.04242	0.13543
Asparagine	7.989E-06	-0.00021594	-0.00050284	5.5646E-05	0.0012199	0.0020339	-0.0014977	-0.00044996
Cholate	1.0662	-0.72894	0.063881	0.036307	-0.054707	0.010488	-0.012233	0.014912
Citrulline	-0.0054375	-0.018203	-0.025963	-0.01941	0.081616	-0.22198	0.1181	-0.096024
Creatinine	-9.4437E-05	-0.00022645	0.00030448	-0.00081947	0.0012367	-0.0027746	0.0015939	-5.6002E-06
Cysteate	-0.00039753	-0.0006645	-0.0017813	-0.0028291	0.0024634	0.0065941	-0.0054221	0.0026804
dAMP	-2.304E-05	-4.7465E-05	-0.00089467	0.00041337	-0.00043416	-0.00072498	0.00028607	-7.6174E-05
dCMP	-0.00015151	-0.00020773	-0.001285	0.00098364	-0.0014378	-0.0021108	0.00031212	-0.00053262
Deoxyinosine	0.0019242	-0.0033492	-0.0069793	0.013181	-0.0012005	0.0004113	-0.01913	-0.074107
D-Gluconate	-0.061887	-0.12984	-0.043761	-0.35572	0.57136	-0.75815	0.5427	-0.10049
FAD	-0.00089049	-0.00070833	-0.018187	0.0083777	-0.0042956	-0.015122	0.0048004	-0.0048205
Folate	6.6579E-05	-5.4431E-05	-0.00011425	0.0010013	0.00012653	0.0014213	0.00030299	0.0012097
Glucosamine	2.8653E-05	-8.3458E-05	-0.00031499	0.00018713	0.00046757	-0.00077738	-1.9903E-05	-0.0018638
Glucose 1-phosphate	-0.0014189	-0.0022837	0.0067076	0.0031978	-0.013342	0.0066919	-0.0042738	-0.013546
Glucose 6-phosphate	-0.0014176	-0.0022813	0.0067895	0.0031963	-0.013305	0.0068698	-0.0042746	-0.013421
Glutamine	-0.00053713	-0.0010972	-0.0085725	0.0036065	0.0048364	-0.0099084	0.0035718	-0.014878
Glutathione disulfide	9.6842E-05	1.0668E-05	-0.00078091	-0.0024541	0.00053639	0.011165	0.00077703	0.0092587
Guanosine	0.00014598	-0.00015165	-0.0012524	0.00038451	4.7915E-05	-0.0006048	-0.00066499	-0.00087639
Homocysteic acid	-3.9428E-05	-4.9954E-05	-0.00093601	0.00086048	0.0001686	-3.636E-05	-0.00014246	-0.0019312
Homoserine/Threonine	-0.0051138	-0.020177	-0.0024849	0.010551	0.015595	-0.055018	0.050821	-0.06751
Homovanillic acid	-0.085322	-0.19064	0.36211	-0.032392	-0.18885	-0.47936	0.50333	-0.68977
Hydroxyisocaproic acid	-0.28536	-0.51372	0.72039	-0.10625	-0.0020407	-0.057186	-0.37808	0.57542
Hydroxyproline	-3.0245E-05	-0.00014521	-0.00030672	-0.00056739	0.00058002	0.0010962	-0.00067752	0.00085452
Hypoxanthine	0.00026303	-0.011	-0.052174	0.071854	0.028256	-0.028826	0.02374	-0.044644
Kynurenic acid	-0.00081695	-0.002805	-0.0099963	0.010876	-0.0085394	0.0038311	-0.0030101	-0.033031
Leucine/Isoleucine	-0.026757	-0.12072	-0.35971	0.19994	0.34343	0.057694	-0.18193	-0.037952
L-Methionine	0.0025674	-0.0049531	-0.014189	0.016432	0.014298	-0.017898	-0.012039	-0.052748
Lysine	0.00042024	-0.00022428	-0.00043958	0.00095171	0.0010367	-0.0018882	-0.001508	-0.0029046
Malate	-0.13667	-0.20533	-0.80841	0.35417	-0.38829	0.65865	-0.44002	-0.046938
N-Acetylglucosamine	0.0035944	-0.0027726	-0.0019313	0.005563	-0.0040429	0.0056999	-0.00051647	-0.0062433
N-Acetylglutamate	-0.066353	-0.076039	-0.68722	0.57053	-0.24398	-0.33337	0.3503	0.11049
N-Acetylornithine	0.00089119	-0.0026418	-0.0059285	0.0071693	0.010328	0.013441	-0.010575	-0.024451
Nicotinate	-0.0031937	-0.0035286	-0.056063	0.043373	0.0012113	-0.036329	0.025723	0.023101
Ornithine	-0.0019701	-0.0056893	-0.02069	0.0097641	0.024263	-0.067506	0.031453	-0.019445
Orotate	-0.010799	-0.017264	-0.0064113	0.027427	-0.022177	0.028339	0.0018775	-0.11812
Pantothenate	-0.019393	-0.0477	-0.76632	0.46809	-0.12069	-0.57799	0.30131	0.3173
Phenylalanine	-0.016957	-0.1076	-0.19881	-0.24286	0.37208	0.23832	-0.35688	0.13633

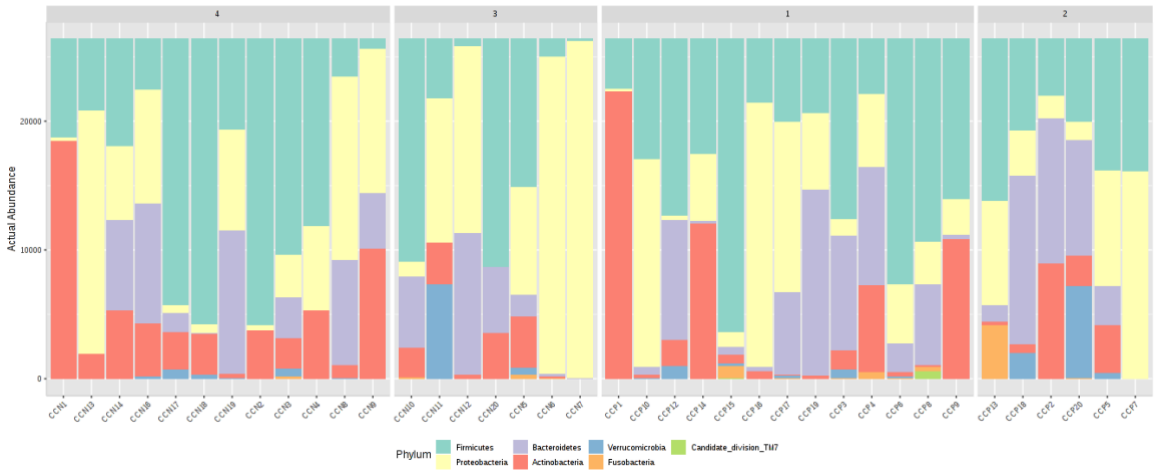
Appendix 2- 16. PLS-DA Loadings



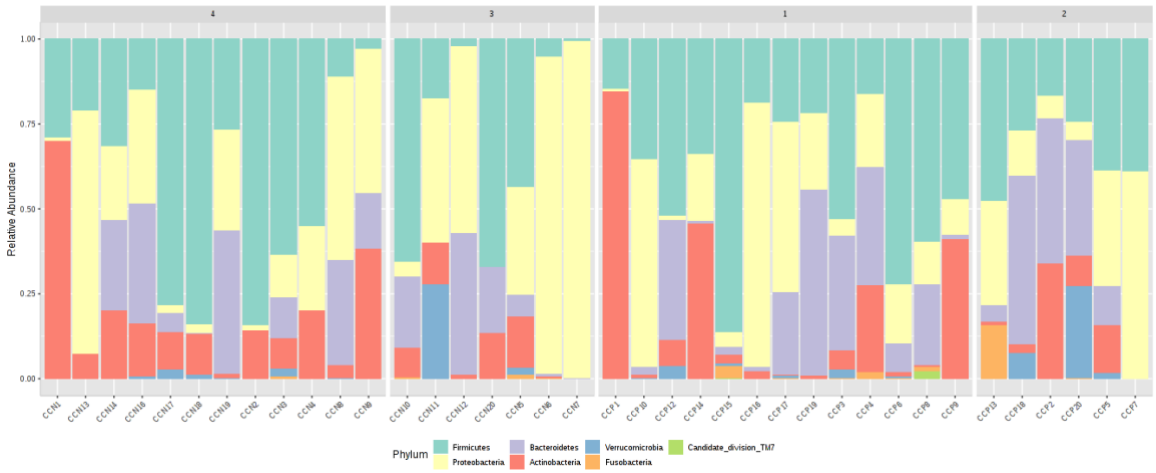
Appendix 2- 17. Actual Relative Abundance of Microbiome



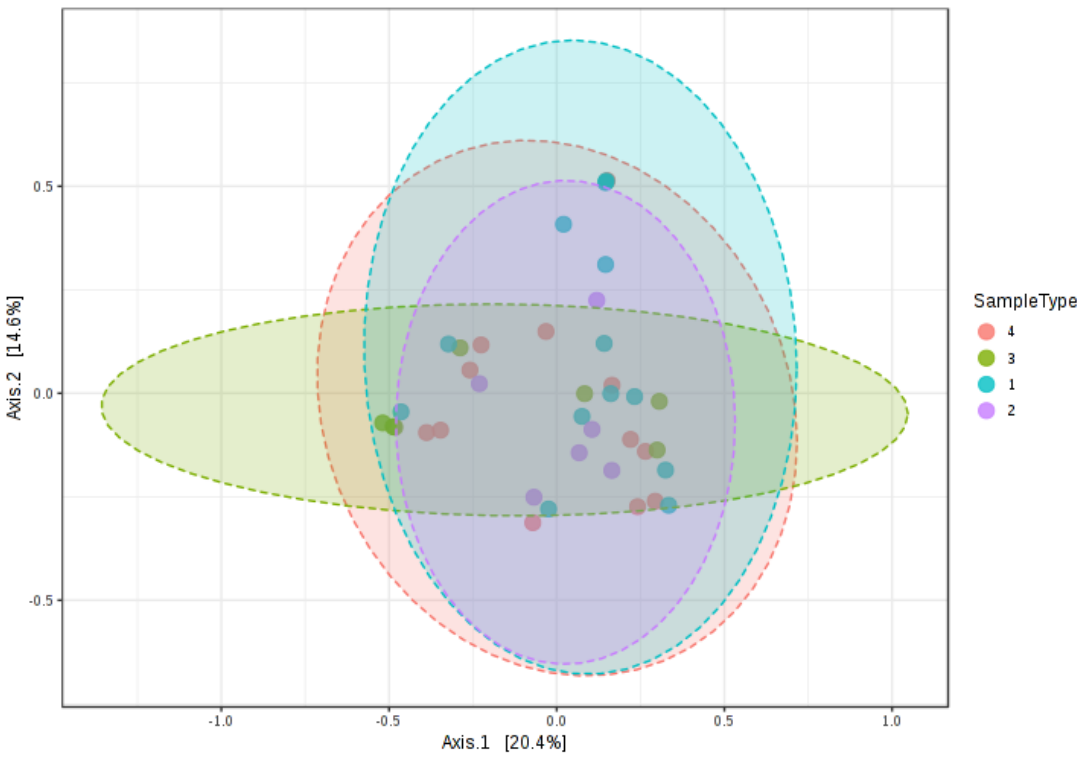
Appendix 2- 18. Relative Abundance of Microbiome Based on Percentage



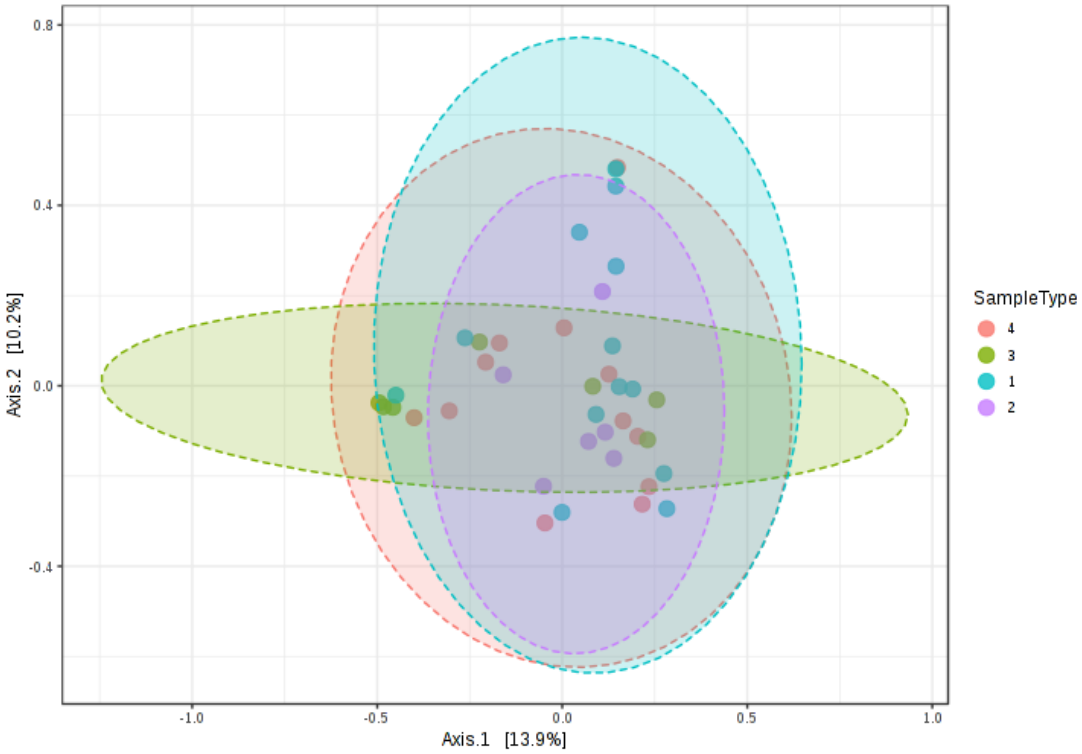
Appendix 2- 19. Actual Relative Abundance of Individual Subjects Microbiomes



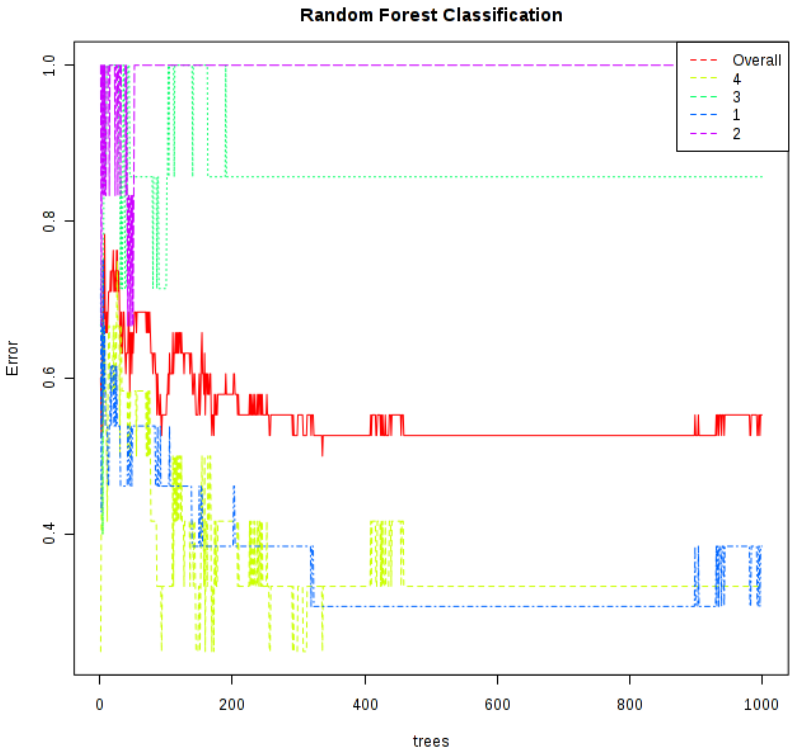
Appendix 2- 20. Relative Abundance of Individual Subjects' Microbiome Based on Percentage



