

**Prediction of Some Health Outcomes Among Advanced Cancer Patient Caregivers Using  
Gut Microbiome**

by

**Ruopu Song**

BS in Biology, Nankai University, China, 2018

Submitted to the Graduate Faculty of the

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH  
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

**Ruopu Song**

It was defended on

March 20, 2020

and approved by

Jeanine Buchanich, PhD, Research Associate Professor, Biostatistics  
Graduate School of Public Health, University of Pittsburgh

Jennifer Steel, PhD, Professor, Surgery, Psychiatry and Psychology  
Kenneth P. Dietrich School of Arts & Sciences, University of Pittsburgh

**Thesis Advisor:** Shyamal Peddada, PhD, Professor and Chair, Biostatistics  
Graduate School of Public Health, University of Pittsburgh

Copyright © by Ruopu Song

2020

# **Prediction of Some Health Outcomes Among Advanced Cancer Patient Caregivers Using Gut Microbiome**

Ruopu Song, MS

University of Pittsburgh, 2020

## **Abstract**

Increasingly researchers are discovering that many health outcomes are associated with gut microbiome. This was the motivation for the research conducted in this thesis with focus on psychosocial and metabolic health of caregivers of cancer patients. The microbiome data used in this study were obtained from cancer caregivers who participated in a study focusing on the relationship between psychosocial and behavioral predictors, and metabolic syndrome. Our goal was to determine how well microbiome alone can predict the psychosocial and metabolic health of a caregiver, as defined by depression, stress, hostility and patient-caregiver relationship, and metabolic syndrome.

We explored two different prediction (or classification) procedures, namely Fisher's Linear Discriminant Analysis (LDA) and logistic regression. The predictors consisted of the five important bacterial phyla that constitute 95% of the adult gut flora, namely, Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Verrucomicrobia, with the remaining phyla combined as "Other". Aitchison's log-ratios were used to transform the data from the simplex to the Euclidean space before applying LDA and logistic regression.

For the five health characteristics, we found that the logistic regression had generally higher total correct prediction/classification rates for classifying subjects into their correct health categories than LDA. We found the overall correct classification rates for depression and patient-caregiver relationship to be 80% and 67%, respectively. Thus, there is an 80% chance of correctly

predicting the depression symptom category (low or high) of a caregiver using the microbial phyla. The correct classification rates for caregiver stress, hostility and metabolic syndrome were about 63%, 60% and 53%, respectively.

This appears to be the first study that attempted to predict psychosocial characteristics and metabolic health of a caregiver using the stool microbial phyla in cancer patient caregiver population. Although the overall success is modest, the results are encouraging to conduct a larger follow-up study which can have major clinical implications. If successful, like blood and urine tests, the stool microbiome can potentially be used to diagnose the above noted psychosocial characteristics and metabolic syndrome for caregivers. The public health relevance of this study is that we provided a direction on developing new microbiome-based intervention for helping cancer patient caregivers relieve mental and physical health symptoms.

## Table of Contents

Preface.....	ix
1.0 Introduction.....	1
2.0 Methodology .....	4
2.1 Microbiome Data .....	4
2.1.1 Dimension Reduction .....	5
2.1.2 Estimation of Mean and Covariance Matrix .....	7
2.1.3 Dataset.....	9
2.2 Instruments .....	10
2.2.1 Psychological and Social Variables .....	10
2.2.2 Metabolic Syndrome.....	11
2.3 LDA Classification Model.....	12
2.4 Logistic Regression Model.....	13
2.3.1 Diagnostics for Logistic Regression Models .....	15
3.0 Results .....	18
3.1 Classification .....	19
4.0 Discussion.....	25
Appendix Supplemental Table.....	29
Bibliography .....	30

## List of Tables

<b>Table 1 Medians for each of the five variables</b> .....	<b>11</b>
<b>Table 2 Demographic Variables Summary</b> .....	<b>19</b>
<b>Table 3 Percentages of Correct Classification</b> .....	<b>20</b>
<b>Table S 1 Odds Ratios of Logistic Model</b> .....	<b>29</b>

## List of Figures

<b>Figure 1 Boxplot of Ages in Each Gender of Caregivers .....</b>	<b>18</b>
<b>Figure 2 ROC Curve for Stress Model (AUC = 0.67).....</b>	<b>22</b>
<b>Figure 3 ROC Curve for Depression Model (AUC = 0.80) .....</b>	<b>22</b>
<b>Figure 4 ROC Curve for Hostility Model (AUC = 0.67) .....</b>	<b>23</b>
<b>Figure 5 ROC Curve for DAS Model (AUC = 0.74) .....</b>	<b>23</b>
<b>Figure 6 ROC curve for metabolic syndrome model (AUC = 0.6) .....</b>	<b>24</b>

## Preface

I would like to thank my advisor Dr. Peddada for his excellent guidance and support during the process of my thesis. I am very indebted to Dr. Buchanich and Dr. Steel, members of my thesis committee, who were extremely patient with my writing and made numerous edits that elevated the quality of my presentation. I am very grateful to Dr. Steel who provided me the cancer patient caregiver data that were analyzed. I am also grateful to her for introducing to this very important and interesting area of research. I wish to thank all three members of my committee, without whose corporation and help, I would not have been able to conduct the research described in this thesis.

I would like to thank my girlfriend Anni Guo and my best friend Bo Liu, without their help and support, I would not have been able to be confident and strong enough to solve the problems I met during the process of my work. Last but not least, I would like to thank my families for their support during the two years of my master study in a foreign country.

Ruopu Song

Mar. 20. 2020

## 1.0 Introduction

With the advent of high throughput sequencing technologies, the past decade has witnessed an exponential growth in human microbiome studies for understanding the role of microbiome in human health (Hadrich, 2018). A variety of diseases and health conditions have been demonstrated to be associated, and even caused, by human. To put things in proper perspective, between 2013 and 2017 about 13,000 papers related to gut microbiome were published and in 2017 alone there were about 4000 publications (Cani, 2018). The 13,000 publications during the period 2013 to 2017 account for nearly 80% of total publications on the subject since 1977.

This exponential growth in literature is not surprising because humans are largely microbial organisms (Cani, 2018) and therefore human mental and physical health may partly be affected by the imbalances and changes in microbial composition. Increasingly researchers are discovering functions and mode of action of different bacteria in various diseases. Some popular examples are gram-negative bacteria such as some members of the phylum Proteobacteria, including *Escherichia coli*, *Salmonella*, *Shigella*, *Enterobacteriaceae*, *Pseudomonas*, *Moraxella*, *Helicobacter*, and so on, that are associated with or cause various diseases (Shin, Whon, & Bae, 2015). The outer membrane of gram-negative bacteria has lipopolysaccharides (LPS) which are also known as lipoglycans and endotoxins. They are involved in the production of pro-inflammatory cytokines such as interleukin-1B (IL-1B), IL-2, IL-6, interferon gamma (IFN- $\gamma$ ), and tumor necrosis factor (TNF) (Kany, Vollrath, & Relja, 2019). The LPS pathway is not limited to pro-inflammatory gram-negative bacteria alone, but there are other bacteria which are also involved in triggering the LPS. For example, it is well-known that increase in phyla Firmicutes (and perhaps Actinobacteria) and a decrease in Bacteroidetes may potentially increase the

absorption of lipopolysaccharide (LPS) causing the activation of Toll-like receptors 2 and 4, among others, thus leading to inflammatory pathways and hence insulin resistance and metabolic syndrome (Caricilli & Saad, 2013).

