Distribution Agreement

Signature:

______________________________	______________
Qi Meng	Date

Analyses for the association between Business travel and BMI

By

Qi Meng
Master of Public Health

Department of Biostatistics and Bioinformatics

Reneé H. Moore
Committee Chair

Michele Marcus
Committee Member

Sharon Bergquist
Committee Member

Analyses for the association between Business travel and BMI

By

Qi Meng

B.S.
Tianjin University
2018

Thesis Committee Chair: Reneé H. Moore, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Department of Biostatistics and Bioinformatics
2020

Abstract

Analyses for the association between Business travel and BMI
By Qi Meng

Background: Employees are traveling more as part of their work nowadays. Traveling can add stress and represents a challenge to maintaining healthy habits. Crossing time zones poses additional risks of jet lag and circadian rhythm disruption. Numerous research studies show that business travelers have a higher probability of contracting infections. However, only a few studies have assessed the noninfectious outcomes associated with business travel. In this study, we focused on the health risks associated with high-frequent business travel (domestic, international travel, as well the combination of two types), and assessed the specific association between business travel frequency and BMI/obesity.

Methods: A total of 795 executives that were seen at Emory Executive Health between January 2017 and June 2018 were examined in this retrospective study. Summary statistics of demographic and lifestyle characteristics are presented overall and by gender. Univariate associations with travel frequencies (in two ways: continuous, four categories, three types: domestic, international, combination) and BMI outcomes (in three ways: continuous, Binary Obesity, Multilevel Obesity) were conducted. Univariate and multivariable analyses using multinomial logistic regression model, cumulative logit model and linear regression model assessed the specific association between travel frequency and BMI (continuous, Binary Obesity, Multilevel Obesity). Further, multinomial logistic regression model, cumulative logit model and linear regression model were compared to give a recommendation for future studies.

Conclusion: International travel frequency had a significant effect on BMI (continuous or Multilevel Obesity) among males. Specifically, the health outcomes were worse for those not travelling, travelling seldomly and those traveling extensively (more than 13 times per month in this study). In other words, the BMI displayed a U-shaped pattern of associations with the business travel frequency. In addition, among three modeling methods, multinomial logistic regression models performed best for assessing the specific association between travel frequency and BMI (Multilevel Obesity).

KEYWORDS: travel frequency, BMI, obesity, multinomial logistic regression, cumulative logit model

Analyses for the association between Business travel and BMI

By

Qi Meng

B.S.
Tianjin University
2018

Thesis Committee Chair: Reneé H. Moore, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Department of Biostatistics and Bioinformatics
2020

Contents

1. Introduction

With the development of economic globalization, businesses are expanding into more areas, and many employees are traveling as part of their work. According to Global Business Travel Association[1] and National Travel & Tourism Office[2], Americans took more than 500 million domestic business trips and more than 4.5 million international business trips in 2016. Traveling can add stress, making it harder to maintain healthy eating, exercising, and sleeping habits, leading to worse health.[3] Additionally, crossing time zones poses the additional risks of jet lag and circadian rhythm disruption.[4]

There are numerous research studies on the association between business travel and health risk, showing that business travelers have a higher possibility of getting infections, including malaria, acute unspecified diarrhea, and viral syndromes. This trend is especially pronounced in those who take international trips to some developing countries in sub-Saharan Africa, Southeast Asia, South Central Asia.[5-8] According to these findings, multiple measures have been taken to help prevent travel-related disease: providing corresponding immunizations in advance, researching travel medicines, and giving pre-travel health assessment.[9] However, only a few research studies have studied the noninfectious outcomes associated with business travel.