Appendix 2- 21. PCoA Plot of Beta-Diversity of Four Cohorts

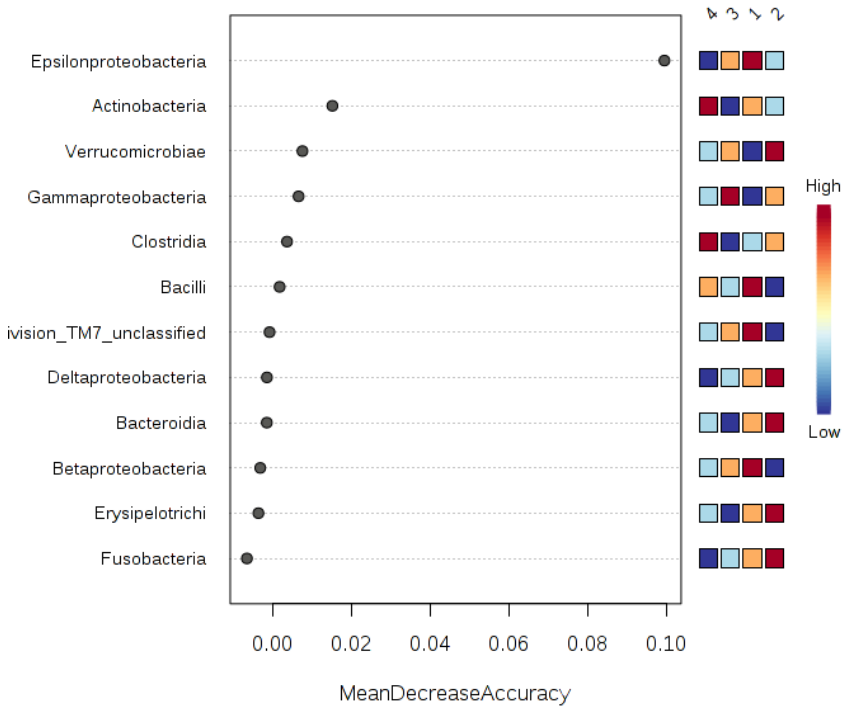


Appendix 2- 22. PCoA Plot of Beta-Diversity of Four Cohorts



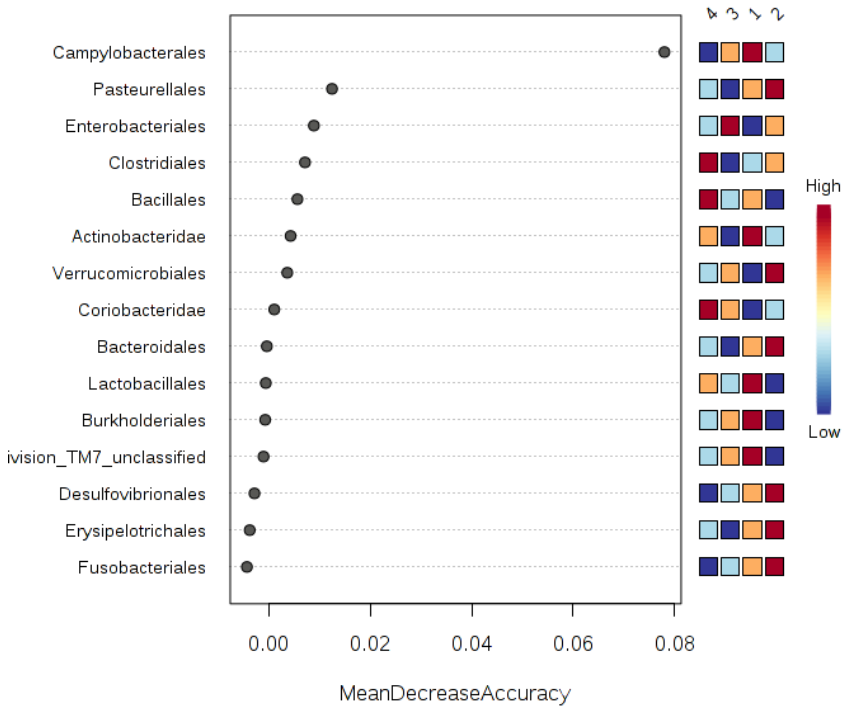
Appendix 2- 23. Random Forest of Microbiome Classification Plot

The number of trees to grow was set to 1000, with 7 predictors, and randomness setting turned on.



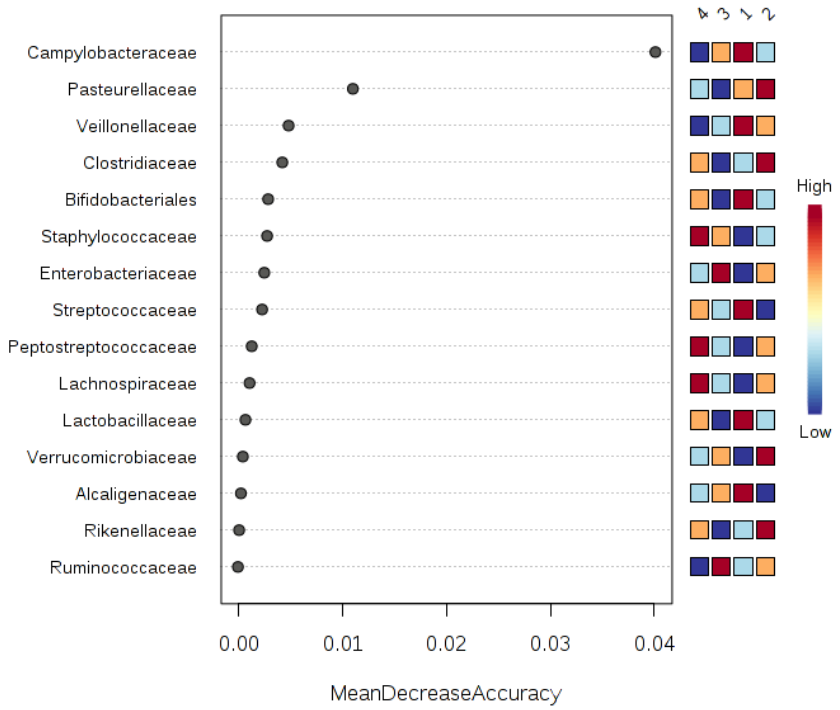
Appendix 2- 24. Random Forest Results at Class Level

The number of trees to grow was set to 1000, with 7 predictors, and randomness setting turned on.



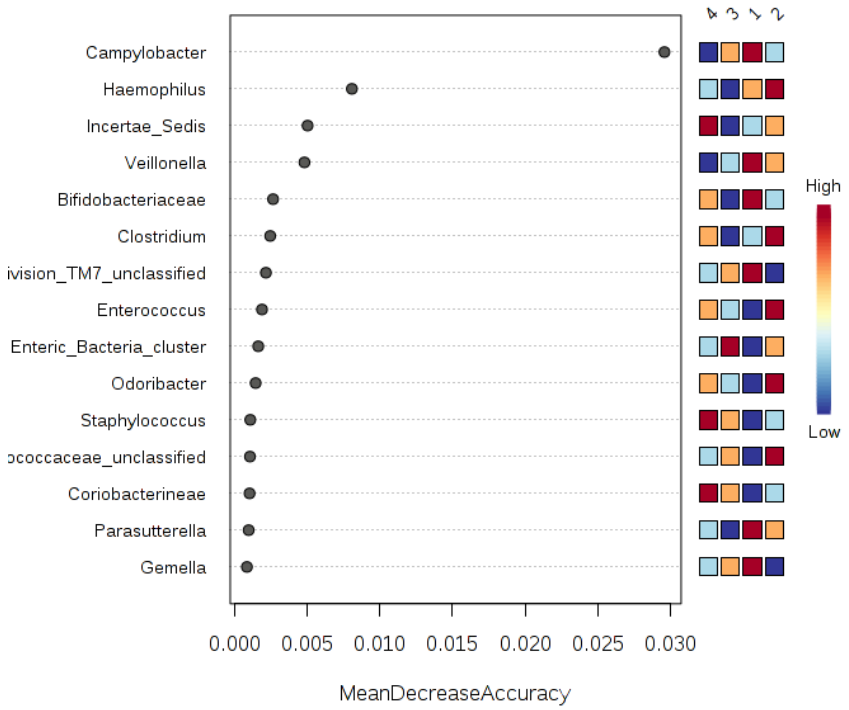
Appendix 2- 25. Random Forest Results at Order Level

The number of trees to grow was set to 1000, with 7 predictors, and randomness setting turned on.



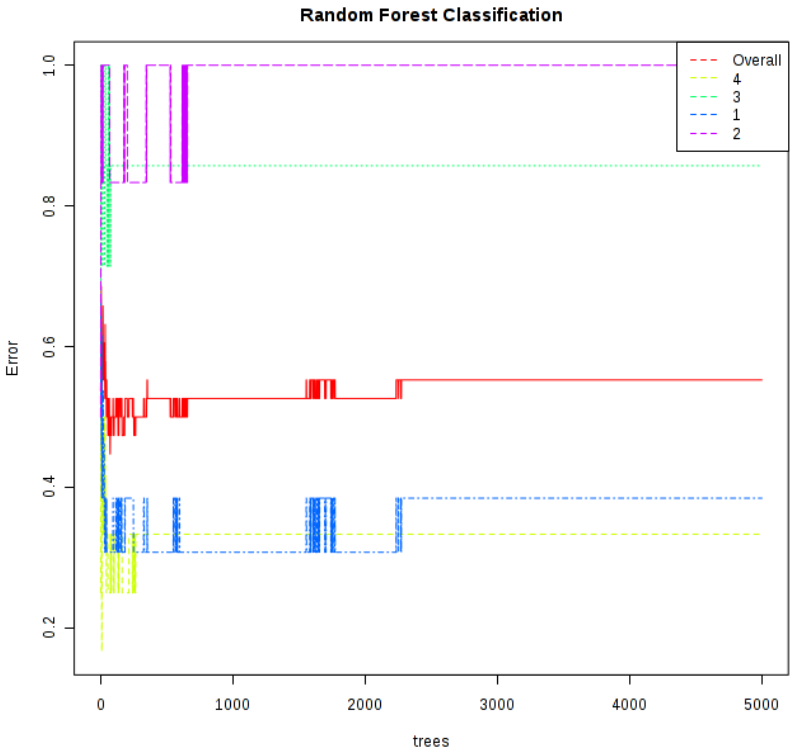
Appendix 2- 26. Random Forest Results at Family Level

The number of trees to grow was set to 1000, with 7 predictors, and randomness setting turned on.



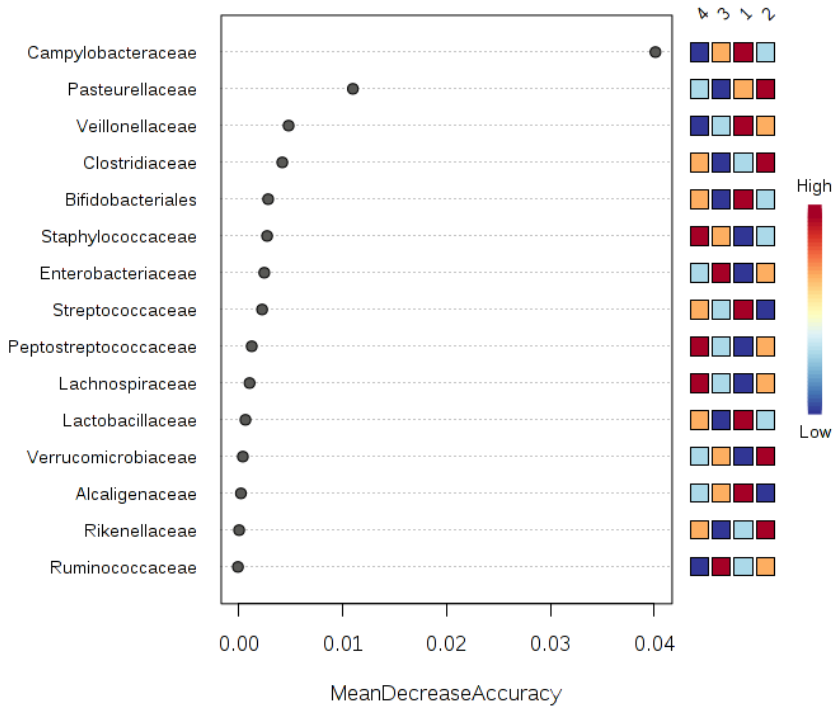
Appendix 2- 27. Random Forest Results at Genus Level

The number of trees to grow was set to 1000, with 7 predictors, and randomness setting turned on.



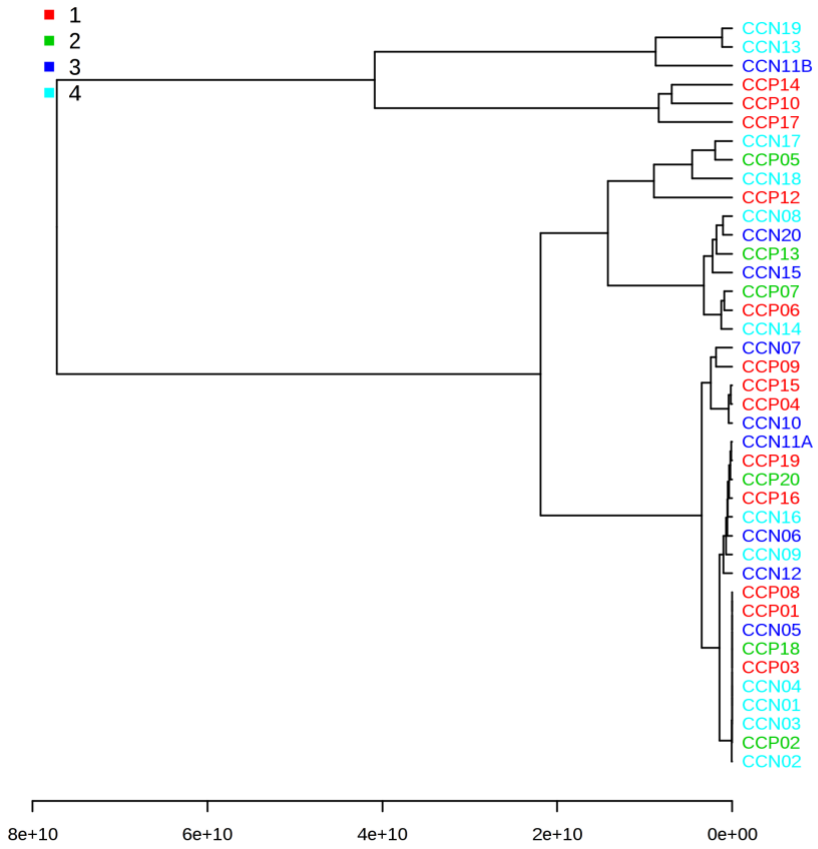
Appendix 2- 28. Random Forest Classification Plot

The number of trees to grow was set to 5000, with 7 predictors, and randomness setting turned on.



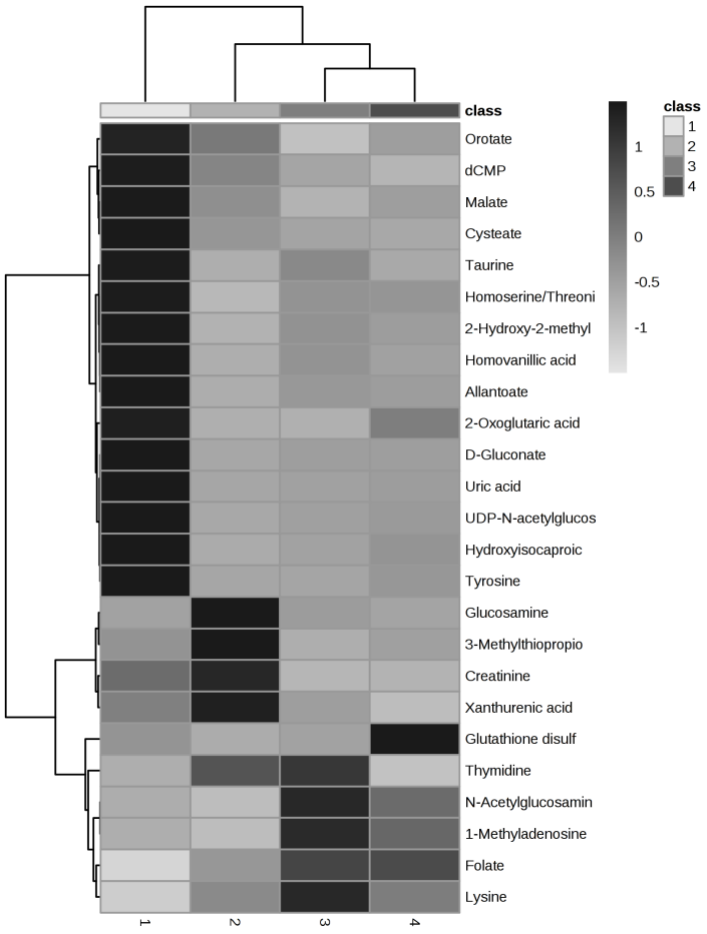
Appendix 2- 29. Random Forest Results at Family Level

The number of trees to grow was set to 5000, with 7 predictors, and randomness setting turned on.