Many of these are known to be involved in various risk factors of metabolic syndrome such as insulin resistance and obesity (De Luca & Olefsky, 2008). Metabolic syndrome is a combination of several well-known cardiovascular risk factors including insulin resistance, diabetes, stroke, measured by blood pressure and abdominal girth (Huang, 2009). Psychosocial factors including stress, depression and hostility are associated with the components of metabolic syndrome, such as increased insulin resistance, hypertension and obesity (Vaccarino et al., 2008). And cancer patient caregivers were found under the risk of coronary heart disease and stroke, which are the components of metabolic syndrome (Ji, Zöller, Sundquist, & Sundquist, 2012). Therefore, cancer caregivers are of interests in this study because they are under the risks of some mental health problems which were linked to metabolic syndrome.

Several mental health problems such as depression, stress, and anxiety are linked to dysbiosis in the gut bacteria (Reber et al., 2016). Depression, for example, is correlated with increased levels of IL-6 and TNF- $\alpha$  (Berk et al., 2013). The gut microbiome are known to affect the production of these cytokines. An increase in the relative abundances of Actinobacteria and Bacteroidetes, and a decrease in the relative abundance of Firmicutes was observed in patients with depression symptoms compared to the healthy controls (Jiang et al., 2015). In a CD-1 mouse study (Bailey et al., 2011), the authors noticed a decrease in the abundance of Bacteroidetes to be associated with stress.

Thus, there is considerable evidence to suggest that the microbiome is associated with metabolic syndrome as well as with psychosocial factors. However, these associations have not

been well studied in the caregiver population of cancer patients. By understanding these associations, and if possible causal pathways, one may potentially derive new microbiome-based treatments for caregivers, such as probiotics or even fecal matter transplantation (FMT) which is becoming increasingly popular method of treatment for many diseases. Such approaches may be necessary considering the fact that the average age of caregivers is over 60 and hence may not be adaptable to new diets, physical and social activities which are often necessary for improving health.

Therefore, in this thesis we take the first step towards understanding how well the gut microbiome predicts various physical and psychosocial characteristics. Specifically, we are interested in predicting characteristics such as depression, stress, hostility and patient-caregiver relationship, and metabolic syndrome using gut microbiome. Similar to how physicians predict the various health characteristics of patients using blood and urine samples, we are interested in knowing whether gut microbiome can be used to predict a caregiver's health in terms of the above characteristics.

This thesis is organized as follows. In Chapter 2 we describe the microbiome data obtained from a pilot study, followed by the statistical methodology used in this thesis to predict the psychosocial and metabolic health of a caregiver. In particular, we describe Fisher's linear discriminant analysis (LDA) and logistic regression-based methodology. Results of the analyses of the pilot data are summarized in Chapter 3. Concluding remarks along with strengths and limitations of the study are provided in Chapter 4.

## 2.0 Methodology

In this section, we begin by introducing some notations and background information for classifying subjects into different categories of a phenotype. For simplicity of exposition, we shall use the generic term “phenotype” to mean either depression, stress, hostility, patient-caregiver relationship, or metabolic syndrome.

We consider two classification or class prediction procedures, namely, the Linear Discriminant Analysis (LDA) and the logistic regression analysis.

### 2.1 Microbiome Data

The typical microbiome data are obtained using two different technologies, one based on 16s ribosomal RNA (16s rRNA) and the other based on metagenomics. Although the statistical methodology described in this thesis is broadly applicable to data obtained from either technology, in this thesis we focus on data obtained from 16s rRNA. Typically, the 16s rRNA technology yields operational taxonomic unit (OTU) categories, which may be viewed as surrogates for various bacteria available in various databases. These OTUs may be summarized at different levels of phylogeny, such as phylum, genus, family, etc. The dominant phyla in the human gut are Actinobacteria, Bacteroidetes, Proteobacteria, and Firmicutes (Kho & Lal, 2018). For simplicity of exposition, throughout this thesis we use the term “taxa” to represent bacteria at any level of the phylogeny. The singular of taxa is “taxon”.

The OTU count data is a matrix where the rows (denoted by  $p$ ) are the various OTUs and the columns are the samples or subjects (denoted by  $n$ ) belonging to different study groups. Often  $n < p$  and hence the curse of dimensionality. The OTU data matrix is often called the OTU table. The OTU table has three important characteristics. Firstly, it is a sparse table – meaning a large number of entries are zeros (or unobserved). As described in Kaul et al. (2017), zeros arise in the OTU tables for a number of reasons (Kaul, Mandal, Davidov, & Peddada, 2017). Kaul et al. identified three major sources of zeros, namely, structural zeros, outlier zeros and zeros due to sampling depth or library size. A second important characteristic of the OTU table is that the observed data are necessarily compositional, i.e. they are relative abundances of taxa that sum to a constant. Consequently, standard Euclidean space-based methods are not appropriate for these data as they reside inside a  $p$ -dimensional simplex. One needs to transform the simplex space-based data into appropriate  $(p-1)$ -dimensional Euclidean space data and then apply Euclidean space methods (Kaul, Davidov, & Peddada, 2017; Kaul, Mandal, et al., 2017; Mandal et al., 2015). The third important characteristic of the OTU table is that the library size across samples (i.e. total number of bacterial counts) is not constant. This variation can be troubling when the library sizes are highly variable. As will be described in the next subsection, the Aitchison’s log-ratio based methodology used in this thesis overcomes several of the above issues.

### **2.1.1 Dimension Reduction**

Since the number of taxa is usually large, often larger than the sample size, many classification methods including LDA and logistic regression model cannot be applied directly without performing some notion of dimension reduction. While one may use the computational methodology developed in Kaul et al. (2017), we take a more biological approach to dimension

reduction. Rather than working with OTUs which are often very sparse, i.e., about 90% of OTU tables consist zeros and the sample size is often substantially smaller than the number of OTUs (dimension of the problem), we perform classification of samples at a higher level of the taxonomy, namely the phyla level. As often done in microbiome literature, we removed phyla that were observed in only one subject. Among the remaining phyla, we combined phyla that were observed in less than 30% of the subjects (i.e. 9 subjects out of 30) to further reduce the dimension of phyla table. This was necessary because we have a very small sample size and if we did not establish such filters then we would be dealing with very small and sparse data to draw any useful conclusions. After completing these pre-processing steps, we summarized the taxa into five phyla, which constitute more than 95% of the taxa in an adult gut, namely, Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Verrucomicrobia, with the remaining taxa were combined into a single group called Others. By doing so, we also addressed the sparsity issue related to OTU tables.