A study in 2010, comparing thousands of international travelers and nontravelers, found frequent international business travel was significantly associated with lower body composition index, increased alcohol consumption, as well as sleep deprivation. Frequent international business travel also diminished employee’s confidence to keep up with the pace of work.[10] The study in the medical literature, authored by Dr. Andrew Rundle and his colleagues at the Mailman School of Public Health at Columbia, found a strong correlation between frequency of travel and physical and emotional health risks.[11,12] Looking at electronic medical records of 13,000 business travelers from EHE International, Inc (EHE), they found that compared to those who spent one to six nights a month away from home for business travel, those who spent 14 or more nights away from home per month had significantly higher body mass index, obesity index, anxiety, depression, lack of physical activity, alcohol use, smoking, trouble sleeping, as well as poor self-rated health. In addition, those who spent 21 or more nights away from home per month had slightly higher diastolic blood pressure and lower high-density lipoprotein.[11,12] The association between business travel and mental health can also be suggested by the fact that anxiety and depression are some of the top issues reported by travelers according to call records of risk management companies. Nearly 40 percent of calls to International SOS, the largest medical assistance and security company worldwide, involve issues of mental health, and this frequency is increasing continuously.[13]

We believe high-frequency business travel results in excess risks for poor health and want to further identify the health risks associated with business travel. This thesis is a retrospective study of executives that were seen at Emory Executive Health between January 2017 and June 2018. The dataset was derived from a project with our Biostatistics Collaboration Core and investigators Dr. Sharon Bergquist and Dr. Michelle Marcus. By analyzing data obtained through the routine course of executives' medical care, we examined whether frequent business travel was significantly associated with risk factors for chronic disease including obesity, BMI, body composition (e.g. body fat percentage), and lifestyle habits (alcohol and tobacco use, exercise, amount of work and sleep). In particular, we assessed the specific relationship between business travel frequency and BMI/obesity, adjusting for covariates related to body composition, lifestyle habits.

2. Materials and Methods

2.1 Data Collection

Much of the details for data collection outlined below are directly from Dr. Bergquist's study proposal for the project with Biostatistics Collaboration Core, the study was approved by Emory IRB.

Annually, Emory Executive Health provides physical examinations to 1300 business executives requiring them to complete a health history questionnaire, routine labs and preventive tests. The health history questionnaire records the information about travel frequency, exercise habits, amount of sleep, as well as tobacco and alcohol use, while physical examination and lab tests records the information for blood pressure, body mass index, body composition, cholesterol, and blood sugar, assessing the risks of potential diseases.

This study is a chart review of routine medical care records for executives between January 2017 and June 2018. Relevant clinical data is stored in four databases: 1) the Emory Healthcare computer data warehouse (CDW), 2) SalesForce (used to administer health history questionnaires and facilitate visit scheduling at Executive Health, 3) the body composition analyzer machine (SECA), 4) REDCap from a prior IRB approved study (IRB 00092476 Epigenetics and Stress), which includes stress and resilience surveys and measures on a subset of patients will be utilized. Among extracted data sets, the Health

History Questionnaires (HHQ), and SECA body composition were the main datasets utilized.

In order to link the records from different data sets, variables include medical record numbers (MRN), date of birth (DOB), and date of service (DOS) are used as identifiers. No specimens have been intentionally collected or stored for this retrospective study. The data used will continue to be stored in patient medical records as appropriate for their medical care.

2.2 Data Organization

All data organization were conducted using SAS 9.4 (Cary, NC).

2.2.1 Data Cleaning

For HHQ Lifestyle dataset:

The **HHQ Lifestyle** dataset contains the information related to the business travel frequency and lifestyle habits of observations. First, we checked and deleted duplicated rows since some records were recorded twice with mistake. Second, we cleaned business travel related variables including domestic travel frequency, international travel frequency and total travel frequency. For all of these variables, we transformed non-standard units to days/month, regarded entries indicating rare frequencies as zero, entries with slash as intervals, and entries without unit/not providing specific frequencies/greater than 30 as missing, as well as special cases including ‘depends’, ‘varies’ as ‘<1’ (refer to **Table 2.2.1** for examples of each of these). For each travel frequency, we calculated the midpoint for records with time intervals. We created a new variable-summed travel frequency, which is equal to the summation of the domestic travel frequency and international travel frequency.