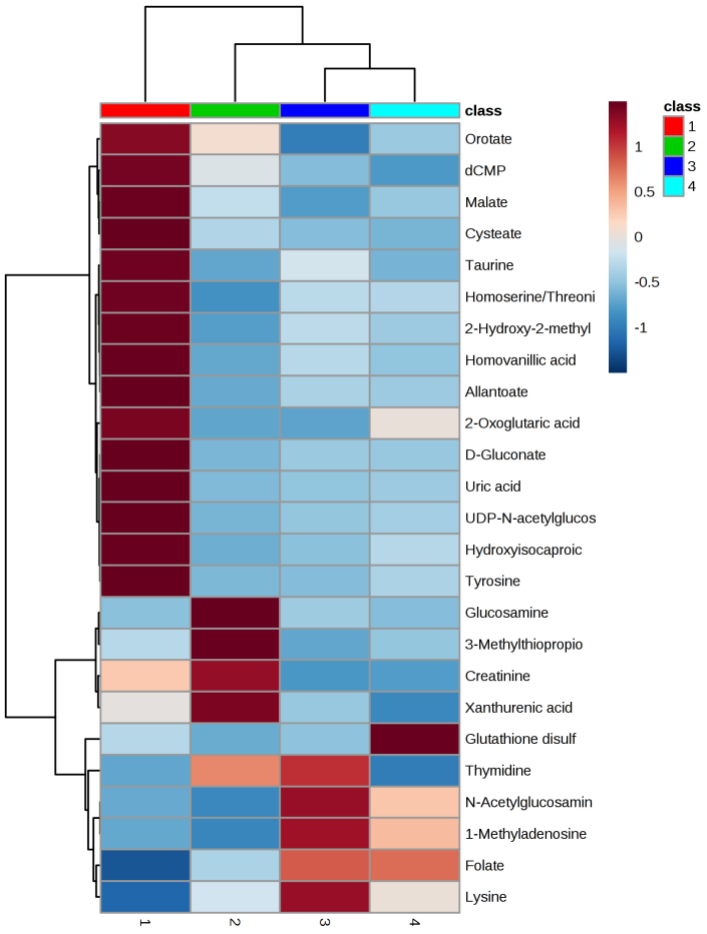
Appendix 2- 30. Dendrogram based on Metabolomic Analysis

This dendrogram was created using Euclidean ward model.



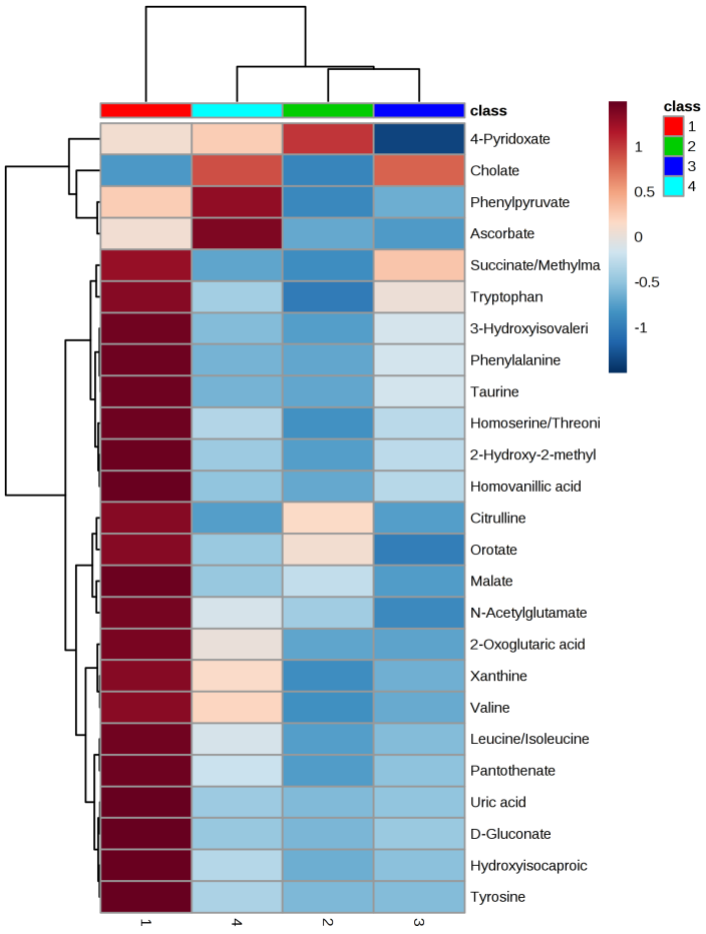
Appendix 2- 31. Metabolome Heat Map 1

This heat map was created based on the distance measure of Euclidean ward and statistical method of T-Test/ANOVA.



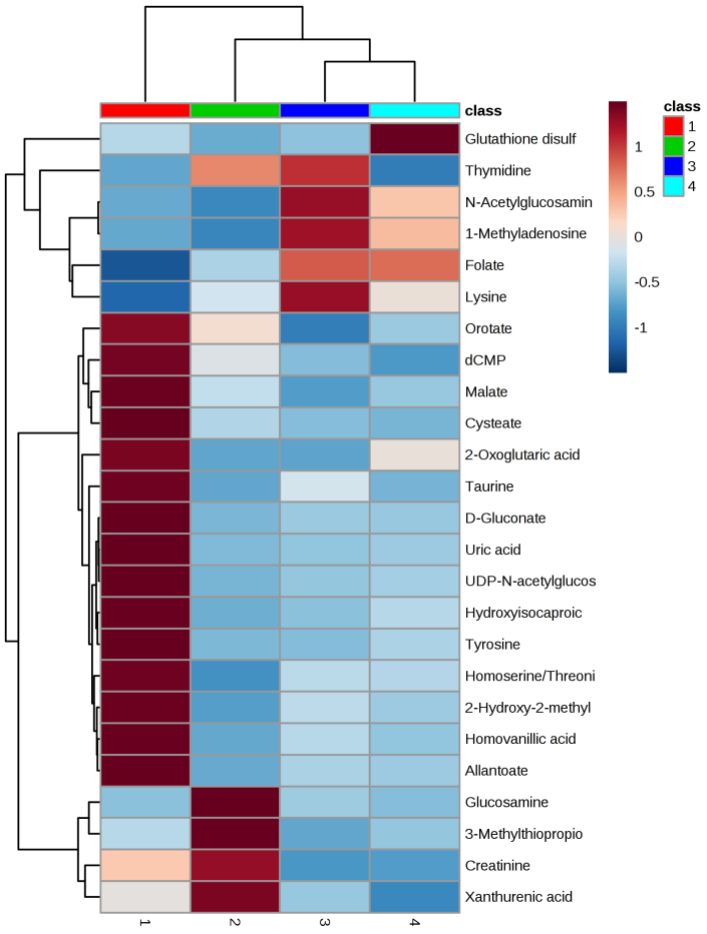
Appendix 2- 32. Metabolome Heatmap 2

This heat map was created based on the distance measure of Euclidean ward and statistical method of T-Test/ANOVA. Same as Heatmap 1, but in color.



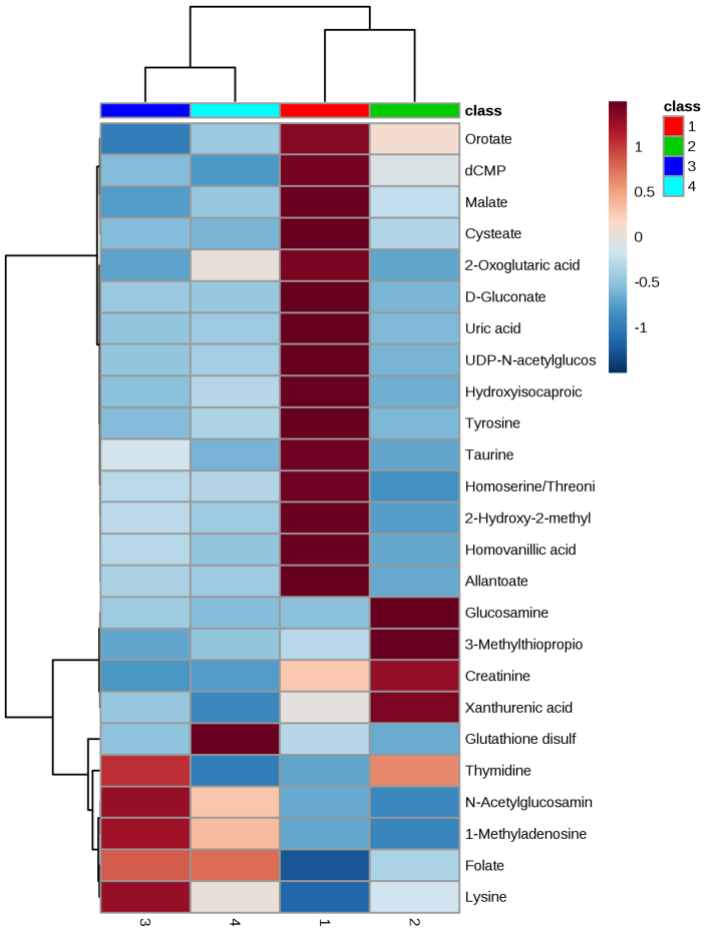
Appendix 2- 33. Metabolome Heatmap 3

This heat map was created based on the distance measure of Euclidean ward and VIP results.



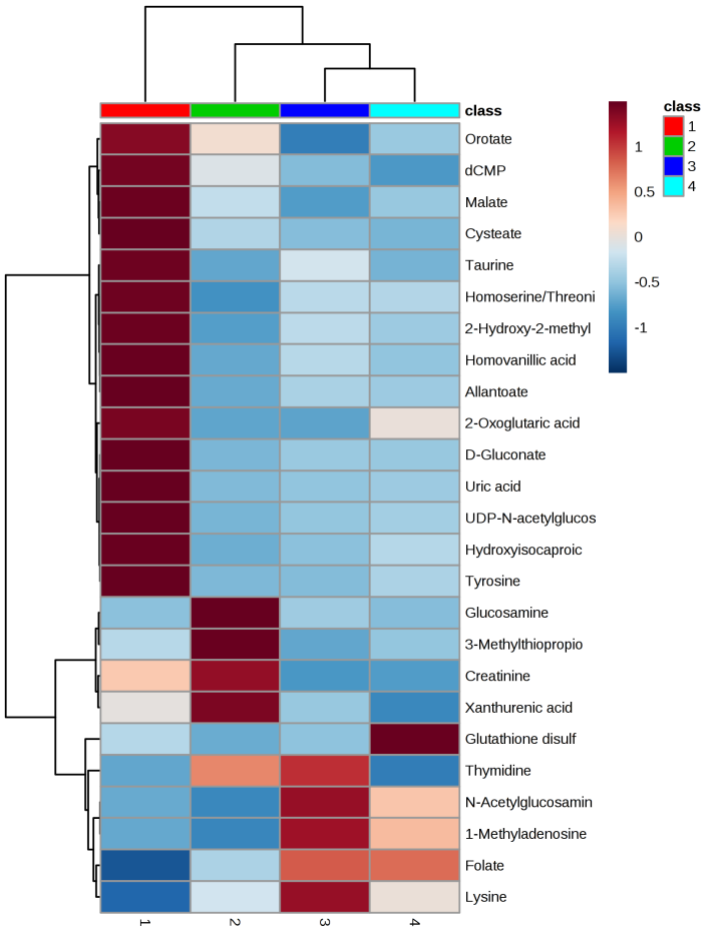
Appendix 2- 34. Metabolome Heatmap 4

This heat map was created based on the distance measure of Euclidean average and statistical method of T-Test/ANOVA.



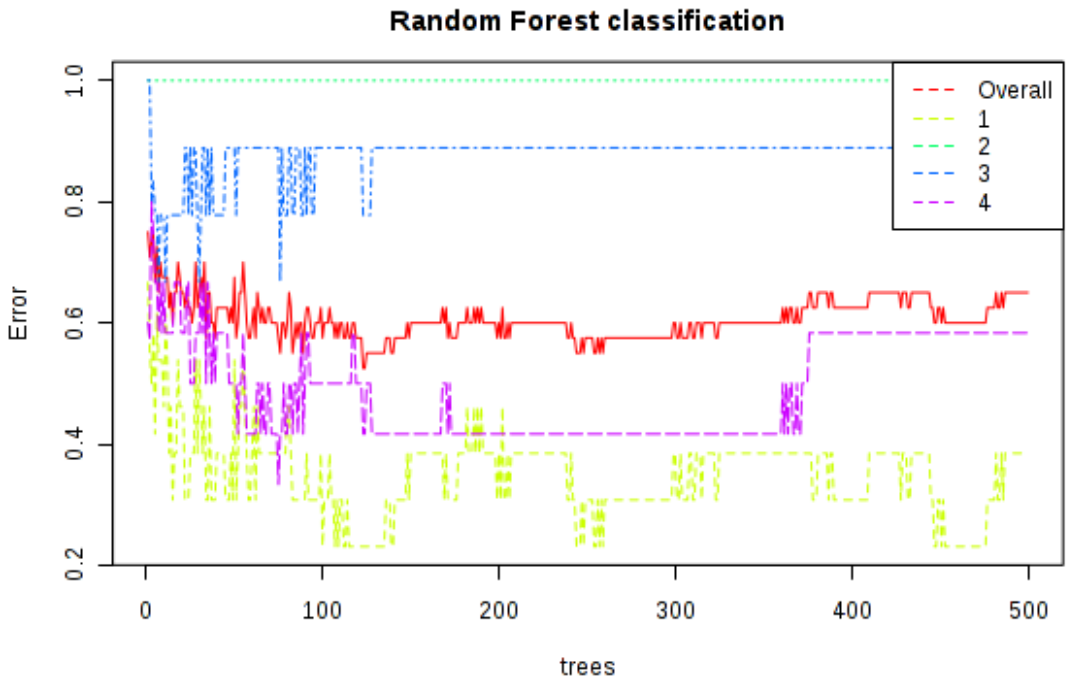
Appendix 2- 35. Metabolome Heatmap 5

This heat map was created based on the distance measure of Pearson ward and statistical method of T-Test/ANOVA.



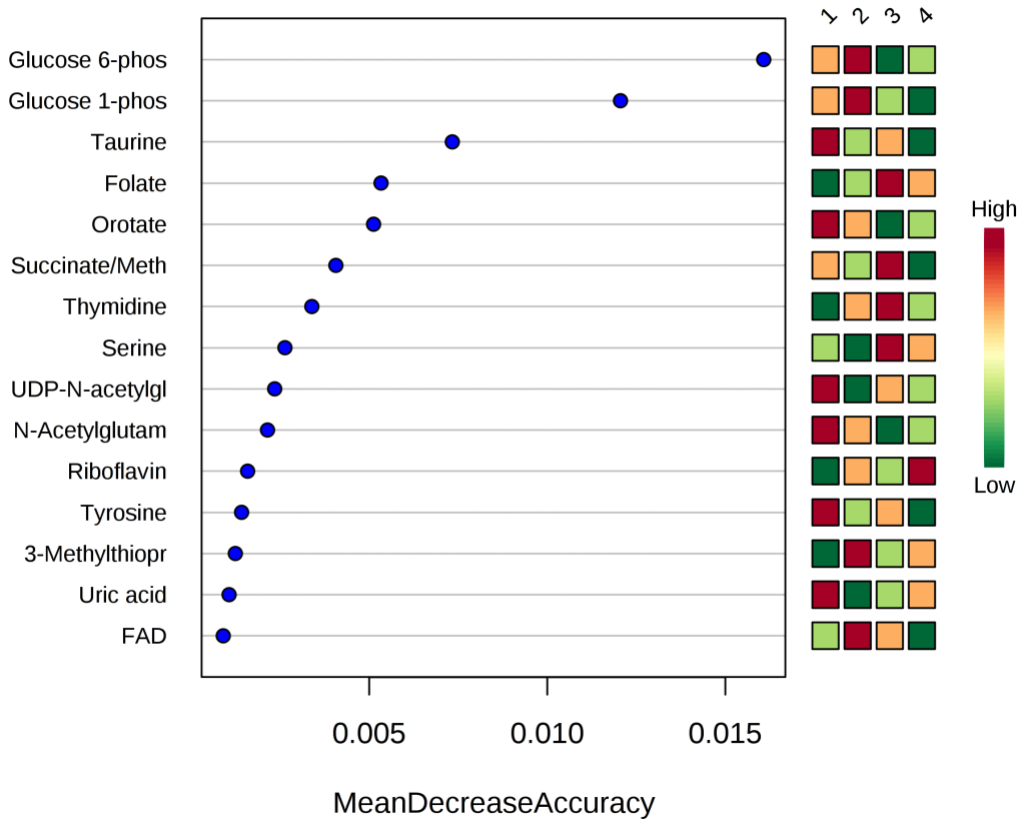
Appendix 2- 36. Metabolome Heatmap 6

This heat map was created based on the distance measure of Minkowsi ward and statistical method of T-Test/ANOVA.