As noted earlier, the relative abundances of microbiome data sum to 1. Thus, the data reside inside a simplex. To be precise, suppose for the  $i^{th}$  subject,  $i = 1, 2, \dots, n$ ,  $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{id+1})^T$  denotes the abundance of  $(d+1)$  taxa. Since  $Y_i$  is a point inside a simplex, i.e.  $\sum_{j=1}^{d+1} Y_{ij} = C$ , we therefore use Aitchison (1982) transformation of the data to a  $d$  – dimensional Euclidean space by converting data into log-ratios using one of the taxa as the reference taxon. Without loss of generality, suppose  $(d + 1)^{th}$  taxon is the reference taxon, then we transform  $Y_i$  to  $X_i$ , where  $X_{ij} = \ln\left(\frac{Y_{ij}}{Y_{id+1}}\right)$ . Thus, all our calculations are based on these log-transformed data. It is important to note that some of the abundances  $Y_{ij}$  may be zero. Because Bacteroidetes and Firmicutes are two common phyla in every subject, we treated the Bacteroidetes as the reference taxon to perform log-transformation. In such cases, following Kaul et al. (2017), we shall classify the data into one of the three types of zeros using the methodology described in Kaul et al. (2017).

If it is either a structural zero or an outlier zero, we shall treat that observation as a missing value in the rest of the calculations. If it is a sampling zero then it will be replaced by 1 before computing logarithms. While imputing the missing counts by 1 is arbitrary, and may even lead to slight inflation of false positive and false negative rates (Kaul et al., 2017), standard imputation methods are not valid for microbiome data because of differential library sizes across samples (Kaul et al., 2017) and this is the common approach used in the microbiome literature (Mandal et al., 2015). Therefore, the transformed microbiome data matrix used to do analysis in following sections has 5 dimensions.

### 2.1.2 Estimation of Mean and Covariance Matrix

As we mentioned in the previous section, the structural zeros and outlier zeros were treated as missing value in the transformed data matrix, thus the distribution of the missing patterns was needed to be defined. Kaul et al. (2017) derived the distribution of missing pattern  $M_i$ , and supposed that the missingness has independent but not identical probability to satisfy the definition of structural zeros and outlier zeros. The observed data  $X_{ij}$  after log-transformation is needed to be defined by an index set  $A_i = \{j, M_{ij} = 1\}$  (Kaul, Davidov, et al., 2017). Therefore, the estimation of sub-vector mean  $\mu$  and covariance matrix  $\Sigma$  can be calculated by the components of  $X_{ij}$  in the index set  $A_i$ , which are non-zero components of the observation. Kaul claimed that the zero components of an observation do not affect the distribution of non-zero components in the same observation. Under this definition, we can derive the estimation of mean vector of non-zero components, for each  $1 \leq l, m \leq d$ , where  $n_{(l)}$  is defined as the number of subjects where  $l$ th component is observed.

$$\widehat{\mu}_l = \frac{1}{n_{(l)}} \sum_{i \in n_{(l)}} X_{il}, \quad 1 \leq l \leq d$$

The estimation of covariance matrix  $\widehat{\Sigma}$  is defined as follow, where  $n_{(l,m)}$  is defined as the number of subjects where both  $l$ th and  $m$ th components are observed. The precision matrix  $\widehat{\Omega}$  is the inverse of the estimator of the covariance matrix  $\widehat{\Sigma}$ .

$$\widehat{\sigma}_{lm} = \frac{\sum_{i \in n_{(l,m)}} (X_{il} - \widehat{\mu}_l)(X_{im} - \widehat{\mu}_m)}{|n_{(l,m)}|}, \quad \widehat{\Sigma} = [\widehat{\sigma}_{lm}]_{l,m=1,\dots,d}$$

However, the estimation of the precision matrix is not guaranteed to be positive definite due to the estimation of covariance matrix is not guaranteed to be positive definite. A mild assumption on the missing structure was made to proceed the estimation of covariance matrix and precision matrix with any  $0 \leq q < 1$ . We assume the covariance matrix and precision matrix belongs to the following classes of matrices respectively,

$$M(q, s_0(d), K) = \left\{ \Sigma: \sigma_{ii} \leq K, \max_{1 \leq i \leq d} \sum_{j=1}^d |\sigma_{ij}|^q \leq s_0(d) \right\}$$

$$U(q, s_0(d), K) = \left\{ \Omega: \Omega > 0, \|\Omega\|_{L_1} \leq K, \max_{1 \leq i \leq d} \sum_{j=1}^d |\omega_{ij}|^q \leq s_0(d) \right\}$$

where  $s_0(d)$  is depend on  $d$ ,  $\|\Omega\|_{L_1} = \sum_{ij} |\omega_{ij}|$  is the elementwise  $l_1$  norm. The estimator  $\widehat{\Omega}$  converges to positive definite limit with asymptotic probability 1 by the theorem developed in Kaul study (Kaul, Davidov, et al., 2017). The theorem indicates if  $X_i$ ,  $1 \leq i \leq n$  follow the Gaussian distribution and  $\Omega \in U$ , then the *Sup* norm of  $\widehat{\Omega} - \Omega$  is bounded with the probability at least  $1 - c_1 \exp(-c_2 \log d)$ ,

$$\|\widehat{\Omega} - \Omega\|_{\infty} = o\left(\sqrt{\frac{\log d}{n}}\right)$$

### 2.1.3 Dataset

We use gut microbiome data derived from stool samples obtained from thirty caregivers who participated in a pilot study conducted by Dr. Steel. These thirty caregivers were a sample from a larger trial (Steel et al., 2019). The caregivers with their spouses who had cancer were involved in a prospective study and evaluated on mental and physical health by using questionnaires, physical exams and blood tests at the entry of the study and at 6- and 12-month follow-up (Steel et al., 2019). The microbiome data and instruments were collected at baseline. The dataset for analysis and classification is summarized as a combination of the phyla table, and the mental and physical health outcomes. The phyla table has 30 observations and 5 variables including the counts of each of four phyla, namely, Actinobacteria, Firmicutes, Proteobacteria, Verrucomicrobia, and Others. The mental and physical health outcomes are caregiver stress, depression, hostility and patient-caregiver relationship, and metabolic syndrome. Each of these mental and physical health outcomes will be introduced in the following section. The phyla table is used to create classification models according to each of five health outcomes. As noted earlier, log-ratios were calculated with respect to Bacteroidetes. Hence, throughout this chapter from this point on we shall not explicitly mention Bacteroidetes. All remaining phyla are understood to be relative to Bacteroidetes.

## 2.2 Instruments

### 2.2.1 Psychological and Social Variables

Caregiver stress was evaluated by the reversed score obtained from the Caregiver Quality of Life Index-Cancer (CQOLC) scale with 35 items of cancer-specific instruments. It assessed the life quality of cancer caregivers (Weitzner, Jacobsen, Wagner, Friedland, & Cox, 1999). The scores were reversed so that a high score represents high caregiver stress or poor quality of life (Steel et al., 2019). The subjects were classified into low stress level group if stress score was less than the median of its distribution, which was 23, otherwise classified into high stress level group (Table 1).