Based on the midpoint values, we further created new variables to categorize above four travel frequencies into 4 levels: <1, 1-6, 7-13, >13. Third, for the lifestyle characteristic working hours, we transformed all non-standard units into hours/week, regarded entries > 100 as missing, and then calculated the midpoint. Based on the value of midpoint, we further categorized work variable into 9 levels: <30, [30-35), [35-40), [40-45), [45-50), [50-55), [55-60), [60-65), [65-70), 70+. Also, we created a new variable, worker, to identify whether the participant is a current worker or not. Here, all observations recorded as retired and with working hours less than 30 were regarded as non-worker, including 'stay home mom' and 'housewife' as this study is interested in full time work outside the home. Fourth, for sleep variable, we regarded entries <= 1 as missing. Fifth, for average # of drinks/week, we reduced the number of missing by finding correspondence between never drank and avg # of drinks, and then recategorized the variable into 6 levels: 0, 1-2, 3-5, 6-10, 11-15, 16+ drinks/week. Detailed rules for data cleaning are as shown in **Table 2.2.1**. We kept MRN, DOB, DOS, business travel frequency, as well as lifestyle habits including sleeping hours, alcohol fields, tobacco fields for further merging steps.

Table 2.2.1 Data Cleaning Rules for HHQ Lifestyle Dataset

Actions	**Details/Examples**
Continuous variables	
Business Travel Frequency:	
Deleted duplicated rows	
Transformed non-standard units to days/month	Special case: 'Nights' was coded as days; 'per vacation' regarded as per year; 'hours' was coded as 1 day.
Coded entries indicating rare frequencies as Zero	"N/A", "NA", "n/a", "0/month", "None", "none", "O", "o", "0/0", "00/00", "Minimal", "Mimimal"…
Coded entries with slash as intervals	'3/4' was coded as '3-4' Special case: '15/30' was coded as 15 out of 30 days.
Coded entries without unit, not providing specific frequencies, and ≥ 30 as missing	'a', 'yes', '??', ' ', 'Don?t know', 'Not sure', , 'alot'
Coded special entries as '<1'	'depends', 'varies'
Coded some special cases	1) mrn-13344213, changed ITF from 30 to 3;

Actions	**Details/Examples**
	2) mrn-51949895, assumed unit for DTF as ‘per year’
Calculated the midpoint.	Midpoint for interval ‘0-1’ would be 0.5.
Created a new variable - Summed Travel Frequency (STF)	STF = DTF + ITF
Work:	
Set threshold for large entry as 100.	Entries larger than 100 would be coded as missing.
Transformed characters indicating numbers into numeric numbers.	Transformed ‘None’ into 0; Transformed ‘Fifty’ into ’50.
Ignored the ‘+’	Included ‘40+’ as ‘40’.
Took the level of working hours even though marked as retired	Included ‘5-retired’ as ‘5’; ‘Retired, but 50-70’ as ’50-70’.
Calculated the midpoint.	
Worker: (created variable to identify current worker or not)	
Coded those with working hours <30 as non-workers	'O', '0/week', 'None', 'NA', 'n/a', 'none', ‘retired’, ‘unemployed’, ‘housewife’, ‘stay home mom’
Coded as worker with missing working frequencies	‘work at home’, ‘varies’, ’a’, large entries >100
Coded all retired observations as non-worker	‘5-retired’, ‘retired, but 50-70’。
Sleep:	
Coded entries <= 1 as missing.	
Categorical variables	
Travel frequency: <1, 1-6, 7-13, >13 (days/month) **Work:** <30, [30-35), [35-40), [40-45), [45-50), [50-55), [55-60), [60-65), [65-70), 70+ (hours/week)	
Approximated midpoint to the closest integer.	'0-1' has midpoint 0.5, categorized to group '<1'; '5-8' has midpoint 6.5, categorized to group '7-13'.
If sum travel frequency>30, then code all travel frequencies as missing	DTF = 15, ITF = 20, then STF = 35 > 30. Therefore, viewed all travel frequencies as missing.
Categorized variables based on above levels.	Based on values of midpoints, categorized variables into corresponding levels.
Average # of drinks/week: 0, 1-2, 3-5, 6-10, 11-15, 16+.	
Reduced the number of missing	Found correspondence between Never Drank and Avg # of drinks, and 0 drinks/week for observations that never drank.
Re-categorized the variable	Split 0-2 to two levels: 0, 1-2; and combine last three levels to 16+.