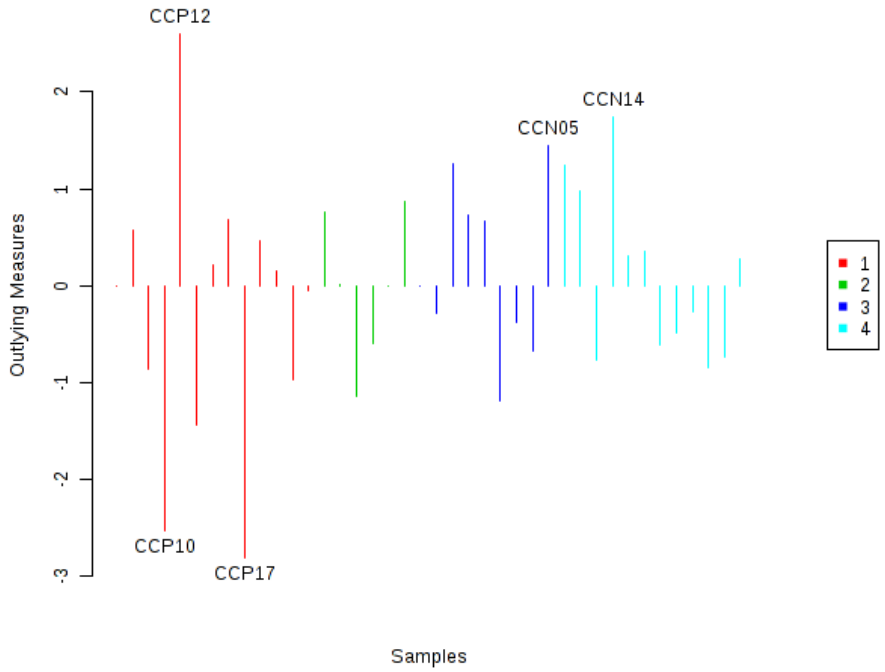
Appendix 2- 37. Metabolome Random Forest Classification Plot

The number of trees was set to 5000, with 7 predictors, and randomness turned on.



Appendix 2- 38. Metabolome Random Forest Results

The number of trees was set to 5000, with 7 predictors, and randomness turned on. The OOB error is equal to .625.



Appendix 2- 39. Metabolome Random Forest Outlier Detection Results

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by diarrhea_present		
	diarrhea_present(diarrhea_present)		
Case	0	1	Total
No	12 30.00 60.00 66.67	8 20.00 40.00 36.36	20 50.00
Yes	6 15.00 30.00 33.33	14 35.00 70.00 63.64	20 50.00
Total	18 45.00	22 55.00	40 100.00

Statistics for Table of Case by diarrhea_present

Statistic	DF	Value	Prob
Chi-Square	1	3.6364	0.0565
Likelihood Ratio Chi-Square	1	3.6961	0.0545
Continuity Adj. Chi-Square	1	2.5253	0.1120
Mantel-Haenszel Chi-Square	1	3.5455	0.0597
Phi Coefficient		0.3015	
Contingency Coefficient		0.2887	
Cramer's V		0.3015	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.9876
Right-sided Pr >= F	0.0555
Table Probability (P)	0.0431
Two-sided Pr <= P	0.1110

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.5000	0.9448	12.9659
Relative Risk (Column 1)	2.0000	0.9362	4.2727
Relative Risk (Column 2)	0.5714	0.3109	1.0502

Sample Size = 40

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by Parent_saw_blood_number					
	Parent_saw_blood_number(Parent_saw_blood_number)				
Case	0	1	2	3	Total
No	17 42.50 85.00 54.84	3 7.50 15.00 42.86	0 0.00 0.00 0.00	0 0.00 0.00 0.00	20 50.00
Yes	14 35.00 70.00 45.16	4 10.00 20.00 57.14	1 2.50 5.00 100.00	1 2.50 5.00 100.00	20 50.00
Total	31 77.50	7 17.50	1 2.50	1 2.50	40 100.00

Statistics for Table of Case by Parent_saw_blood_number

Statistic	DF	Value	Prob
Chi-Square	3	2.4332	0.4875
Likelihood Ratio Chi-Square	3	3.2067	0.3608
Mantel-Haenszel Chi-Square	1	2.1402	0.1435
Phi Coefficient		0.2466	
Contingency Coefficient		0.2395	
Cramer's V		0.2466	
WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by vomiting			
	vomiting(vomiting)		
Case	0	1	Total
No	13 32.50 65.00 56.52	7 17.50 35.00 41.18	20 50.00
Yes	10 25.00 50.00 43.48	10 25.00 50.00 58.82	20 50.00
Total	23 57.50	17 42.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by fever

Statistic	DF	Value	Prob
Chi-Square	1	3.6364	0.0565
Likelihood Ratio Chi-Square	1	3.6961	0.0545
Continuity Adj. Chi-Square	1	2.5253	0.1120
Mantel-Haenszel Chi-Square	1	3.5455	0.0597
Phi Coefficient		0.3015	
Contingency Coefficient		0.2887	
Cramer's V		0.3015	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.9876
Right-sided Pr >= F	0.0555
Table Probability (P)	0.0431
Two-sided Pr <= P	0.1110

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.5000	0.9448	12.9659
Relative Risk (Column 1)	2.0000	0.9362	4.2727
Relative Risk (Column 2)	0.5714	0.3109	1.0502

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by race			
	race(race)		
Case	0	1	Total
No	6 15.00 30.00 50.00	14 35.00 70.00 50.00	20 50.00
Yes	6 15.00 30.00 50.00	14 35.00 70.00 50.00	20 50.00
Total	12 30.00	28 70.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by antidiarrheal_14days

Statistic	DF	Value	Prob
Chi-Square	1	2.8490	0.0914
Likelihood Ratio Chi-Square	1	2.9048	0.0883
Continuity Adj. Chi-Square	1	1.8234	0.1769
Mantel-Haenszel Chi-Square	1	2.7778	0.0956
Phi Coefficient		0.2669	
Contingency Coefficient		0.2579	
Cramer's V		0.2669	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.9796
Right-sided Pr >= F	0.0880
Table Probability (P)	0.0676
Two-sided Pr <= P	0.1760

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.2727	0.8023	13.3499
Relative Risk (Column 1)	1.4545	0.9247	2.2879
Relative Risk (Column 2)	0.4444	0.1633	1.2100

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by bloody_stool			
	bloody_stool(bloody_stool)		
Case	0	1	Total
No	16 40.00 80.00 50.00	4 10.00 20.00 50.00	20 50.00
Yes	16 40.00 80.00 50.00	4 10.00 20.00 50.00	20 50.00
Total	32 80.00	8 20.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by bloody_stool

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.6526
Right-sided Pr >= F	0.6526
Table Probability (P)	0.3052
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.2124	4.7091
Relative Risk (Column 1)	1.0000	0.7335	1.3633
Relative Risk (Column 2)	1.0000	0.2895	3.4542

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by mucus_stool			
	mucus_stool(mucus_stool)		
Case	0	1	Total
No	10 25.00 50.00 66.67	10 25.00 50.00 40.00	20 50.00
Yes	5 12.50 25.00 33.33	15 37.50 75.00 60.00	20 50.00
Total	15 37.50	25 62.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by race

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	6
Left-sided Pr <= F	0.6345
Right-sided Pr >= F	0.6345
Table Probability (P)	0.2689
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.2586	3.8671
Relative Risk (Column 1)	1.0000	0.3880	2.5773
Relative Risk (Column 2)	1.0000	0.6665	1.5004

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by antidiarrheal_14days			
	antidiarrheal_14days(antidiarrheal_14days)		
Case	0	1	Total
No	16 40.00 80.00 59.26	4 10.00 20.00 30.77	20 50.00
Yes	11 27.50 55.00 40.74	9 22.50 45.00 69.23	20 50.00
Total	27 67.50	13 32.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by mucus_stool

Statistic	DF	Value	Prob
Chi-Square	1	2.6667	0.1025
Likelihood Ratio Chi-Square	1	2.7058	0.1000
Continuity Adj. Chi-Square	1	1.7067	0.1914
Mantel-Haenszel Chi-Square	1	2.6000	0.1069
Phi Coefficient		0.2582	
Contingency Coefficient		0.2500	
Cramer's V		0.2582	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	10
Left-sided Pr <= F	0.9758
Right-sided Pr >= F	0.0954
Table Probability (P)	0.0712
Two-sided Pr <= P	0.1908

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.0000	0.7864	11.4447
Relative Risk (Column 1)	2.0000	0.8325	4.8051
Relative Risk (Column 2)	0.6667	0.4019	1.1058

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by stoolwbc			
	stoolwbc(stoolwbc)		
Case	0	1	Total
No	12 30.00 60.00 63.16	8 20.00 40.00 38.10	20 50.00
Yes	7 17.50 35.00 36.84	13 32.50 65.00 61.90	20 50.00
Total	19 47.50	21 52.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by stoolwbc

Statistic	DF	Value	Prob
Chi-Square	1	2.5063	0.1134
Likelihood Ratio Chi-Square	1	2.5334	0.1115
Continuity Adj. Chi-Square	1	1.6040	0.2053
Mantel-Haenszel Chi-Square	1	2.4436	0.1180
Phi Coefficient		0.2503	
Contingency Coefficient		0.2428	
Cramer's V		0.2503	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.9719
Right-sided Pr >= F	0.1025
Table Probability (P)	0.0744
Two-sided Pr <= P	0.2049

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.7857	0.7727	10.0434
Relative Risk (Column 1)	1.7143	0.8545	3.4392
Relative Risk (Column 2)	0.6154	0.3292	1.1505

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by entamoeba_hystolitica		
	entamoeba_hystolitica(entamoeba_hystolitica)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by crypto

Case	crypto(crypto) 0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency
Percent
Row Pct
Col Pct

Table of Case by giardia

Case	giardia(giardia) 0	1	Total
No	20 50.00 100.00 52.63	0 0.00 0.00 0.00	20 50.00
Yes	18 45.00 90.00 47.37	2 5.00 10.00 100.00	20 50.00
Total	38 95.00	2 5.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by giardia

Statistic	DF	Value	Prob
Chi-Square	1	2.1053	0.1468
Likelihood Ratio Chi-Square	1	2.8779	0.0898
Continuity Adj. Chi-Square	1	0.5263	0.4682
Mantel-Haenszel Chi-Square	1	2.0526	0.1519
Phi Coefficient		0.2294	
Contingency Coefficient		0.2236	
Cramer's V		0.2294	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.2436
Table Probability (P)	0.2436
Two-sided Pr <= P	0.4872

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	1.1111	0.9601	1.2859
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by entcoli		
	entcoli(entcoli)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Table of Case by enthart

Frequency Percent Row Pct Col Pct	enthart(enthart)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Table of Case by endona

Frequency Percent Row Pct Col Pct	endona(endona)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Table of Case by blastocytis

Frequency Percent Row Pct Col Pct	blastocytis(blastocytis)		
Case	0	1	Total
No	18 45.00 90.00 47.37	2 5.00 10.00 100.00	20 50.00
Yes	20 50.00 100.00 52.63	0 0.00 0.00 0.00	20 50.00
Total	38 95.00	2 5.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by blastocytis

Statistic	DF	Value	Prob
Chi-Square	1	2.1053	0.1468
Likelihood Ratio Chi-Square	1	2.8779	0.0898
Continuity Adj. Chi-Square	1	0.5263	0.4682
Mantel-Haenszel Chi-Square	1	2.0526	0.1519
Phi Coefficient		-0.2294	
Contingency Coefficient		0.2236	
Cramer's V		-0.2294	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.2436
Right-sided Pr >= F	1.0000
Table Probability (P)	0.2436
Two-sided Pr <= P	0.4872

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	0.9000	0.7777	1.0416
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by chilome		
	chilome(chilome)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by shigsp	
	shigsp(shigsp)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by salmsp	
	salmsp(salmsp)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by yersent	
	yersent(yersent)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by campspp		
	campspp(campspp)		
Case	0	1	Total
No	20 50.00 100.00 100.00	0 0.00 0.00 0.00	20 50.00
Yes	0 0.00 0.00 0.00	20 50.00 100.00 100.00	20 50.00
Total	20 50.00	20 50.00	40 100.00

Statistics for Table of Case by campspp

Statistic	DF	Value	Prob
Chi-Square	1	40.0000	<.0001
Likelihood Ratio Chi-Square	1	55.4518	<.0001
Continuity Adj. Chi-Square	1	36.1000	<.0001
Mantel-Haenszel Chi-Square	1	39.0000	<.0001
Phi Coefficient		1.0000	
Contingency Coefficient		0.7071	
Cramer's V		1.0000	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

One or more statistics not computed – zero cell.

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by ecoli		
	ecoli(ecoli)		
Case	0	1	Total
No	7 17.50 35.00 77.78	13 32.50 65.00 41.94	20 50.00
Yes	2 5.00 10.00 22.22	18 45.00 90.00 58.06	20 50.00
Total	9 22.50	31 77.50	40 100.00

Statistics for Table of Case by ecoli

Statistic	DF	Value	Prob
Chi-Square	1	3.5842	0.0583
Likelihood Ratio Chi-Square	1	3.7519	0.0527
Continuity Adj. Chi-Square	1	2.2939	0.1299
Mantel-Haenszel Chi-Square	1	3.4946	0.0616
Phi Coefficient		0.2993	
Contingency Coefficient		0.2868	
Cramer's V		0.2993	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	7
Left-sided Pr <= F	0.9902
Right-sided Pr >= F	0.0637
Table Probability (P)	0.0539
Two-sided Pr <= P	0.1274

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	4.8462	0.8628	27.2212
Relative Risk (Column 1)	3.5000	0.8259	14.8328
Relative Risk (Column 2)	0.7222	0.5073	1.0282

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by pathogenicecoli		
	pathogenicecoli(pathogenicecoli)		
Case	0	1	Total
No	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Yes	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Total	36 90.00	4 10.00	40 100.00

Statistics for Table of Case by pathogenicecoli

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.6975
Right-sided Pr >= F	0.6975
Table Probability (P)	0.3950
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.1267	7.8931
Relative Risk (Column 1)	1.0000	0.8133	1.2295
Relative Risk (Column 2)	1.0000	0.1558	6.4198

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by pathotypeno		
	pathotypeno(pathotypeno)		
Case	0	1	Total
No	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Yes	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Total	36 90.00	4 10.00	40 100.00

Statistics for Table of Case by pathotypeno

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.6975
Right-sided Pr >= F	0.6975
Table Probability (P)	0.3950
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.1267	7.8931
Relative Risk (Column 1)	1.0000	0.8133	1.2295
Relative Risk (Column 2)	1.0000	0.1558	6.4198

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by pathotypeother	
	pathotypeother(pathotypeother)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by ecolitype		
	ecolitype(ecolitype)		
Case	0	1	Total
No	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Yes	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Total	36 90.00	4 10.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by ecolitype

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.6975
Right-sided Pr >= F	0.6975
Table Probability (P)	0.3950
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.1267	7.8931
Relative Risk (Column 1)	1.0000	0.8133	1.2295
Relative Risk (Column 2)	1.0000	0.1558	6.4198

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by emergecoli		
	emergecoli(emergecoli)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by norovirus		
	norovirus(norovirus)		
Case	0	1	Total
No	20 50.00 100.00 54.05	0 0.00 0.00 0.00	20 50.00
Yes	17 42.50 85.00 45.95	3 7.50 15.00 100.00	20 50.00
Total	37 92.50	3 7.50	40 100.00

Statistics for Table of Case by norovirus

Statistic	DF	Value	Prob
Chi-Square	1	3.2432	0.0717
Likelihood Ratio Chi-Square	1	4.4024	0.0359
Continuity Adj. Chi-Square	1	1.4414	0.2299
Mantel-Haenszel Chi-Square	1	3.1622	0.0754
Phi Coefficient		0.2847	
Contingency Coefficient		0.2739	
Cramer's V		0.2847	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.1154
Table Probability (P)	0.1154
Two-sided Pr <= P	0.2308

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	1.1765	0.9786	1.4143
One or more statistics not computed -- zero cell.			

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by norogenoty		
	norogenoty(norogenoty)		
Case	0	2	Total
No	20 50.00 100.00 54.05	0 0.00 0.00 0.00	20 50.00
Yes	17 42.50 85.00 45.95	3 7.50 15.00 100.00	20 50.00
Total	37 92.50	3 7.50	40 100.00

Statistics for Table of Case by norogenoty

Statistic	DF	Value	Prob
Chi-Square	1	3.2432	0.0717
Likelihood Ratio Chi-Square	1	4.4024	0.0359
Continuity Adj. Chi-Square	1	1.4414	0.2299
Mantel-Haenszel Chi-Square	1	3.1622	0.0754
Phi Coefficient		0.2847	
Contingency Coefficient		0.2739	
Cramer's V		0.2847	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.1154
Table Probability (P)	0.1154
Two-sided Pr <= P	0.2308

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	1.1765	0.9786	1.4143
One or more statistics not computed -- zero cell.			