Depressive symptoms were measured by the caregivers reported score obtained from the Center for Epidemiologic Studies-Depression (CES-D), which is a 20-item self-report questionnaire of depression symptoms (Radloff, 1977). The caregivers were evaluated by the self-report score on a 4-point scale by every week of depressive symptoms (“rarely”, “some days”, “occasionally”, or “most days”) (Steel et al., 2019). The subjects were classified into the different group of two depression categories according to the CES-D score. When depression score obtained from CES-D was less than 16, the subject was classified into the group representing the low depression symptoms levels, otherwise classified into the group representing the high depression symptoms levels.

Hostility was defined by mistrust, cynicism, and negative beliefs and attributions concerning others (Smith, 2003). The same measurement, the Cook-Medley Hostility Scale with 50 items. The 50 self-report questions were true-false items from the Minnesota Multiphasic Personality Inventory (Scherwitz, Perkins, Chesney, & Hughes, 1991). The subjects were

classified into a binary outcome with low and high level of hostility by the median of the hostility scores (Table 1).

Relationship between the caregiver and the patient was evaluated by the score obtained from the short form of the Dyadic Adjustment Scale (DAS-7) with 7 items (Hunsley, Best, Lefebvre, & Vito, 2001; Spanier, 1976). DAS is used to evaluate the satisfactory of the relationship between caregivers and their spouses with cancer (Spanier, 1976). The subject was classified into the DAS group with two categories representing the low and high level of the relationship by the median of the DAS scores, respectively (Table 1).

### 2.2.2 Metabolic Syndrome

As we mentioned in the previous section, metabolic syndrome was a group of risk factors for cardiovascular disease (CVD). The metabolic syndrome score was evaluated by five metabolic abnormalities: (a) abdominal girth  $\geq 40$  inches in men and  $\geq 35$  inches in women; (b) elevated serum triglycerides  $\geq 150$  mg/dL; (c) HDL  $< 40$  mg/dL in men; (d) blood pressure  $\geq 130/85$  mmHg or drug treatment for high blood pressure; and (e) nonfasting glucose  $\geq 140$  mg/dL (Steel et al., 2019). The metabolic syndrome score is the number of how many metabolic abnormalities were reported for caregivers, which was obtained from the pilot study (Steel et al., 2019). We classified the subjects with two or less risk factors as the no metabolic syndrome group, while classified the subjects with three or more risk factors as the metabolic syndrome group.

**Table 1 Medians for each of the five variables**

Caregiver Stress	Depression	Hostility	Patient-caregiver Relationship	Metabolic Syndrome
23	11	12.5	25	3

The mental and physical health variables were defined as binary outcomes. The cutoff point in each phenotype is the median of that variable, except for the depression and metabolic syndrome. The cutoff point of depression is 16 from the score of the questionnaire due to the extensive research that has demonstrated that a score of 16 has a high sensitivity and specificity for detecting a DSM diagnosis of Major Depressive Disorder assessed using a structured clinical interview. A subject with metabolic score less than or equal to 2 was classified as “no metabolic syndrome”, and a subject with score more than two was classified as “have metabolic syndrome”. Mental and physical health table has 30 observations and 5 binary outcomes. The classification model which will be introduced in the following sections has been performed for the phyla table according to each of five binary outcomes.

### **2.3 LDA Classification Model**

LDA is the method to find linear combinations of predictors that define or separate two or more classes of subjects (Fisher, 1936). To simplify and reduce the dimension of the data matrix, our analysis is based on “phylum” level, thus the dimension ( $d$ ) is less than the sample size ( $n = 30$ ). In this case, the precision matrix  $\Omega$  can be estimated and performed in the LDA based classification according to each phenotype.

Since we have a small sample size, we estimate the correct classification rate of LDA method using the Leave-One-Out (LOO) method instead of creating test and training sets (Johnson & Wichern, 2002). The LOO method is implemented as follows. From the 30 observations, we hold back one observation and use the remaining 29 observations to create the linear discriminant rule. Using this rule, we classify the observation that was left out. We repeat this process with all

30 observations and obtain the proportion of correct classification rates. Three different types of classification rates were calculated; (a) correct classification rate for subjects in group 1 (e.g. low stress group), (b) correct classification rate for subjects in group 2 (e.g. high stress group), (c) overall correct classification rate for both groups combined.

Classification of subjects into two groups using LDA method is performed as follows. Using the methodology described in Chapter, we obtain the sample mean vectors  $\hat{\mu}_1$  and  $\hat{\mu}_2$  for group 1 and group 2, respectively, and the precision matrix  $\hat{\Omega}$ , under the assumption that two groups have the same population covariance matrix  $\Sigma$ . Let  $X = (X_1, \dots, X_d)^T$  denote a vector corresponding to a new subject to be classified into group 1 or group 2. Let

$$\delta_r(X) = X^T \Omega \hat{\mu}_r - \frac{1}{2} \hat{\mu}_r^T \Omega \hat{\mu}_r, r = 1, 2$$

Then observation  $X$  is classified into group 1 if  $\delta_1(X) > \delta_2(X)$ , otherwise classified in group 2. The correct classification rate for group  $r$  is calculated by the number of subjects classified into the group  $r$  where they are observed in. For example, the correct classification rate for low depression symptoms level group is 66.7%, because out of 15 caregivers who belong to the low depression symptoms group, 10 were correctly classified into that group.

## 2.4 Logistic Regression Model

In addition to the LDA method, logistic regression was also performed. The generalized logistic regression model defined the linear predictors  $\eta_i$  calculated by the logistic regression equation. We defined each of the five mental and physical health outcomes as a binary response  $Y_i$ , where  $i = 1, \dots, n$ , and the probability of  $Y_i = 1$  is modeled by the equation  $\pi_i =$

$P(Y_i = 1 | \mathbf{x}_i)$ , where  $Y_i = 1$  represents the subject  $i$  is classified into the high-level group of the phenotype defined in previous sections, and  $\mathbf{x}_i$  is the observed phyla data after log-ratio transformation obtained on subject  $i$ . The general linear logistic regression model is

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha + \boldsymbol{\beta}^T \mathbf{x}_i$$

where  $\alpha$  is the intercept parameter, and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$  is the vector of  $d$  slope parameter. The linear predictor  $\eta_i = \alpha + \boldsymbol{\beta}^T \mathbf{x}$  is estimated by

$$\hat{\eta}_i = \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \mathbf{x}_i$$

where  $\hat{\alpha}$  is the maximum likelihood estimator (MLE) of  $\alpha$ , and  $\hat{\boldsymbol{\beta}}$  is the MLE of  $\boldsymbol{\beta}$ . The PROC LOGISTIC procedure calculates the predicted probability  $\hat{\pi}_i$  of classifying a subject into the “high level” category of the phenotype using the following formula:

$$\hat{\pi}_i = \frac{1}{1 + e^{-\hat{\eta}_i}}$$

If the predicted probability  $\hat{\pi}_i$  is greater than 0.5, the subject  $i$  will be classified into the high-level group according to each of the five outcome variables, otherwise it will be classified into the low-level group.

In PROC LOGISTIC procedure, the parameters of the logistic regression model were re-estimated by using LOO method. The new estimated parameters were applied to classify the subjects. Therefore, the model was fitted by leaving out each observation one at a time. In practice, the linear predictors and estimates of parameters can be stored by using “OUTPUT OUT = ...” in the PROC LOGISTIC procedure. The classification table can be called out by using TABLE statements in a PROC FREQ procedure.