For HHQ Exercise dataset:

The **HHQ Exercise** dataset records the exercise time of different types of activity for each participant (identified by MRN). First, since some exercise activities were recorded multiple times for the same date with identical information, we verified identical information and deleted these duplicated rows. Second, we regarded some special entries as missing. Finally, we summed up exercise time for each type of activity for each MRN to get their total exercise time and changed the unit to hours per week. Detailed rules for data cleaning are as shown in **Table 2.2.2**. We kept MRN and total exercise time for the further merging step.

Table 2.2.2 Data Cleaning Rule for HHQ Exercise Dataset

Actions	**Details**
Deleted the duplicated rows with same activity details.	Some exercise activities are recorded more than once by mistake, verified identical information and deleted real duplicated rows.
Some special cases	Coded entry-3019 in 'minutes our session' and entries >= 21 in 'times per week' as missing.
Calculated the total exercise time for each MRN.	Summed up exercise time for each type of activity.
Changed the unit to hours/week	Divided previous values of minutes/week by 60, kept two decimal points.
Set the threshold for implausible entry as 25 hours/week, given FT workers	Values > 25 hours per week, coded as missing.

For SECA (SECA-Lifestyle) dataset:

The **SECA** dataset records the information related to body composition. Due to the non-standard MRN format in **SECA** dataset, we conducted data cleaning steps for **SECA** variables after merging it with the **HHQ Lifestyle** dataset. In other words, we cleaned **SECA-Lifestyle** dataset. (More details can be found in following data merging part.) First, we calculated the average of all variables related to BMI and body composition for the observations with same MRNs, and only kept the averaged values for those observations (one row for each MRN). Then, we derived four variables: 1) obesity, which identifies whether BMI ≥ 30 or not, represented by 1 or 0; 2) obesity3, which identify different classes of obesity based on the value of BMI, with 3 levels (**Table 2.2.3)**; 3) body fat percentage (%), which equals to absolute fat mass value divided by weight value; 4) age, which equals to DOS (date of service) minus DOB (date of birth). Detailed rules for data cleaning are as shown in **Table 2.2.3**. We kept BMI fields, body composition, DOB, DOS, age, gender, ethnicity for further merging step.

Table 2.2.3 Data Cleaning Rule for SECA (SECA-Lifestyle) dataset

Actions	**Details**
Averaged the values of variables related to BMI or body composition for observations with duplicated MRNs.	There are some observations with same MRNs in this dataset, we decided to calculate the average of variables related to BMI or body composition for the observations with same MRNs, and only keep the averaged values for those observations.
Obesity (new variable to identify obese or not)	
Derived new variable based on BMI	If BMI ≥ 30 then obesity=1, otherwise obesity=0.
Obesity3 (new variable to identify different classes of obese)	
Derived new categorical variable based on BMI value	Normal (BMI < 25) Overweight (25 ≤ BMI < 30) Obesity (BMI ≥ 30)
Body fat percentage (%)	
Derived new variable	Equals to absolute fat mass value / weight value.
Age	
Derived new variable	Calculated from DOS (date of service) and DOB (date of birth)

2.2.2 Data Merging

Due to a large portion of non-standard MRN inputs in **SECA** table (i.e. different MRN coding than MRN in clinic visit and HHQ tables), we utilized date of birth (DOB) and date of service (DOS) in order to merge the **HHQ Lifestyle** and **SECA** tables. Specifically, we first merged cleaned **HHQ Lifestyle** and **clinic visits w MRN** tables by MRNs to link DOB and DOS to the **HHQ Lifestyle** table and obtained **LifestyleDOB** table. Then we further merged **LifestyleDOB** table with **SECA** table by DOB and DOS to obtain so called **SECA-Lifestyle** dataset. We then cleaned **SECA-Lifestyle** dataset and merged it with cleaned **HHQ Exercise** table. As our interest was on the effect of business travels, we only kept current worker (worker=1) during this merging step for our analysis. The finalized data contained 795 observations. Detailed merging steps were as shown in **Figure 2.2.1**.