Sample Size = 40

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by rotavirus

Case	rotavirus(rotavirus) 0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency
Percent
Row Pct
Col Pct

Table of Case by sapovirus

Case	sapovirus(sapovirus) 0	1	Total
No	20 50.00 100.00 51.28	0 0.00 0.00 0.00	20 50.00
Yes	19 47.50 95.00 48.72	1 2.50 5.00 100.00	20 50.00
Total	39 97.50	1 2.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by sapovirus

Statistic	DF	Value	Prob
Chi-Square	1	1.0256	0.3112
Likelihood Ratio Chi-Square	1	1.4119	0.2347
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	1.0000	0.3173
Phi Coefficient		0.1601	
Contingency Coefficient		0.1581	
Cramer's V		0.1601	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.5000
Table Probability (P)	0.5000
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	1.0526	0.9519	1.1640
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by astrovirus		
	astrovirus(astrovirus)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by adenovirus		
	adenovirus(adenovirus)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency
Percent
Row Pct
Col Pct

Table of Case by abdominalpain			
	abdominalpain(abdominalpain)		
Case	0	1	Total
No	15 37.50 75.00 51.72	5 12.50 25.00 45.45	20 50.00
Yes	14 35.00 70.00 48.28	6 15.00 30.00 54.55	20 50.00
Total	29 72.50	11 27.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by abdominalpain

Statistic	DF	Value	Prob
Chi-Square	1	0.1254	0.7233
Likelihood Ratio Chi-Square	1	0.1255	0.7231
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.1223	0.7266
Phi Coefficient		0.0560	
Contingency Coefficient		0.0559	
Cramer's V		0.0560	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	15
Left-sided Pr <= F	0.7599
Right-sided Pr >= F	0.5000
Table Probability (P)	0.2599
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.2857	0.3194	5.1748
Relative Risk (Column 1)	1.0714	0.7308	1.5707
Relative Risk (Column 2)	0.8333	0.3029	2.2929

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by breastfeedinghx			
	breastfeedinghx(breastfeedinghx)		
Case	0	1	Total
No	1 2.50 5.00 50.00	19 47.50 95.00 50.00	20 50.00
Yes	1 2.50 5.00 50.00	19 47.50 95.00 50.00	20 50.00
Total	2 5.00	38 95.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by breastfeedinghx

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	1
Left-sided Pr <= F	0.7564
Right-sided Pr >= F	0.7564
Table Probability (P)	0.5128
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.0582	17.1812
Relative Risk (Column 1)	1.0000	0.0671	14.9039
Relative Risk (Column 2)	1.0000	0.8675	1.1528

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by vaccinerota			
	vaccinerota(vaccinerota)		
Case	0	1	Total
No	2 5.00 10.00 100.00	18 45.00 90.00 47.37	20 50.00
Yes	0 0.00 0.00 0.00	20 50.00 100.00 52.63	20 50.00
Total	2 5.00	38 95.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by vaccinerota

Statistic	DF	Value	Prob
Chi-Square	1	2.1053	0.1468
Likelihood Ratio Chi-Square	1	2.8779	0.0898
Continuity Adj. Chi-Square	1	0.5263	0.4682
Mantel-Haenszel Chi-Square	1	2.0526	0.1519
Phi Coefficient		0.2294	
Contingency Coefficient		0.2236	
Cramer's V		0.2294	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.2436
Table Probability (P)	0.2436
Two-sided Pr <= P	0.4872

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 2)	0.9000	0.7777	1.0416
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by vaccinerotadoseno				
	vaccinerotadoseno(vaccinerotadoseno)			
Case	0	1	2	Total
No	5 13.51 29.41 83.33	11 29.73 64.71 36.67	1 2.70 5.88 100.00	17 45.95
Yes	1 2.70 5.00 16.67	19 51.35 95.00 63.33	0 0.00 0.00 0.00	20 54.05
Total	6 16.22	30 81.08	1 2.70	37 100.00
Frequency Missing = 3				

The FREQ Procedure

Statistics for Table of Case by vaccinerotadoseno

Statistic	DF	Value	Prob
Chi-Square	2	5.5935	0.0610
Likelihood Ratio Chi-Square	2	6.2132	0.0448
Mantel-Haenszel Chi-Square	1	1.7959	0.1802
Phi Coefficient		0.3888	
Contingency Coefficient		0.3624	
Cramer's V		0.3888	
WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

One or more statistics not computed – zero cell.

Effective Sample Size = 37
Frequency Missing = 3

Frequency Percent Row Pct Col Pct	Table of Case by race1		
	race1(race1)		
Case	0	1	Total
No	6 15.00 30.00 50.00	14 35.00 70.00 50.00	20 50.00
Yes	6 15.00 30.00 50.00	14 35.00 70.00 50.00	20 50.00
Total	12 30.00	28 70.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by race1

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	6
Left-sided Pr <= F	0.6345
Right-sided Pr >= F	0.6345
Table Probability (P)	0.2689
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.2586	3.8671
Relative Risk (Column 1)	1.0000	0.3880	2.5773
Relative Risk (Column 2)	1.0000	0.6665	1.5004

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by anyanimalexpo			
	anyanimalexpo(anyanimalexpo)		
Case	0	1	Total
No	12 30.00 60.00 54.55	8 20.00 40.00 44.44	20 50.00
Yes	10 25.00 50.00 45.45	10 25.00 50.00 55.56	20 50.00
Total	22 55.00	18 45.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by anyanimalexpo

Statistic	DF	Value	Prob
Chi-Square	1	0.4040	0.5250
Likelihood Ratio Chi-Square	1	0.4048	0.5246
Continuity Adj. Chi-Square	1	0.1010	0.7506
Mantel-Haenszel Chi-Square	1	0.3939	0.5302
Phi Coefficient		0.1005	
Contingency Coefficient		0.1000	
Cramer's V		0.1005	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.8297
Right-sided Pr >= F	0.3756
Table Probability (P)	0.2053
Two-sided Pr <= P	0.7512

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.5000	0.4287	5.2483
Relative Risk (Column 1)	1.2000	0.6815	2.1130
Relative Risk (Column 2)	0.8000	0.4001	1.5997

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by birdexpo			
	birdexpo(birdexpo)		
Case	0	1	Total
No	13 32.50 65.00 39.39	7 17.50 35.00 100.00	20 50.00
Yes	20 50.00 100.00 60.61	0 0.00 0.00 0.00	20 50.00
Total	33 82.50	7 17.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by birdexpo

Statistic	DF	Value	Prob
Chi-Square	1	8.4848	0.0036
Likelihood Ratio Chi-Square	1	11.2002	0.0008
Continuity Adj. Chi-Square	1	6.2338	0.0125
Mantel-Haenszel Chi-Square	1	8.2727	0.0040
Phi Coefficient		-0.4606	
Contingency Coefficient		0.4183	
Cramer's V		-0.4606	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	13
Left-sided Pr <= F	0.0042
Right-sided Pr >= F	1.0000
Table Probability (P)	0.0042
Two-sided Pr <= P	0.0083

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	0.6500	0.4712	0.8966
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by catexpo			
	catexpo(catexpo)		
Case	0	1	Total
No	18 45.00 90.00 58.06	2 5.00 10.00 22.22	20 50.00
Yes	13 32.50 65.00 41.94	7 17.50 35.00 77.78	20 50.00
Total	31 77.50	9 22.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by catexpo

Statistic	DF	Value	Prob
Chi-Square	1	3.5842	0.0583
Likelihood Ratio Chi-Square	1	3.7519	0.0527
Continuity Adj. Chi-Square	1	2.2939	0.1299
Mantel-Haenszel Chi-Square	1	3.4946	0.0616
Phi Coefficient		0.2993	
Contingency Coefficient		0.2868	
Cramer's V		0.2993	

WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.9902
Right-sided Pr >= F	0.0637
Table Probability (P)	0.0539
Two-sided Pr <= P	0.1274

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	4.8462	0.8628	27.2212
Relative Risk (Column 1)	1.3846	0.9726	1.9712
Relative Risk (Column 2)	0.2857	0.0674	1.2108

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by dogexpo			
	dogexpo(dogexpo)		
Case	0	1	Total
No	12 30.00 60.00 50.00	8 20.00 40.00 50.00	20 50.00
Yes	12 30.00 60.00 50.00	8 20.00 40.00 50.00	20 50.00
Total	24 60.00	16 40.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by dogexpo

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	12
Left-sided Pr <= F	0.6262
Right-sided Pr >= F	0.6262
Table Probability (P)	0.2525
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.2822	3.5436
Relative Risk (Column 1)	1.0000	0.6029	1.6587
Relative Risk (Column 2)	1.0000	0.4681	2.1363

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by chickenexpo			
	chickenexpo(chickenexpo)		
Case	0	1	Total
No	18 45.00 90.00 51.43	2 5.00 10.00 40.00	20 50.00
Yes	17 42.50 85.00 48.57	3 7.50 15.00 60.00	20 50.00
Total	35 87.50	5 12.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by chickenexpo

Statistic	DF	Value	Prob
Chi-Square	1	0.2286	0.6326
Likelihood Ratio Chi-Square	1	0.2299	0.6316
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.2229	0.6369
Phi Coefficient		0.0756	
Contingency Coefficient		0.0754	
Cramer's V		0.0756	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.8292
Right-sided Pr >= F	0.5000
Table Probability (P)	0.3292
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.5882	0.2356	10.7044
Relative Risk (Column 1)	1.0588	0.8371	1.3393
Relative Risk (Column 2)	0.6667	0.1244	3.5714

Sample Size = 40

Frequency Percent Row Pct Col Pct	Table of Case by pigexpo	
	pigexpo(pigexpo)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by cowsexpo	
	cowsexpo(cowsexpo)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by reptilesexpo	
	reptilesexpo(reptilesexpo)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by rodentexpo	
	rodentexpo(rodentexpo)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by otheranimalexpo	
	otheranimalexpo(otheranimalexpo)	
Case	0	Total
No	20 50.00 100.00 50.00	20 50.00
Yes	20 50.00 100.00 50.00	20 50.00
Total	40 100.00	40 100.00

Frequency Percent Row Pct Col Pct	Table of Case by farmanimalexpo		
	farmanimalexpo(farmanimalexpo)		
Case	0	1	Total
No	19 47.50 95.00 52.78	1 2.50 5.00 25.00	20 50.00
Yes	17 42.50 85.00 47.22	3 7.50 15.00 75.00	20 50.00
Total	36 90.00	4 10.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by farmanimalexpo

Statistic	DF	Value	Prob
Chi-Square	1	1.1111	0.2918
Likelihood Ratio Chi-Square	1	1.1577	0.2820
Continuity Adj. Chi-Square	1	0.2778	0.5982
Mantel-Haenszel Chi-Square	1	1.0833	0.2980
Phi Coefficient		0.1667	
Contingency Coefficient		0.1644	
Cramer's V		0.1667	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	19
Left-sided Pr <= F	0.9470
Right-sided Pr >= F	0.3025
Table Probability (P)	0.2495
Two-sided Pr <= P	0.6050

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.3529	0.3179	35.3643
Relative Risk (Column 1)	1.1176	0.9062	1.3785
Relative Risk (Column 2)	0.3333	0.0378	2.9389

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by travelhx			
	travelhx(travelhx)		
Case	0	1	Total
No	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Yes	18 45.00 90.00 50.00	2 5.00 10.00 50.00	20 50.00
Total	36 90.00	4 10.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by travelhx

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.6975
Right-sided Pr >= F	0.6975
Table Probability (P)	0.3950
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.1267	7.8931
Relative Risk (Column 1)	1.0000	0.8133	1.2295
Relative Risk (Column 2)	1.0000	0.1558	6.4198

Sample Size = 40

Frequency Percent Row Pct Col Pct

Table of Case by outsiderexpo1			
	outsiderexpo1(outsiderexpo1)		
Case	0	1	Total
No	16 40.00 80.00 53.33	4 10.00 20.00 40.00	20 50.00
Yes	14 35.00 70.00 46.67	6 15.00 30.00 60.00	20 50.00
Total	30 75.00	10 25.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by outsiderexpo1

Statistic	DF	Value	Prob
Chi-Square	1	0.5333	0.4652
Likelihood Ratio Chi-Square	1	0.5361	0.4640
Continuity Adj. Chi-Square	1	0.1333	0.7150
Mantel-Haenszel Chi-Square	1	0.5200	0.4708
Phi Coefficient		0.1155	
Contingency Coefficient		0.1147	
Cramer's V		0.1155	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.8633
Right-sided Pr >= F	0.3582
Table Probability (P)	0.2215
Two-sided Pr <= P	0.7164

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.7143	0.4004	7.3399
Relative Risk (Column 1)	1.1429	0.7965	1.6398
Relative Risk (Column 2)	0.6667	0.2213	2.0087

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by diarrheaexpo			
	diarrheaexpo(diarrheaexpo)		
Case	0	1	Total
No	20 50.00 100.00 51.28	0 0.00 0.00 0.00	20 50.00
Yes	19 47.50 95.00 48.72	1 2.50 5.00 100.00	20 50.00
Total	39 97.50	1 2.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by diarrheaexpo

Statistic	DF	Value	Prob
Chi-Square	1	1.0256	0.3112
Likelihood Ratio Chi-Square	1	1.4119	0.2347
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	1.0000	0.3173
Phi Coefficient		0.1601	
Contingency Coefficient		0.1581	
Cramer's V		0.1601	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	20
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	0.5000
Table Probability (P)	0.5000
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Relative Risk (Column 1)	1.0526	0.9519	1.1640
One or more statistics not computed -- zero cell.			

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by inhomediarrheaexpo			
	inhomediarrheaexpo(inhomediarrheaexpo)		
Case	0	1	Total
No	19 47.50 95.00 50.00	1 2.50 5.00 50.00	20 50.00
Yes	19 47.50 95.00 50.00	1 2.50 5.00 50.00	20 50.00
Total	38 95.00	2 5.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by inhomediarrheaexpo

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	19
Left-sided Pr <= F	0.7564
Right-sided Pr >= F	0.7564
Table Probability (P)	0.5128
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.0582	17.1812
Relative Risk (Column 1)	1.0000	0.8675	1.1528
Relative Risk (Column 2)	1.0000	0.0671	14.9039

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by daycareexpo			
	daycareexpo(daycareexpo)		
Case	0	1	Total
No	15 37.50 75.00 48.39	5 12.50 25.00 55.56	20 50.00
Yes	16 40.00 80.00 51.61	4 10.00 20.00 44.44	20 50.00
Total	31 77.50	9 22.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by daycareexpo

Statistic	DF	Value	Prob
Chi-Square	1	0.1434	0.7050
Likelihood Ratio Chi-Square	1	0.1436	0.7047
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.1398	0.7085
Phi Coefficient		-0.0599	
Contingency Coefficient		0.0598	
Cramer's V		-0.0599	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	15
Left-sided Pr <= F	0.5000
Right-sided Pr >= F	0.7747
Table Probability (P)	0.2747
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.7500	0.1688	3.3331
Relative Risk (Column 1)	0.9375	0.6708	1.3102
Relative Risk (Column 2)	1.2500	0.3920	3.9855

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by diaperuse			
	diaperuse(diaperuse)		
Case	0	1	Total
No	6 15.00 30.00 66.67	14 35.00 70.00 45.16	20 50.00
Yes	3 7.50 15.00 33.33	17 42.50 85.00 54.84	20 50.00
Total	9 22.50	31 77.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by diaperuse

Statistic	DF	Value	Prob
Chi-Square	1	1.2903	0.2560
Likelihood Ratio Chi-Square	1	1.3102	0.2524
Continuity Adj. Chi-Square	1	0.5735	0.4489
Mantel-Haenszel Chi-Square	1	1.2581	0.2620
Phi Coefficient		0.1796	
Contingency Coefficient		0.1768	
Cramer's V		0.1796	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	6
Left-sided Pr <= F	0.9363
Right-sided Pr >= F	0.2253
Table Probability (P)	0.1616
Two-sided Pr <= P	0.4506

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.4286	0.5124	11.5108
Relative Risk (Column 1)	2.0000	0.5790	6.9084
Relative Risk (Column 2)	0.8235	0.5856	1.1581

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by houseincome					
	houseincome(houseincome)				
Case	0	1	2	3	Total
No	2 5.00 10.00 100.00	5 12.50 25.00 83.33	12 30.00 60.00 40.00	1 2.50 5.00 50.00	20 50.00
Yes	0 0.00 0.00 0.00	1 2.50 5.00 16.67	18 45.00 90.00 60.00	1 2.50 5.00 50.00	20 50.00
Total	2 5.00	6 15.00	30 75.00	2 5.00	40 100.00

The FREQ Procedure

Statistics for Table of Case by houseincome

Statistic	DF	Value	Prob
Chi-Square	3	5.8667	0.1183
Likelihood Ratio Chi-Square	3	6.8918	0.0754
Mantel-Haenszel Chi-Square	1	4.3333	0.0374
Phi Coefficient		0.3830	
Contingency Coefficient		0.3576	
Cramer's V		0.3830	
WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by zone

Case	zone(zone) 0	1	2	3	4	Total
No	2 5.00 10.00 100.00	2 5.00 10.00 25.00	9 22.50 45.00 64.29	7 17.50 35.00 46.67	0 0.00 0.00 0.00	20 50.00
Yes	0 0.00 0.00 0.00	6 15.00 30.00 75.00	5 12.50 25.00 35.71	8 20.00 40.00 53.33	1 2.50 5.00 100.00	20 50.00
Total	2 5.00	8 20.00	14 35.00	15 37.50	1 2.50	40 100.00