### **2.3.1 Diagnostics for Logistic Regression Models**

Sensitivity is the proportion of observations the model predicts to be in the high-level group of a phenotype when they indeed belong to the high-level group. Specificity is the proportion of observations the model predicts to be in the low-level group of a phenotype when they indeed belong to the low-level group. The receiver operating characteristic (ROC) curve plots the sensitivity against 1 minus the specificity. The estimated probabilities for high level groups are listed in descending order so that the ROC curve is non-decreasing. For a logistic regression model with high predictive accuracy, the ROC curve will be above the diagonal line. The diagonal line corresponds to the case when the model randomly classifies subjects into the two groups, similar to making decisions by tossing a fair coin. Thus, the area under the curve (AUC) is large for a model with high predictive accuracy. Conversely, the ROC curve will be close to the diagonal line and have a small AUC, if logistic regression model has low prediction accuracy.

## **2.4 Other Classification Methods**

### **2.4.1 Decision tree**

Decision tree model is one of the machine learning methods developed for solving regression and classification problems, so that it is also called Classification and Regression Tree (CART) model. The CART is one type of supervised learning algorithm that can be performed on predicting categorical and continuous outcome variables. The main idea of CART is to partitioning the data into multiple sub-spaces repeatedly to obtain the best split containing the most

homogeneous outcome in each final sub-space. The results of this algorithm are generally visualized as a binary tree composed of decision nodes, branches and leaf nodes. Each of the predictor variables introduced into CART model is corresponded to a decision node with the best split rule produced by the algorithm. The leaf nodes are used to make predictions of the outcome variable. The branch is an entire sub-partition. The recursive partitioning continues until all leaf nodes are homogeneous with one single class, or a pre-specified minimum number of observations cannot be classified to each of the leaf nodes.

In classification tree model, three measures of purity are generally used, including classification error rate, Gini index and the entropy. In this study, we performed the Gini index to measure how many observations in the training set in a particular region belongs to a single class. Instead of separating data into training and testing sets randomly, we used LOO method to maintain consistency and reasonable size. If the observations in a region  $R$  are mostly from a single class  $c$ ,  $c = 1, 2$ , then the Gini index can be calculated by the following equation:

$$G = \sum_{c=1}^2 \hat{p}_c (1 - \hat{p}_c)$$

where  $\hat{p}_c$  is the proportion of the training data in region  $R$  that belong to class  $c$ .

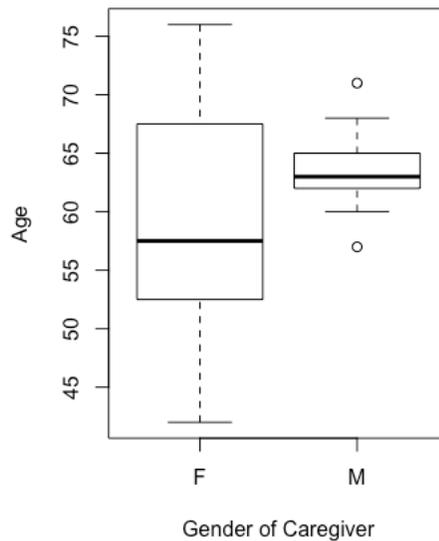
## 2.4.2 Random forest

Random forest (RF) is another tree-based machine learning algorithm for classification. The main idea of the RF model is to choose the classification by growing many decision trees using some of the predictor variables including the five phyla variables we transferred in the previous section. We assume the size of the training set is  $m$ , and randomly choose  $m$  samples from the original data set with replacement as the training data. Then we select  $p$  predictors out of

all five variables we have in the original data set (e.g. 4 out of 5). Each of the classification trees will be fully grown using the training data with  $p$  dimensions. The results of the classification trees are aggregated to predict the mental and physical health outcome including stress, depression, hostility and patient-caregiver relationship, and metabolic syndrome. The decision rule for each of the classification trees is using the Gini Index method.

### 3.0 Results

The sample consisted of 20 female and 10 male caregivers, and the box-plot of age distribution by gender is provided in Figure 1. The overall mean age of all participants was 60.5 year, the mean age of males was 63.5 and that of females was 59.1 years. Other demographic variables are summarized in Table 2.



**Figure 1** Boxplot of Ages in Each Gender of Caregivers

**Table 2 Demographic Variables Summary**

Characteristics	N (%)
Marital Status	
Never married	3 (10.0)
Married	24 (80.0)
Living with others	3 (10.0)
Race	
African American	1 (3.3)
White	28 (93.3)
White/American Indian	1 (3.3)
Education	
High School	10 (33.3)
Graduate	7 (23.3)
Some college	13 (43.3)

### **3.1 Classification**

The analysis is based on “phylum” level of bacterial phylogeny. The sample size of two groups for each phenotype and percentage of correctly classified observations by LDA model are summarized in Table 3. The range of test sizes is from 14 to 16. According to the results, the LDA classification performs well for some categories of these variables. In classification of caregiver stress, the LDA did not perform well in low level of stress with only 37.5% correction. The correct classification rate in depression is greater than 50% for low level of depression. The classification in hostility does not perform well for both two categories of this variable. The results also show good predictions for low level of patient-caregiver relationship and for having metabolic syndrome. But the classification does not perform well in high level of patient-caregiver relationship and no metabolic syndrome groups.

**Table 3 Percentages of Correct Classification**

Phenotype	Group	Size (n)	LDA Correct %	LR Correct %	CART Correct %	RF Correct %
Caregiver Stress	Low Stress	16	37.5	75.0	31.3	37.5
	High Stress	14	50.0	50.0	64.3	14.3
	Total	30	43.3	63.3	46.7	26.7
Depression	Low Depression	21	52.4	90.5	85.7	95.2
	High Depression	9	44.4	55.6	11.1	11.1
	Total	30	50.0	80.0	63.3	70.0
Hostility	Low Hostility	15	40.0	66.7	33.3	53.3
	High Hostility	15	40.0	53.3	13.3	46.7
	Total	30	40.0	60.0	23.3	50.0
Patient-caregiver Relationship	Low DAS	16	87.5	75.0	68.8	43.8
	High DAS	14	14.3	57.1	35.7	50.0
	Total	30	53.3	66.7	53.3	46.7
Metabolic Syndrome	No Syndrome	14	21.4	42.9	50.0	35.7
	Have Syndrome	15	86.7	66.7	60.0	46.7
	Total	29	55.2	53.3	55.2	41.4

According to the results in Table 3, logistic model generally outperformed LDA in terms of total correct classification rates for all phenotypes considered in this study. For some phenotypes, such as depression, the gains were substantial (80% vs. 50%). The logistic model classified subjects into low stress group with 75% accuracy, and 50% accuracy into high stress group. The accuracy of classification into low depression was 90.5%, whereas it was 55.6% for classification into high depression group. In the classification of hostility, subjects with low hostility level were classified with 66.7% accuracy, while subjects with high hostility level were classified with 53.3% accuracy. Subjects with low level of patient-caregiver relationship were classified correctly with 75% rate, and the subjects with high level of patient-caregiver relationship