Statistics for Table of Case by zone

Statistic	DF	Value	Prob
Chi-Square	4	6.2095	0.1840
Likelihood Ratio Chi-Square	4	7.4775	0.1127
Mantel-Haenszel Chi-Square	1	0.2553	0.6134
Phi Coefficient		0.3940	
Contingency Coefficient		0.3666	
Cramer's V		0.3940	
WARNING: 60% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by ownhouse		
	ownhouse(ownhouse)		
Case	0	1	Total
No	19 47.50 95.00 61.29	1 2.50 5.00 11.11	20 50.00
Yes	12 30.00 60.00 38.71	8 20.00 40.00 88.89	20 50.00
Total	31 77.50	9 22.50	40 100.00

Statistics for Table of Case by ownhouse

Statistic	DF	Value	Prob
Chi-Square	1	7.0251	0.0080
Likelihood Ratio Chi-Square	1	7.7920	0.0052
Continuity Adj. Chi-Square	1	5.1613	0.0231
Mantel-Haenszel Chi-Square	1	6.8495	0.0089
Phi Coefficient		0.4191	
Contingency Coefficient		0.3865	
Cramer's V		0.4191	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	19
Left-sided Pr <= F	0.9994
Right-sided Pr >= F	0.0098
Table Probability (P)	0.0092
Two-sided Pr <= P	0.0197

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	12.6667	1.4022	114.4194
Relative Risk (Column 1)	1.5833	1.0918	2.2961
Relative Risk (Column 2)	0.1250	0.0172	0.9093

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by headgender		
	headgender(headgender)		
Case	0	1	Total
No	15 37.50 75.00 46.88	5 12.50 25.00 62.50	20 50.00
Yes	17 42.50 85.00 53.13	3 7.50 15.00 37.50	20 50.00
Total	32 80.00	8 20.00	40 100.00

Statistics for Table of Case by headgender

Statistic	DF	Value	Prob
Chi-Square	1	0.6250	0.4292
Likelihood Ratio Chi-Square	1	0.6304	0.4272
Continuity Adj. Chi-Square	1	0.1563	0.6926
Mantel-Haenszel Chi-Square	1	0.6094	0.4350
Phi Coefficient		-0.1250	
Contingency Coefficient		0.1240	
Cramer's V		-0.1250	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	15
Left-sided Pr <= F	0.3474
Right-sided Pr >= F	0.8824
Table Probability (P)	0.2298
Two-sided Pr <= P	0.6948

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.5294	0.1079	2.5983
Relative Risk (Column 1)	0.8824	0.6453	1.2065
Relative Risk (Column 2)	1.6667	0.4587	6.0559

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by headgender		
	headgender(headgender)		
Case	0	1	Total
No	15 37.50 75.00 46.88	5 12.50 25.00 62.50	20 50.00
Yes	17 42.50 85.00 53.13	3 7.50 15.00 37.50	20 50.00
Total	32 80.00	8 20.00	40 100.00

Statistics for Table of Case by headgender

Statistic	DF	Value	Prob
Chi-Square	1	0.6250	0.4292
Likelihood Ratio Chi-Square	1	0.6304	0.4272
Continuity Adj. Chi-Square	1	0.1563	0.6926
Mantel-Haenszel Chi-Square	1	0.6094	0.4350
Phi Coefficient		-0.1250	
Contingency Coefficient		0.1240	
Cramer's V		-0.1250	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	15
Left-sided Pr <= F	0.3474
Right-sided Pr >= F	0.8824
Table Probability (P)	0.2298
Two-sided Pr <= P	0.6948

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.5294	0.1079	2.5983
Relative Risk (Column 1)	0.8824	0.6453	1.2065
Relative Risk (Column 2)	1.6667	0.4587	6.0559

Sample Size = 40

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by edprimary

Case	edprimary(edprimary) 1	2	3	4	Total
No	3 7.69 15.00 42.86	10 25.64 50.00 52.63	4 10.26 20.00 66.67	3 7.69 15.00 42.86	20 51.28
Yes	4 10.26 21.05 57.14	9 23.08 47.37 47.37	2 5.13 10.53 33.33	4 10.26 21.05 57.14	19 48.72
Total	7 17.95	19 48.72	6 15.38	7 17.95	39 100.00

Frequency Missing = 1

Statistics for Table of Case by edprimary

Statistic	DF	Value	Prob
Chi-Square	3	0.9800	0.8061
Likelihood Ratio Chi-Square	3	0.9933	0.8029
Mantel-Haenszel Chi-Square	1	0.0118	0.9134
Phi Coefficient		0.1585	
Contingency Coefficient		0.1566	
Cramer's V		0.1585	

WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

Effective Sample Size = 39
Frequency Missing = 1

Frequency
Percent
Row Pct
Col Pct

Table of Case by edhead

Case	edhead(edhead) 1	2	3	4	Total
No	0 0.00 0.00 0.00	12 30.00 60.00 50.00	4 10.00 20.00 66.67	4 10.00 20.00 57.14	20 50.00
Yes	3 7.50 15.00 100.00	12 30.00 60.00 50.00	2 5.00 10.00 33.33	3 7.50 15.00 42.86	20 50.00
Total	3 7.50	24 60.00	6 15.00	7 17.50	40 100.00

The FREQ Procedure

Statistics for Table of Case by edhead

Statistic	DF	Value	Prob
Chi-Square	3	3.8095	0.2828
Likelihood Ratio Chi-Square	3	4.9818	0.1731
Mantel-Haenszel Chi-Square	1	1.6045	0.2053
Phi Coefficient		0.3086	
Contingency Coefficient		0.2949	
Cramer's V		0.3086	
WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

Frequency
Percent
Row Pct
Col Pct

Table of Case by headmari				
	headmari(headmari)			
Case	0	1	4	Total
No	4 10.00 20.00 66.67	10 25.00 50.00 71.43	6 15.00 30.00 30.00	20 50.00
Yes	2 5.00 10.00 33.33	4 10.00 20.00 28.57	14 35.00 70.00 70.00	20 50.00
Total	6 15.00	14 35.00	20 50.00	40 100.00

Statistics for Table of Case by headmari

Statistic	DF	Value	Prob
Chi-Square	2	6.4381	0.0400
Likelihood Ratio Chi-Square	2	6.6275	0.0364
Mantel-Haenszel Chi-Square	1	5.8276	0.0158
Phi Coefficient		0.4012	
Contingency Coefficient		0.3723	
Cramer's V		0.4012	
WARNING: 33% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by sex		
	sex(sex)		
Case	0	1	Total
No	14 35.00 70.00 53.85	6 15.00 30.00 42.86	20 50.00
Yes	12 30.00 60.00 46.15	8 20.00 40.00 57.14	20 50.00
Total	26 65.00	14 35.00	40 100.00

Statistics for Table of Case by sex

Statistic	DF	Value	Prob
Chi-Square	1	0.4396	0.5073
Likelihood Ratio Chi-Square	1	0.4407	0.5068
Continuity Adj. Chi-Square	1	0.1099	0.7403
Mantel-Haenszel Chi-Square	1	0.4286	0.5127
Phi Coefficient		0.1048	
Contingency Coefficient		0.1043	
Cramer's V		0.1048	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	14
Left-sided Pr <= F	0.8399
Right-sided Pr >= F	0.3705
Table Probability (P)	0.2104
Two-sided Pr <= P	0.7411

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.5556	0.4199	5.7626
Relative Risk (Column 1)	1.1667	0.7375	1.8456
Relative Risk (Column 2)	0.7500	0.3180	1.7689

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by prematurity		
	prematurity(prematurity)		
Case	0	1	Total
No	16 40.00 80.00 48.48	4 10.00 20.00 57.14	20 50.00
Yes	17 42.50 85.00 51.52	3 7.50 15.00 42.86	20 50.00
Total	33 82.50	7 17.50	40 100.00

Statistics for Table of Case by prematurity

Statistic	DF	Value	Prob
Chi-Square	1	0.1732	0.6773
Likelihood Ratio Chi-Square	1	0.1737	0.6769
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.1688	0.6812
Phi Coefficient		-0.0658	
Contingency Coefficient		0.0657	
Cramer's V		-0.0658	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.5000
Right-sided Pr >= F	0.7963
Table Probability (P)	0.2963
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.7059	0.1362	3.6582
Relative Risk (Column 1)	0.9412	0.7069	1.2531
Relative Risk (Column 2)	1.3333	0.3413	5.2085

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by fever1		
	fever1(fever1)		
Case	0	1	Total
No	19 47.50 95.00 51.35	1 2.50 5.00 33.33	20 50.00
Yes	18 45.00 90.00 48.65	2 5.00 10.00 66.67	20 50.00
Total	37 92.50	3 7.50	40 100.00

Statistics for Table of Case by fever1

Statistic	DF	Value	Prob
Chi-Square	1	0.3604	0.5483
Likelihood Ratio Chi-Square	1	0.3668	0.5447
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.3514	0.5533
Phi Coefficient		0.0949	
Contingency Coefficient		0.0945	
Cramer's V		0.0949	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	19
Left-sided Pr <= F	0.8846
Right-sided Pr >= F	0.5000
Table Probability (P)	0.3846
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.1111	0.1758	25.3487
Relative Risk (Column 1)	1.0556	0.8840	1.2604
Relative Risk (Column 2)	0.5000	0.0492	5.0831

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by insurance		
	insurance(insurance)		
Case	0	1	Total
No	2 5.00 10.00 66.67	18 45.00 90.00 48.65	20 50.00
Yes	1 2.50 5.00 33.33	19 47.50 95.00 51.35	20 50.00
Total	3 7.50	37 92.50	40 100.00

Statistics for Table of Case by insurance

Statistic	DF	Value	Prob
Chi-Square	1	0.3604	0.5483
Likelihood Ratio Chi-Square	1	0.3668	0.5447
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.3514	0.5533
Phi Coefficient		0.0949	
Contingency Coefficient		0.0945	
Cramer's V		0.0949	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	2
Left-sided Pr <= F	0.8846
Right-sided Pr >= F	0.5000
Table Probability (P)	0.3846
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	2.1111	0.1758	25.3487
Relative Risk (Column 1)	2.0000	0.1967	20.3322
Relative Risk (Column 2)	0.9474	0.7934	1.1312

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by micro_blood_stool		
	micro_blood_stool(micro_blood_stool)		
Case	0	1	Total
No	16 40.00 80.00 50.00	4 10.00 20.00 50.00	20 50.00
Yes	16 40.00 80.00 50.00	4 10.00 20.00 50.00	20 50.00
Total	32 80.00	8 20.00	40 100.00

Statistics for Table of Case by micro_blood_stool

Statistic	DF	Value	Prob
Chi-Square	1	0.0000	1.0000
Likelihood Ratio Chi-Square	1	0.0000	1.0000
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0000	1.0000
Phi Coefficient		0.0000	
Contingency Coefficient		0.0000	
Cramer's V		0.0000	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	16
Left-sided Pr <= F	0.6526
Right-sided Pr >= F	0.6526
Table Probability (P)	0.3052
Two-sided Pr <= P	1.0000

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.0000	0.2124	4.7091
Relative Risk (Column 1)	1.0000	0.7335	1.3633
Relative Risk (Column 2)	1.0000	0.2895	3.4542

Sample Size = 40

The TTEST Procedure

Variable: daysofdiarrhea (daysofdiarrhea)

Case	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	20	1.0500	1.7911	0.4005	0	7.0000
Yes	20	2.5000	2.5443	0.5689	0	9.0000
Diff (1-2)		-1.4500	2.2002	0.6958		

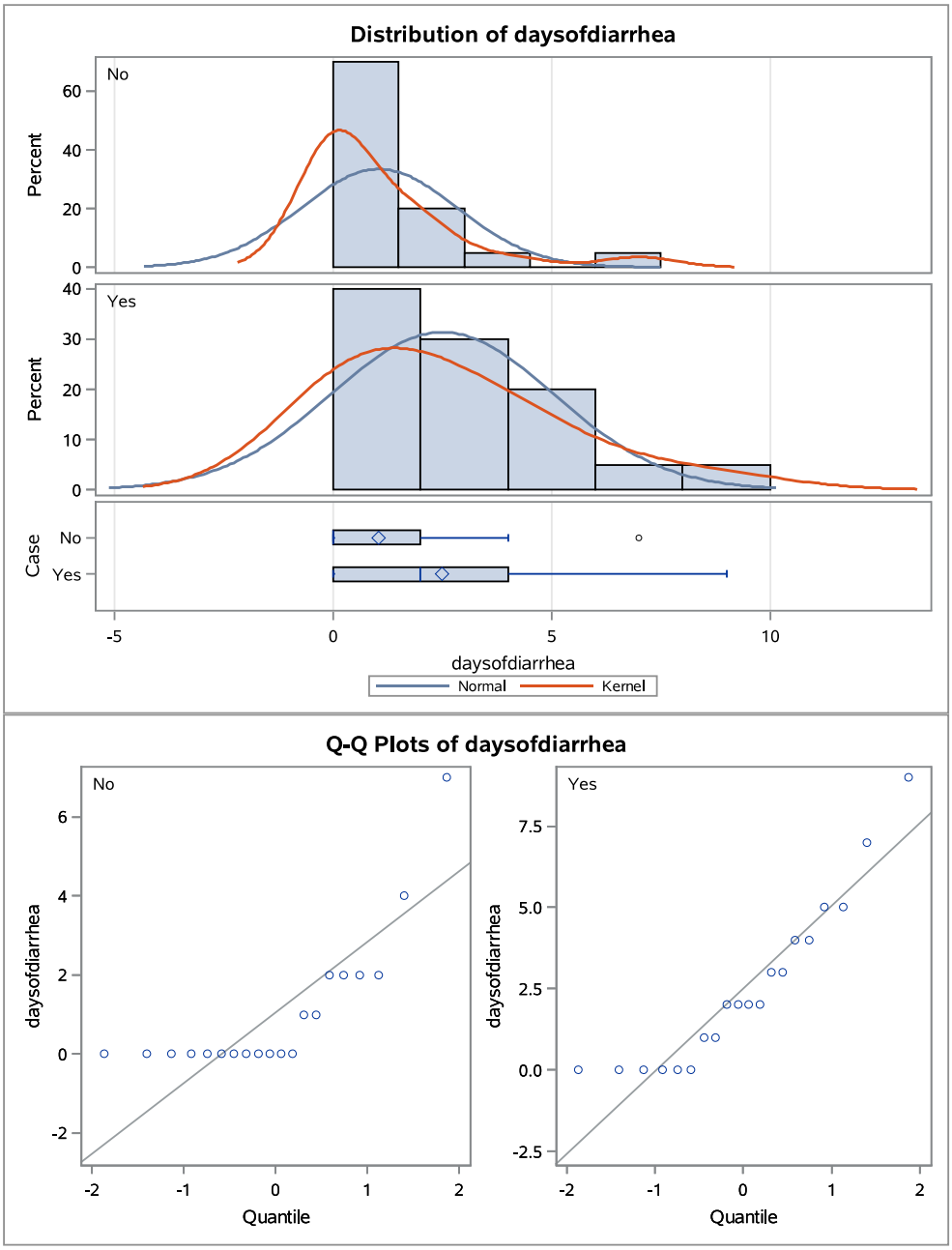
Case	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		1.0500	0.2118	1.8882	1.7911	1.3621	2.6160
Yes		2.5000	1.3092	3.6908	2.5443	1.9349	3.7162
Diff (1-2)	Pooled	-1.4500	-2.8585	-0.0415	2.2002	1.7981	2.8355
Diff (1-2)	Satterthwaite	-1.4500	-2.8638	-0.0362			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	-2.08	0.0439
Satterthwaite	Unequal	34.118	-2.08	0.0447

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	2.02	0.1348

The TTEST Procedure

Variable: daysofdiarrhea (daysofdiarrhea)



The TTEST Procedure

Variable: diarrea_episodes_24h (diarrea_episodes_24h)

Variable: diarrea_episodes_24h (diarrea_episodes_24h)

Case	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	20	5.2000	8.4143	1.8815	0	30.0000
Yes	20	6.4000	5.6326	1.2595	0	15.0000
Diff (1-2)		-1.2000	7.1598	2.2641		

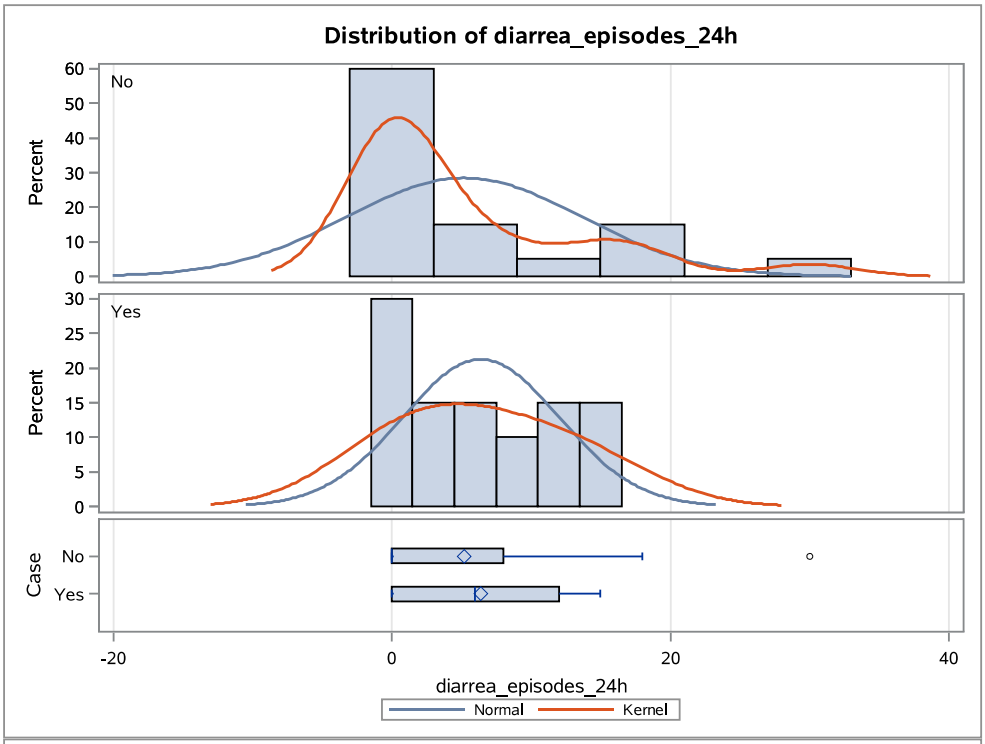
Case	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		5.2000	1.2620	9.1380	8.4143	6.3990	12.2896
Yes		6.4000	3.7639	9.0361	5.6326	4.2835	8.2268
Diff (1-2)	Pooled	-1.2000	-5.7835	3.3835	7.1598	5.8513	9.2274
Diff (1-2)	Satterthwaite	-1.2000	-5.8055	3.4055			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	-0.53	0.5992
Satterthwaite	Unequal	33.181	-0.53	0.5996

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	2.23	0.0883

The TTEST Procedure

Variable: diarrea_episodes_24h (diarrea_episodes_24h)



The TTEST Procedure

Variable: household_number (household_number)

Variable: household_number (household_number)

Case	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	20	5.6500	2.7391	0.6125	2.0000	14.0000
Yes	20	4.8500	1.6631	0.3719	3.0000	10.0000
Diff (1-2)		0.8000	2.2659	0.7165		

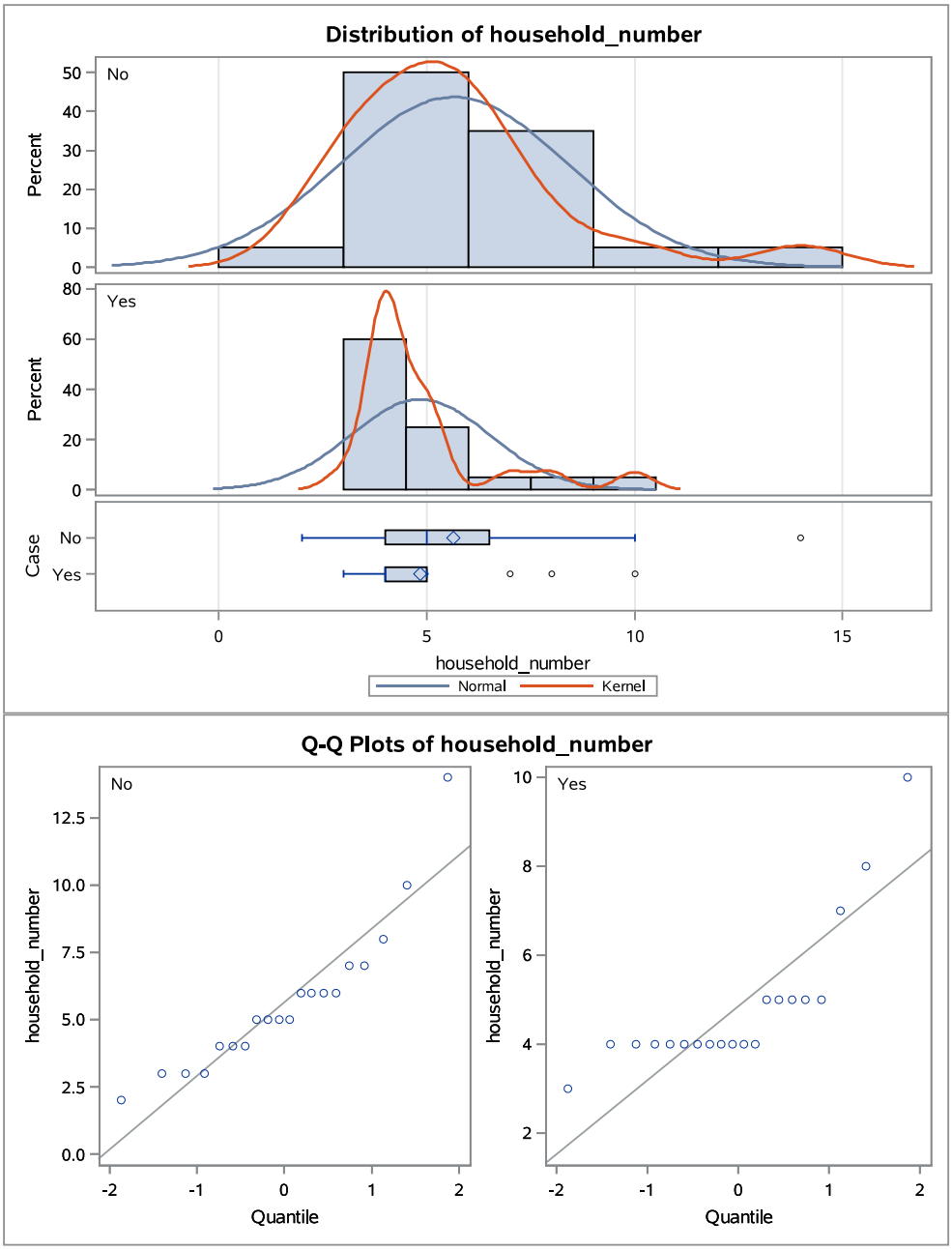
Case	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		5.6500	4.3681	6.9319	2.7391	2.0831	4.0006
Yes		4.8500	4.0717	5.6283	1.6631	1.2647	2.4290
Diff (1-2)	Pooled	0.8000	-0.6505	2.2505	2.2659	1.8518	2.9202
Diff (1-2)	Satterthwaite	0.8000	-0.6608	2.2608			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	1.12	0.2712
Satterthwaite	Unequal	31.332	1.12	0.2727

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	2.71	0.0353

The TTEST Procedure

Variable: household_number (household_number)



The TTEST Procedure

Variable: daycarekidsno (daycarekidsno)

Variable: daycarekidsno (daycarekidsno)

Case	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	20	0.6500	1.1821	0.2643	0	3.0000
Yes	20	0.5000	1.0513	0.2351	0	3.0000
Diff (1-2)		0.1500	1.1186	0.3537		

Case	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		0.6500	0.0968	1.2032	1.1821	0.8990	1.7265
Yes		0.5000	0.00797	0.9920	1.0513	0.7995	1.5355
Diff (1-2)	Pooled	0.1500	-0.5661	0.8661	1.1186	0.9142	1.4417
Diff (1-2)	Satterthwaite	0.1500	-0.5664	0.8664			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	0.42	0.6739
Satterthwaite	Unequal	37.489	0.42	0.6740

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	1.26	0.6144

The TTEST Procedure

Variable: motherage (motherage)

Variable: motherage (motherage)

Case	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	20	27.1000	9.2957	2.0786	0	43.0000
Yes	20	25.3000	5.5260	1.2357	16.0000	35.0000
Diff (1-2)		1.8000	7.6468	2.4181		

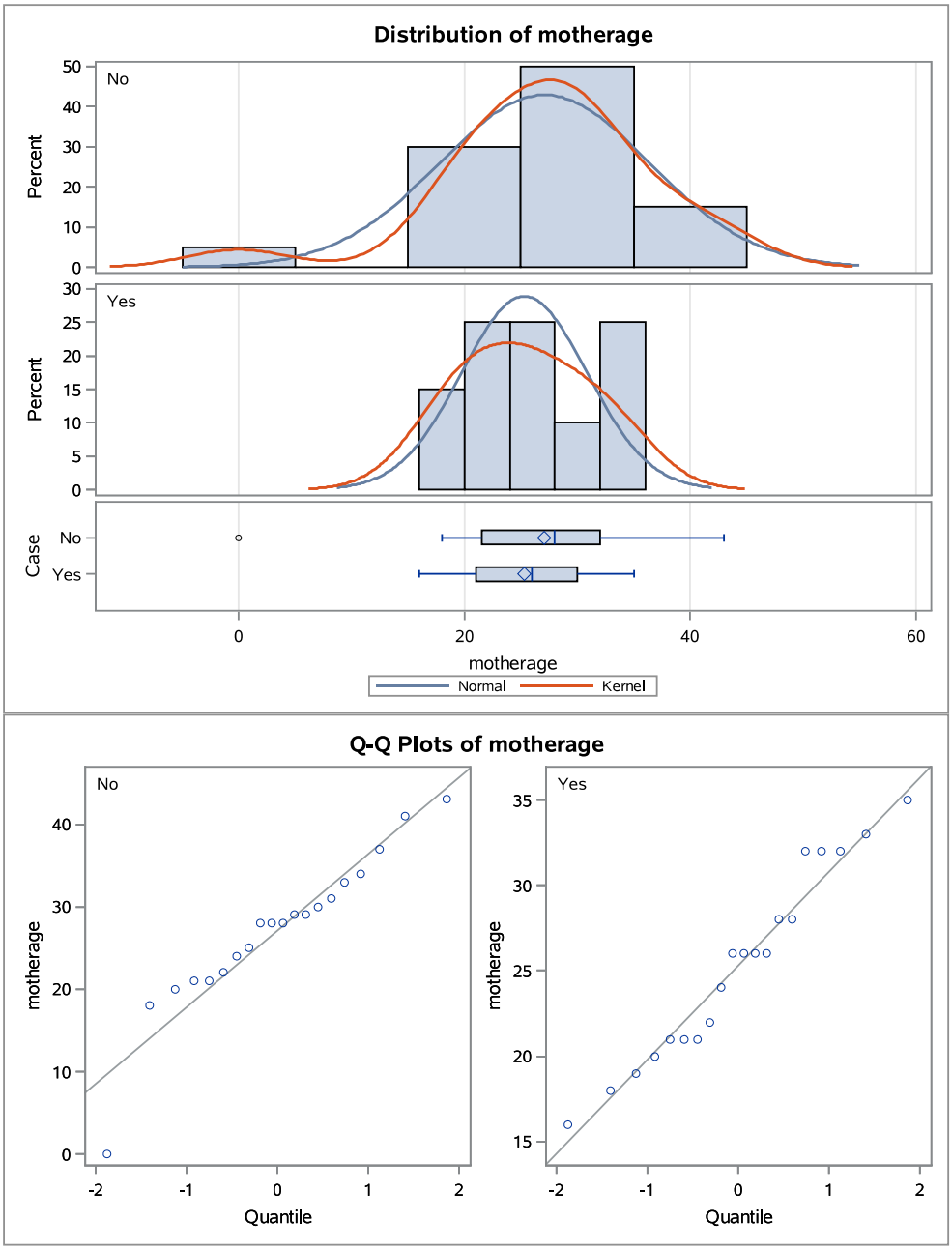
Case	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		27.1000	22.7495	31.4505	9.2957	7.0693	13.5771
Yes		25.3000	22.7137	27.8863	5.5260	4.2025	8.0711
Diff (1-2)	Pooled	1.8000	-3.0953	6.6953	7.6468	6.2493	9.8550
Diff (1-2)	Satterthwaite	1.8000	-3.1322	6.7322			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	38	0.74	0.4612
Satterthwaite	Unequal	30.938	0.74	0.4623

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	2.83	0.0285

The TTEST Procedure

Variable: motherage (motherage)



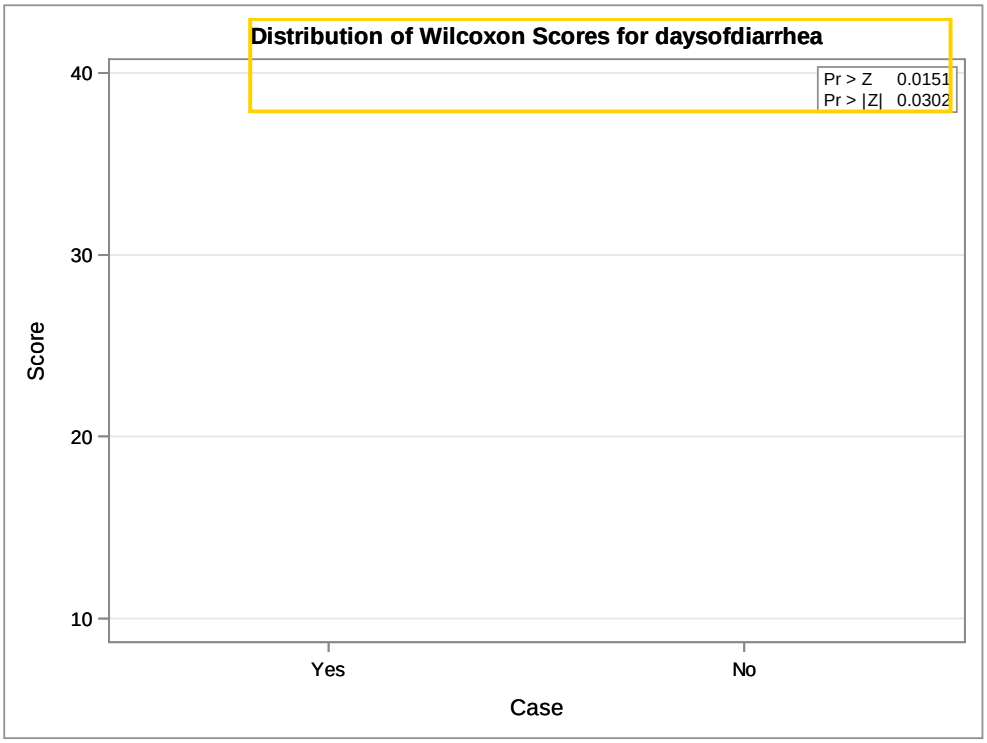
The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable daysofdiarrhea Classified by Variable Case					
Case	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Yes	20	486.50	410.0	35.064044	24.3250
No	20	333.50	410.0	35.064044	16.6750
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	486.5000
Normal Approximation	
Z	2.1675
One-Sided Pr > Z	0.0151
Two-Sided Pr > |Z|	0.0302
t Approximation	
One-Sided Pr > Z	0.0182
Two-Sided Pr > |Z|	0.0364
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	4.7599
DF	1
Pr > Chi-Square	0.0291

The NPAR1WAY Procedure



Appendix 2- 40. Epidemiological Data SAS Output

Raw SAS statistical data outputs for epidemiological data for *Campylobacter* infected in comparison to *Campylobacter* uninfected. (pages 1- 66)

The FREQ Procedure

underweight	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	37	92.50	37	92.50
2	2	5.00	39	97.50
3	1	2.50	40	100.00

wasting	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	38	97.44	38	97.44
2	1	2.56	39	100.00
Frequency Missing = 1				

Stunting	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	35	89.74	35	89.74
2	1	2.56	36	92.31
3	3	7.69	39	100.00
Frequency Missing = 1				

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of Case by underweight				
	underweight			
Case	1	2	3	Total
No	19 47.50 95.00 51.35	1 2.50 5.00 50.00	0 0.00 0.00 0.00	20 50.00
Yes	18 45.00 90.00 48.65	1 2.50 5.00 50.00	1 2.50 5.00 100.00	20 50.00
Total	37 92.50	2 5.00	1 2.50	40 100.00

Statistics for Table of Case by underweight

Statistic	DF	Value	Prob
Chi-Square	2	1.0270	0.5984
Likelihood Ratio Chi-Square	2	1.4133	0.4933
Mantel-Haenszel Chi-Square	1	0.6964	0.4040
Phi Coefficient		0.1602	
Contingency Coefficient		0.1582	
Cramer's V		0.1602	
WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 40

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by wasting		
	wasting		
Case	1	2	Total
No	18 46.15 94.74 47.37	1 2.56 5.26 100.00	19 48.72
Yes	20 51.28 100.00 52.63	0 0.00 0.00 0.00	20 51.28
Total	38 97.44	1 2.56	39 100.00
Frequency Missing = 1			