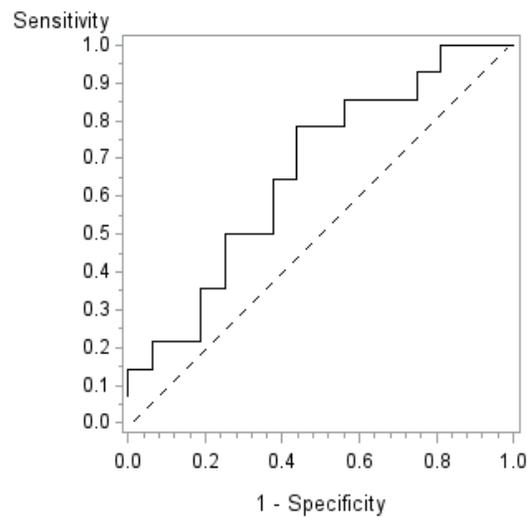
were classified correctly with 57.1%. However, the prediction of low level of metabolic syndrome had correct classification rate less than 50%. The results of logistic regression models are summarized in Table S1. The odds ratios of each covariate in these five models were not statistically significant, because the 95% confidence intervals of the odds ratios contained 1. However, although the p-value for the odds ratios of each of covariates was greater than 0.05, some were close to 0.05 which cannot be ignored. For the depression model, the p-value for the covariate Proteobacteria was 0.12 and for the Others was 0.06. The odds ratio of the covariate Proteobacteria was 0.5, which means that with one unit increase of the proportion of Proteobacteria and Bacteroidetes, the risk of depression will decrease 50%. In the same model, the odds ratio of the covariate Others was 0.61, which means that with one unit increase of the proportion of Others and Bacteroidetes, the risk of depression will decrease 39%. For the caregiver-patient relationship model, the odds ratio of the covariate Proteobacteria was 0.38 with p-value 0.07. With one unit increase of the proportion of abundance of Proteobacteria and Bacteroidetes, the risk of poor relationship will decrease 62%.

According to the results of other two machine learning algorithms on classification, CART models had poor overall correct classification rates on caregiver stress and hostility, which were 46.7% and 23.3% respectively, and RF models had poor performance on predicting caregiver stress, patient-caregiver relationship and metabolic syndrome, which were 26.7%, 46.7% and 41.4% respectively. For some categories of health outcome variables, such as the low-level of depression and patient-caregiver relationship, CART models performed well.

For each of the five phenotypes of interest, ROC plots were obtained for the logistic regression model. Comparison of ROC plots and the area under the curves (Figure 2-6), we note

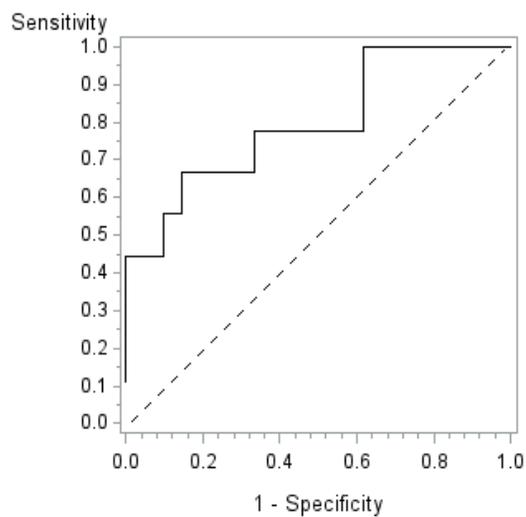
that the model for depression has the highest prediction accuracy while metabolic syndrome model has the lowest predictive accuracy.

ROC curve for caregiver stress is shown in Figure 2. The plot is above the diagonal line with an AUC of 0.67.



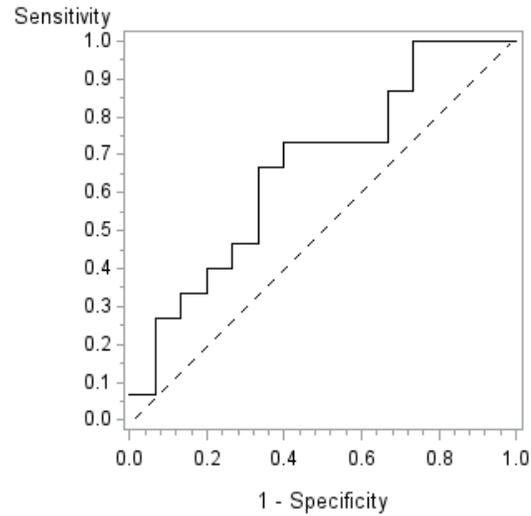
**Figure 2 ROC Curve for Stress Model (AUC = 0.67)**

ROC curve for depression is plotted in Figure 3. The curve grows quickly with an AUC of 0.80, suggesting a very good prediction accuracy.



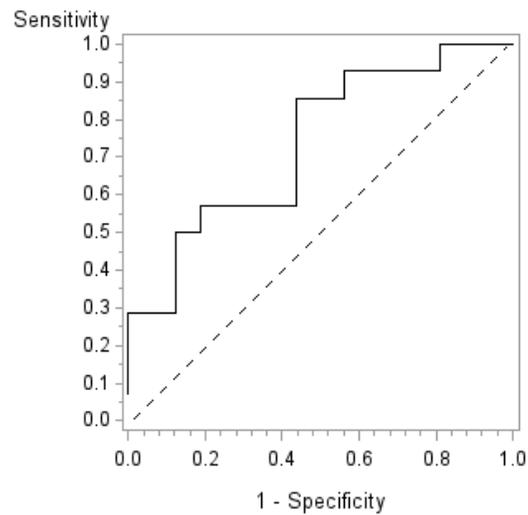
**Figure 3 ROC Curve for Depression Model (AUC = 0.80)**

ROC curve for hostility is provided in Figure 4 which has an AUC of 0.67.



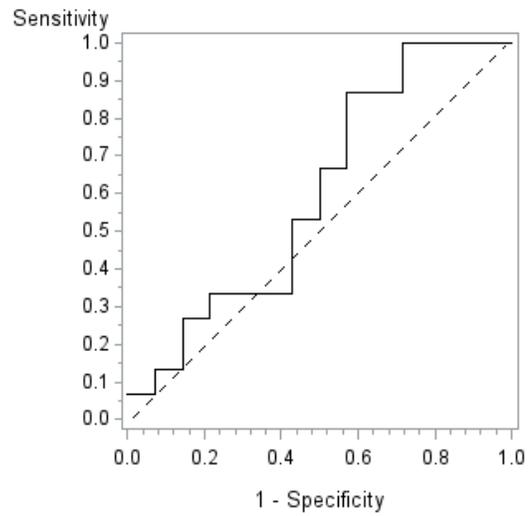
**Figure 4 ROC Curve for Hostility Model (AUC = 0.67)**

ROC curve for patient-caregiver relationship (Figure 5) is also above the diagonal line with a reasonably high AUC value of 0.74.