The FREQ Procedure

Statistics for Table of Case by wasting

Statistic	DF	Value	Prob
Chi-Square	1	1.0803	0.2986
Likelihood Ratio Chi-Square	1	1.4660	0.2260
Continuity Adj. Chi-Square	1	0.0007	0.9793
Mantel-Haenszel Chi-Square	1	1.0526	0.3049
Phi Coefficient		-0.1664	
Contingency Coefficient		0.1642	
Cramer's V		-0.1664	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Fisher's Exact Test	
Cell (1,1) Frequency (F)	18
Left-sided Pr <= F	0.4872
Right-sided Pr >= F	1.0000
Table Probability (P)	0.4872
Two-sided Pr <= P	0.4872

Effective Sample Size = 39
Frequency Missing = 1

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Case by Stunting			
	Stunting			
Case	1	2	3	Total
No	18 46.15 94.74 51.43	0 0.00 0.00 0.00	1 2.56 5.26 33.33	19 48.72
Yes	17 43.59 85.00 48.57	1 2.56 5.00 100.00	2 5.13 10.00 66.67	20 51.28
Total	35 89.74	1 2.56	3 7.69	39 100.00
Frequency Missing = 1				

Statistics for Table of Case by Stunting

Statistic	DF	Value	Prob
Chi-Square	2	1.3371	0.5124
Likelihood Ratio Chi-Square	2	1.7290	0.4213
Mantel-Haenszel Chi-Square	1	0.6605	0.4164
Phi Coefficient		0.1852	
Contingency Coefficient		0.1821	
Cramer's V		0.1852	
WARNING: 67% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Effective Sample Size = 39
Frequency Missing = 1

Appendix 2- 41. Z-Score Analysis of Subjects

Raw SAS statistical data outputs determining subject Z-scores (a value's relationship to the mean of a group of values, measured in terms of standard deviations from the mean). This specifically evaluated the subjects Z-scores in relation to underweight, wasting, and stunting of growth. (n=40, 20 per group)

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value	ANOVA (F-value)
Original	Plylum	Chao1	T-Test/ANOVA	0.70401	0.4717
Original	Class	Chao1	T-Test/ANOVA	0.99023	0.037205
Original	Order	Chao1	T-Test/ANOVA	0.92204	0.16069
Original	Family	Chao1	T-Test/ANOVA	0.92218	0.16046
Original	Genus	Chao1	T-Test/ANOVA	0.63386	0.577736
Original	Species	Chao1	T-Test/ANOVA	0.64322	0.56285
Original	OTU	Chao1	T-Test/ANOVA	0.45222	2.9752
Filtered	Plylum	Chao1	T-Test/ANOVA	0.22058	1.5454
Filtered	Class	Chao1	T-Test/ANOVA	0.3068	1.2505
Filtered	Order	Chao1	T-Test/ANOVA	0.30409	1.2584
Filtered	Family	Chao1	T-Test/ANOVA	0.66041	0.53656
Filtered	Genus	Chao1	T-Test/ANOVA	0.97433	0.072445
Filtered	Species	Chao1	T-Test/ANOVA	0.94355	0.12683
Filtered	OTU	Chao1	T-Test/ANOVA	0.91937	0.164776

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value	KW Value
Original	Plylum	Shannon	Mann-Whitney / Krustal-Wallis	0.22203	0.4393
Original	Class	Shannon	Mann-Whitney / Krustal-Wallis	0.47628	2.4945
Original	Order	Shannon	Mann-Whitney / Krustal-Wallis	0.32863	3.44
Original	Family	Shannon	Mann-Whitney / Krustal-Wallis	0.36704	3.1639
Original	Genus	Shannon	Mann-Whitney / Krustal-Wallis	0.3829	3.0578
Original	Species	Shannon	Mann-Whitney / Krustal-Wallis	0.36865	3.1529
Original	OTU	Shannon	Mann-Whitney / Krustal-Wallis	0.34447	3.3229
Filtered	Plylum	Shannon	Mann-Whitney / Krustal-Wallis	0.1758	4.946
Filtered	Class	Shannon	Mann-Whitney / Krustal-Wallis	0.46558	2.5542
Filtered	Order	Shannon	Mann-Whitney / Krustal-Wallis	0.40643	2.9055
Filtered	Family	Shannon	Mann-Whitney / Krustal-Wallis	0.39615	2.9708
Filtered	Genus	Shannon	Mann-Whitney / Krustal-Wallis	0.39521	2.9769
Filtered	Species	Shannon	Mann-Whitney / Krustal-Wallis	0.32112	3.4973
Filtered	OTU	Shannon	Mann-Whitney / Krustal-Wallis	0.32657	3.4556

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value	KW Value
Original	Plylum	Simpson	Mann-Whitney / Krustal-Wallis	0.27574	3.871
Original	Class	Simpson	Mann-Whitney / Krustal-Wallis	0.3452	3.3176
Original	Order	Simpson	Mann-Whitney / Krustal-Wallis	0.26787	3.9413
Original	Family	Simpson	Mann-Whitney / Krustal-Wallis	0.26787	3.9413
Original	Genus	Simpson	Mann-Whitney / Krustal-Wallis	0.29171	3.7335
Original	Species	Simpson	Mann-Whitney / Krustal-Wallis	0.30581	3.6177
Original	OTU	Simpson	Mann-Whitney / Krustal-Wallis	0.24582	4.149
Filtered	Plylum	Simpson	Mann-Whitney / Krustal-Wallis	0.18889	4.7767
Filtered	Class	Simpson	Mann-Whitney / Krustal-Wallis	0.44666	2.6624
Filtered	Order	Simpson	Mann-Whitney / Krustal-Wallis	0.33791	3.3708
Filtered	Family	Simpson	Mann-Whitney / Krustal-Wallis	0.41877	2.8388
Filtered	Genus	Simpson	Mann-Whitney / Krustal-Wallis	0.48067	2.4704
Filtered	Species	Simpson	Mann-Whitney / Krustal-Wallis	0.34021	3.3539
Filtered	OTU	Simpson	Mann-Whitney / Krustal-Wallis	0.22074	4.407

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value	KW Value
Original	Plylum	Choa1	Mann-Whitney / Krustal-Wallis	0.50643	2.3319
Original	Class	Choa1	Mann-Whitney / Krustal-Wallis	0.93606	0.42008
Original	Order	Choa1	Mann-Whitney / Krustal-Wallis	0.71238	1.3708
Original	Family	Choa1	Mann-Whitney / Krustal-Wallis	0.6353	1.7079
Original	Genus	Choa1	Mann-Whitney / Krustal-Wallis	0.659865	1.6033
Original	Species	Choa1	Mann-Whitney / Krustal-Wallis	0.57933	1.9668
Original	OTU	Choa1	Mann-Whitney / Krustal-Wallis	0.061167	7.3637
Filtered	Plylum	Choa1	Mann-Whitney / Krustal-Wallis	0.17268	4.9879
Filtered	Class	Choa1	Mann-Whitney / Krustal-Wallis	0.07206	6.9949
Filtered	Order	Choa1	Mann-Whitney / Krustal-Wallis	0.03403	8.6691
Filtered	Family	Choa1	Mann-Whitney / Krustal-Wallis	0.6487	1.6472
Filtered	Genus	Choa1	Mann-Whitney / Krustal-Wallis	0.87162	0.70674
Filtered	Species	Choa1	Mann-Whitney / Krustal-Wallis	0.80788	0.9726
Filtered	OTU	Choa1	Mann-Whitney / Krustal-Wallis	0.72175	1.3312

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value	ANOVA (F-value)
Original	Plylum	Shannon	T-Test/ANOVA	0.27549	1.347
Original	Class	Shannon	T-Test/ANOVA	0.22322	1.5348
Original	Order	Shannon	T-Test/ANOVA	0.13213	2.0025
Original	Family	Shannon	T-Test/ANOVA	0.12016	2.0875
Original	Genus	Shannon	T-Test/ANOVA	0.21272	1.5778
Original	Species	Shannon	T-Test/ANOVA	0.16601	1.7987
Original	OTU	Shannon	T-Test/ANOVA	0.13442	1.9871
Filtered	Plylum	Shannon	T-Test/ANOVA	0.20763	1.5994
Filtered	Class	Shannon	T-Test/ANOVA	0.25778	1.4064
Filtered	Order	Shannon	T-Test/ANOVA	0.15757	1.8452
Filtered	Family	Shannon	T-Test/ANOVA	0.13355	1.9929
Filtered	Genus	Shannon	T-Test/ANOVA	0.196998	1.6462
Filtered	Species	Shannon	T-Test/ANOVA	0.14517	1.9184
Filtered	OTU	Shannon	T-Test/ANOVA	0.1169	2.1122

Data Input	Taxonomic Level	Diversity Measue	Statistical Method	p-value
Filtered	OTU	Observed	T-Test/Anova	0.82837
Filtered	OTU	Observed	Mann-Whitney / Krustal-Wallis	0.89257
Filtered	OTU	Ace	T-Test/Anova	0.86653
Filtered	OTU	Ace	Mann-Whitney / Krustal-Wallis	0.79783
Filtered	OTU	Simpson	T-Test/Anova	0.020358
Filtered	OTU	Simpson	Mann-Whitney / Krustal-Wallis	0.21192
Filtered	OTU	Fisher	T-Test/Anova	0.88718
Filtered	OTU	Fisher	Mann-Whitney / Krustal-Wallis	0.892527
Original	OTU	Observed	T-Test/Anova	0.1125
Original	OTU	Observed	Mann-Whitney / Krustal-Wallis	0.23843
Original	OTU	Ace	T-Test/Anova	0.017768
Original	OTU	Ace	Mann-Whitney / Krustal-Wallis	0.033387
Original	OTU	Simpson	T-Test/Anova	0.019404
Original	OTU	Simpson	Mann-Whitney / Krustal-Wallis	0.24582
Original	OTU	Fisher	T-Test/Anova	0.18889
Original	OTU	Fisher	Mann-Whitney / Krustal-Wallis	0.2975

Appendix 2- 42. Alpha-Diversity Statistics Table for the Four Cohorts

The low count filter setting was equal to 4 and prevalence in sample was set to 20%, while the low variance filter was set to remove 10% based on the inter-quantile range. The data was normalized by rarefying but did not scale or transform data. All statistically significant p-vales are highlighted in yellow.

Taxonomic Level	Distance Method	Statistical Method	p-value
Phylum	Bray-Curtis	PERMANOVA	0.2
Phylum	Bray-Curtis	ANOSIM	0.269
Phylum	Bray-Curtis	PERMDISP	0.71599
Class	Bray-Curtis	PERMANOVA	0.2
Class	Bray-Curtis	ANOSIM	0.269
Class	Bray-Curtis	PERMDISP	0.71599
Order	Bray-Curtis	PERMANOVA	0.2
Order	Bray-Curtis	ANOSIM	0.269
Order	Bray-Curtis	PERMDISP	0.71599
Family	Bray-Curtis	PERMANOVA	0.2
Family	Bray-Curtis	ANOSIM	0.269
Family	Bray-Curtis	PERMDISP	0.71599
Genus	Bray-Curtis	PERMANOVA	0.2
Genus	Bray-Curtis	ANOSIM	0.269
Genus	Bray-Curtis	PERMDISP	0.71599
Species	Bray-Curtis	PERMANOVA	0.2
Species	Bray-Curtis	ANOSIM	0.269
Species	Bray-Curtis	PERMDISP	0.71599
OTU	Bray-Curtis	PERMANOVA	0.2
OTU	Bray-Curtis	ANOSIM	0.269
OTU	Bray-Curtis	PERMDISP	0.71599

Taxonomic Level	Distance Method	Statistical Method	p-value
Phylum	Jensen-Shannon Divergence	PERMANOVA	0.227
Phylum	Jensen-Shannon Divergence	ANOSIM	0.249
Phylum	Jensen-Shannon Divergence	PERMDISP	0.68747
Class	Jensen-Shannon Divergence	PERMANOVA	0.227
Class	Jensen-Shannon Divergence	ANOSIM	0.249
Class	Jensen-Shannon Divergence	PERMDISP	0.68747
Order	Jensen-Shannon Divergence	PERMANOVA	0.227
Order	Jensen-Shannon Divergence	ANOSIM	0.249
Order	Jensen-Shannon Divergence	PERMDISP	0.68747
Family	Jensen-Shannon Divergence	PERMANOVA	0.227
Family	Jensen-Shannon Divergence	ANOSIM	0.249
Family	Jensen-Shannon Divergence	PERMDISP	0.68747
Genus	Jensen-Shannon Divergence	PERMANOVA	0.227
Genus	Jensen-Shannon Divergence	ANOSIM	0.249
Genus	Jensen-Shannon Divergence	PERMDISP	0.68747
Species	Jensen-Shannon Divergence	PERMANOVA	0.227
Species	Jensen-Shannon Divergence	ANOSIM	0.249
Species	Jensen-Shannon Divergence	PERMDISP	0.68747
OTU	Jensen-Shannon Divergence	PERMANOVA	0.227
OTU	Jensen-Shannon Divergence	ANOSIM	0.249
OTU	Jensen-Shannon Divergence	PERMDISP	0.68747

Taxonomic Level	Distance Method	Statistical Method	p-value
Phylum	Jaccard Index	PERMANOVA	0.164
Phylum	Jaccard Index	ANOSIM	0.269
Phylum	Jaccard Index	PERMDISP	0.56335
Class	Jaccard Index	PERMANOVA	0.164
Class	Jaccard Index	ANOSIM	0.269
Class	Jaccard Index	PERMDISP	0.56335
Order	Jaccard Index	PERMANOVA	0.164
Order	Jaccard Index	ANOSIM	0.269
Order	Jaccard Index	PERMDISP	0.56335
Family	Jaccard Index	PERMANOVA	0.164
Family	Jaccard Index	ANOSIM	0.269
Family	Jaccard Index	PERMDISP	0.56335
Genus	Jaccard Index	PERMANOVA	0.164
Genus	Jaccard Index	ANOSIM	0.269
Genus	Jaccard Index	PERMDISP	0.56335
Species	Jaccard Index	PERMANOVA	0.164
Species	Jaccard Index	ANOSIM	0.269
Species	Jaccard Index	PERMDISP	0.56335
OTU	Jaccard Index	PERMANOVA	0.164
OTU	Jaccard Index	ANOSIM	0.269
OTU	Jaccard Index	PERMDISP	0.56335

Appendix 2- 43. Beta-Diversity of the Four Cohorts

The low count filter setting was equal to 4 and prevalence in sample was set to 20%, while the low variance filter was set to remove 10% based on the inter-quantile range. The data was normalized by rarefying but did not scale or transform data. All statistically significant p-vales are highlighted in yellow.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.81555
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.8301
OTU	Shannon	T-Test/ANOVA	0.26829
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.111

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.219

Appendix 2- 44. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 2 and prevalence in sample was set to 20%, while the low variance filter was set to remove 5% based on the inter-quantile range. The data was normalized by rarefying but did not scale or transform data. After Filtration there was 404 features remaining.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.81555
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.8301
OTU	Shannon	T-Test/ANOVA	0.26829
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.111

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.219

Appendix 2- 45. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 2 and prevalence in sample was set to 20%, while the low variance filter was set to remove 5% based on the inter-quantile range. The data was normalized by rarefying with total sum scaling.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.93641
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.9874
OTU	Shannon	T-Test/ANOVA	0.26695
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.10842

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.223

Appendix 2- 46. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 2 and prevalence in sample was set to 20%, while the low variance filter was set to remove 5% based on the inter-quantile range. The data was not normalized in anyway.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.81555
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.83301
OTU	Shannon	T-Test/ANOVA	0.26829
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.111

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.219

Appendix 2- 47. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 2 and prevalence in sample was set to 20%, while the low variance filter was set to remove 5% based on the inter-quantile range. The data was normalized by only total sum scaling.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.93641
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.9874
OTU	Shannon	T-Test/ANOVA	0.26695
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.10842

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.223

Appendix 2- 48. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 2 and prevalence in sample was set to 20%, while the low variance filter was set to remove 5% based on the inter-quantile range. The data was normalized by rarefying with relative log expression.

Taxonomic Level	Diversity Measue	Statistical Method	p-value
OTU	Choa1	T-Test/ANOVA	0.82533
OTU	Choa1	Man-Whitney/Krustal-Wallis	0.73211
OTU	Shannon	T-Test/ANOVA	0.34352
OTU	Shannon	Man-Whitney/Krustal-Wallis	0.12077

Taxonomic Level	Distance Method	Statistical Method	p-value
OTU	Bray-Curtis	PERMANOVA	0.201

Appendix 2- 49. Alternative Data Filtering and Normalization Statistics

The low count filter setting was equal to 6 and prevalence in sample was set to 25%, while the low variance filter was set to remove 15% based on the inter-quantile range.

VITA

Jessica Logan Tweedie ("Jessie") was born on April 28th, 1994 in Decatur, Alabama to Paula Brown Couch and Alan David Tweedie. She grew up in Somerville, Alabama on a farm. She attended Auburn University from 2012-2016 studying microbiology and public health. In 2017, she began her graduate studies at The University of Tennessee in Knoxville in Dr. Jeremiah Johnson's laboratory studying *Campylobacter jejuni.*