**Figure 5 ROC Curve for DAS Model (AUC = 0.74)**

The prediction for metabolic syndrome is not good, because the ROC sometimes crosses the diagonal line (Figure 6); the AUC is 0.6, which is the smallest area compared to other models.



**Figure 6 ROC curve for metabolic syndrome model (AUC = 0.6)**

## 4.0 Discussion

Motivation for this study was to examine whether stool microbial phyla can be used to predict diagnose various mental and physical health characteristics of cancer patient caregivers, such as depression, stress, hostility and patient-caregiver relationship, and metabolic syndrome. Thus, similar to how a physician uses blood and urine samples to diagnose various health conditions of a patient, our goal was to assess the diagnostic value of stool microbiome.

With the exception of depression and metabolic syndrome, for all other phenotypes we used the median of the respective distributions as the cutoff for the two binary groups we created. Although this approach provided balanced sample sizes for the two groups within each phenotype it may not be clinically relevant. If we had access to larger sample size then we would create more groups that are clinically relevant. Since we used reversed scores obtained from CQOLC scale, which generally was used to measure caregiver's quality of life, to measure their stress symptoms, there was no specific cutoff point for identification of two categories in this health outcome variable. Compared to the mean of the distribution of CQOLC scores from a larger sample with 239 caregivers, the mean of the scores from our sample was much lower, where the mean of our sample was 50.2 and of the larger sample was 85.1(Schulz & Beach, 1999). It illustrated that the quality of life of the caregivers involved in our study was better than the larger sample. For the cutoff point in separating the caregivers according to depression symptom in a caregiver sample, He et al. chosen 16 in their study of depression-related symptom (He, Dai, Li, He, & Huang, 2019). A 16 or higher score of CES-D scale indicated clinical depression symptom (Delisle et al., 2014).

It is important to note that Bacteroidetes and Firmicutes are two important phyla related to many mental disorders, including stress and depression (Bailey et al., 2011). When we treated

Bacteroidetes as the reference phylum in Aitchison's log-transformation, it may seem as though there is some loss of biological information. However, it is the appropriate thing to do because it is the relative proportion of these bacterial phyla that is of importance and not the individual value (Lach, Schellekens, Dinan, & Cryan, 2018)

A variety of models and tools can be developed to assess the diagnostic value of stool microbiome. In this thesis we considered two very common and simple tools that are widely used in the literature in other contexts (Johnson & Wichern, 2002), namely, LDA and logistic regression. For all five health characteristics, namely, depression, stress, hostility and patient-caregiver relationship, and metabolic syndrome, we found that the logistic regression had generally higher total correct prediction/classification rates for classifying subjects into their correct health categories than LDA. This is possibly due to the fact that, unlike logistic regression, LDA relies on: (a) The assumption that the population covariance matrices for the two groups within a phenotype are equal. This assumption may not be realistic in many applications. Furthermore, when the sample size within each group is small then it is hard to test whether this assumption is valid or not. Perhaps it would be more robust not to make this assumption, however in such a case one needs to estimate the covariance matrices for the two groups separately and then use quadratic discriminant analysis (QDA), a nonlinear function. Unfortunately, however, that would require larger sample sizes within each group to consistently estimate the covariance matrices and would also require normality assumption which may not be reasonable (Johnson & Wichern, 2002). Note that LDA does not make normality assumption (Johnson & Wichern, 2002). (b) Estimator of the precision matrix (inverse of the covariance matrix) is not stable for small sample sizes such as in our present study. For these reasons LDA may not be very stable. On the other hand, as far as prediction problem is concerned, the logistic regression approach does not require explicit

estimation of any such matrices. Hence logistic regression approach for prediction purposes is expected to perform well. Indeed, this appears to be the case in our study. For all variables considered in this project, the total percent correct classification/prediction using logistic regression was much higher than that of LDA. Using the logistic regression approach, we found the overall correct classification rates for depression and patient-caregiver relationship to be 80% and 67%, respectively. Thus, there is an 80% chance of correctly predicting the depression symptom category (low or high) of a caregiver using the microbial phyla. The correct classification rates for caregiver stress, hostility and metabolic syndrome were about 63%, 60% and 53%, respectively. The logistic regression model had higher predictive accuracies for depression and metabolic syndrome compared to other outcomes according to the ROC curves and AUCs. This finding is consistent with a study on the oral microbiome using logistic regression (Si, Lee, & Ko, 2017). Although logistic regression model performed better in total percent correct prediction, LDA model had better predictive abilities on low level group of patient-caregiver relationship and high-level group of metabolic syndrome.

In addition to LDA, QDA and logistic regression analysis, there are numerous other methods available in the literature that were not considered in this thesis. One could explore other classification strategy such as Support Vector Machine (SVM). Last but not least, the limitation due to small sample size cannot be ignored in this study. All of the above methods require larger sample sizes to efficiently perform classification and to estimate efficiently the correct classification rates. The traditional approach of creating training and test sets to estimate the correct classification rates is not feasible with small sample sizes because it will result in large uncertainty estimates for classification rates. To solve this problem, we used the Leave-One-Out method in this study (Johnson & Wichern, 2002). One can explore other resampling based methods such as

the bootstrap-based methodology (Kim, 2009). In future work, we can perform and compare the LOO and resampling strategies to evaluate which one is better to fit the microbiome data. Although we analyzed the data at the phylum level, we could have analyzed the data at other levels of phylogeny such as family or genus using the methodology developed in (Kaul, Davidov, et al., 2017).

This was a proof of concept study using a small pilot data that has potential clinical relevance. Although the results obtained here are based on a small sample size, they are encouraging. A future goal is to create a clinical marker of health for various phenotypes discussed in this study, namely, depression, stress, hostility, patient-caregiver relationship, or metabolic syndrome using a stool sample from a caregiver. To accomplish this, using stool specimens from a large cohort of cancer patient caregivers, one may develop classifiers based on logistic regression or other methods discussed above. Since for each subject we convert the simplex data to Euclidean space data by taking relative abundances of taxa with respect to Bacteroidetes, the resulting classifiers are not affected by any artifacts of batch effects, just like the ANCOM methodology (Mandal et al., 2015). Thus, once a classifier is derived using a large cohort of subjects, it can be applied to a new subject's data. In the pilot study we did not include age, gender and other metadata as potential confounders because we had very small sample size in the pilot study. Since microbiome data as well as some of the phenotypes are sensitive to age and other metadata, one can model age and other potential confounders when building the classifier using a large cohort. Including such metadata will improve the probability of correct classification as well as the AUC.

In summary, research conducted in this study using a small pilot data lays foundation for a larger clinically relevant study in the future.

## Appendix Supplemental Table

**Table S 1 Odds Ratios of Logistic Model**

Model <sup>1</sup>	Covariates <sup>2</sup>	p-value	OR	95% LCI	95% UCI
Stress	Actinobacteria	0.64	0.92	0.64	1.32
	Firmicutes	0.44	1.65	0.46	5.89
	Proteobacteria	0.26	0.64	0.30	1.39
	Verrucomicrobia	0.93	0.98	0.70	1.38
	Others	0.51	1.16	0.75	1.77
Depression	Actinobacteria	0.99	1.00	0.67	1.50
	Firmicutes	0.55	0.62	0.13	2.94
	Proteobacteria	0.12	0.50	0.20	1.21
	Verrucomicrobia	0.28	0.80	0.54	1.20
	Others	0.06	0.61	0.36	1.01
Hostility	Actinobacteria	0.38	1.19	0.81	1.76
	Firmicutes	0.43	0.55	0.13	2.38
	Proteobacteria	0.67	1.14	0.62	2.07
	Verrucomicrobia	0.38	0.84	0.57	1.24
	Others	0.86	0.96	0.63	1.47
Patient-caregiver Relationship	Actinobacteria	0.73	0.93	0.62	1.39
Metabolic Syndrome	Firmicutes	0.82	0.84	0.19	3.70
	Proteobacteria	0.07	0.38	0.13	1.10
	Verrucomicrobia	0.52	0.88	0.59	1.31
	Others	0.71	1.09	0.70	1.70
Metabolic Syndrome	Actinobacteria	0.47	1.15	0.80	1.65
	Firmicutes	0.74	1.23	0.36	4.18
	Proteobacteria	0.81	0.93	0.53	1.64
	Verrucomicrobia	0.43	1.15	0.81	1.62
	Others	0.40	1.21	0.78	1.86

<sup>1</sup> Variables in this column are the outcomes in the logistics model with three levels.

<sup>2</sup> Each covariate was transferred to a natural log of the ratio of the variable and the reference variable.

## Bibliography

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2), 139-160.
- Bailey, M. T., Dowd, S. E., Galley, J. D., Hufnagle, A. R., Allen, R. G., & Lyte, M. (2011). Exposure to a social stressor alters the structure of the intestinal microbiota: implications for stressor-induced immunomodulation. *Brain, behavior, and immunity*, 25(3), 397-407.
- Berk, M., Williams, L. J., Jacka, F. N., O'Neil, A., Pasco, J. A., Moylan, S., . . . Byrne, M. L. (2013). So depression is an inflammatory disease, but where does the inflammation come from? *BMC medicine*, 11(1), 200.
- Cani, P. D. (2018). Human gut microbiome: hopes, threats and promises. *Gut*, 67(9), 1716-1725.
- Caricilli, A. M., & Saad, M. J. (2013). The role of gut microbiota on insulin resistance. *Nutrients*, 5(3), 829-851.
- De Luca, C., & Olefsky, J. M. (2008). Inflammation and insulin resistance. *FEBS letters*, 582(1), 97-105.
- Delisle, V. C., Kwakkenbos, L., Hudson, M., Baron, M., Thombs, B. D., & Group, C. S. R. (2014). An assessment of the measurement equivalence of English and French versions of the Center for Epidemiologic Studies Depression (CES-D) Scale in systemic sclerosis. *PLoS One*, 9(7).
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.
- Hadrich, D. (2018). Microbiome research is becoming the key to better understanding health and nutrition. *Frontiers in genetics*, 9, 212.
- He, J., Dai, W., Li, Y., He, L., & Huang, R. (2019). Frequency of Depression-Related Symptoms in Caregivers of Patients with Silicosis. *Journal of healthcare engineering*, 2019.
- Huang, P. L. (2009). A comprehensive definition for metabolic syndrome. *Dis Model Mech*, 2(5-6), 231-237. doi:10.1242/dmm.001180
- Hunsley, J., Best, M., Lefebvre, M., & Vito, D. (2001). The seven-item short form of the Dyadic Adjustment Scale: Further evidence for construct validity. *American Journal of family therapy*, 29(4), 325-335.
- Ji, J., Zöller, B., Sundquist, K., & Sundquist, J. (2012). Increased risks of coronary heart disease and stroke among spousal caregivers of cancer patients. *Circulation*, 125(14), 1742-1747.
- Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., . . . Shi, J. (2015). Altered fecal microbiota composition in patients with major depressive disorder. *Brain, behavior, and immunity*, 48, 186-194.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5): Prentice hall Upper Saddle River, NJ.
- Kany, S., Vollrath, J. T., & Relja, B. (2019). Cytokines in Inflammatory Disease. *Int J Mol Sci*, 20(23). doi:10.3390/ijms20236008
- Kaul, A., Davidov, O., & Peddada, S. D. (2017). Structural zeros in high-dimensional data with applications to microbiome studies. *Biostatistics*, 18(3), 422-433.
- Kaul, A., Mandal, S., Davidov, O., & Peddada, S. D. (2017). Analysis of microbiome data in the presence of excess zeros. *Frontiers in microbiology*, 8, 2114.

- Kho, Z. Y., & Lal, S. K. (2018). The human gut microbiome—a potential controller of wellness and disease. *Frontiers in microbiology*, *9*, 1835.
- Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational statistics & data analysis*, *53*(11), 3735-3745.
- Lach, G., Schellekens, H., Dinan, T. G., & Cryan, J. F. (2018). Anxiety, depression, and the microbiome: a role for gut peptides. *Neurotherapeutics*, *15*(1), 36-59.
- Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*, *26*(1), 27663.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401.
- Reber, S. O., Siebler, P. H., Donner, N. C., Morton, J. T., Smith, D. G., Kopelman, J. M., . . . Hassell, J. E. (2016). Immunization with a heat-killed preparation of the environmental bacterium *Mycobacterium vaccae* promotes stress resilience in mice. *Proceedings of the National Academy of Sciences*, *113*(22), E3130-E3139.
- Scherwitz, L., Perkins, L., Chesney, M., & Hughes, G. (1991). Cook-Medley Hostility scale and subsets: relationship to demographic and psychosocial characteristics in young adults in the CARDIA study. *Psychosomatic Medicine*.
- Schulz, R., & Beach, S. R. (1999). Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. *Jama*, *282*(23), 2215-2219.
- Shin, N.-R., Whon, T. W., & Bae, J.-W. (2015). Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in biotechnology*, *33*(9), 496-503.
- Si, J., Lee, C., & Ko, G. (2017). Oral microbiota: microbial biomarkers of metabolic syndrome independent of host genetic factors. *Frontiers in cellular and infection microbiology*, *7*, 516.
- Smith, T. W. (2003). Hostility and health: Current status of a psychosomatic hypothesis.
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, 15-28.
- Steel, J. L., Cheng, H., Pathak, R., Wang, Y., Miceli, J., Hecht, C. L., . . . Marsh, W. (2019). Psychosocial and behavioral pathways of metabolic syndrome in cancer caregivers. *Psycho - oncology*, *28*(8), 1735-1742.
- Vaccarino, V., McClure, C., Johnson, B. D., Sheps, D. S., Bittner, V., Rutledge, T., . . . Krantz, D. S. (2008). Depression, the metabolic syndrome and cardiovascular risk. *Psychosomatic Medicine*, *70*(1), 40-48.
- Weitzner, M. A., Jacobsen, P., Wagner, H., Friedland, J., & Cox, C. (1999). The Caregiver Quality of Life Index—Cancer (CQOLC) scale: development and validation of an instrument to measure quality of life of the family caregiver of patients with cancer. *Quality of life Research*, *8*(1-2), 55-63.