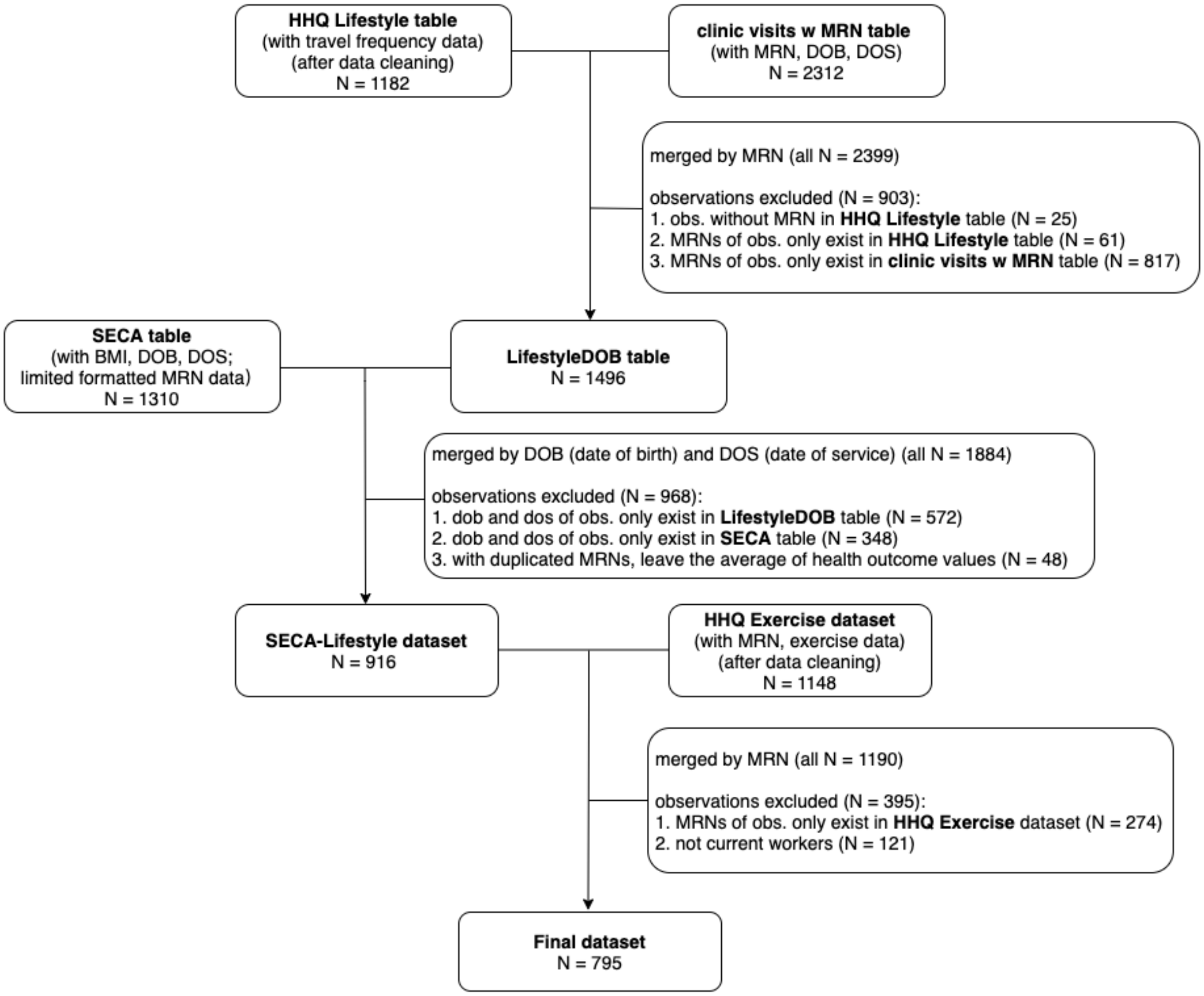

Figure 2.2.1 Flow Chart for data merging steps

2.3 Statistical Methods

All data cleaning, merging, and analyses were conducted using SAS 9.4 (Cary, NC). Descriptive analysis and separate univariate associations with primary exposures and outcomes were carried out with macros: *%DESCRIPTIVE*, *%UNI_CAT*, *%UNI_NUM* [14] created by Dr. Yuan Liu from Emory University, after revising part of the macros to report additional descriptive statistics of interest for this study. Specifically, the macro code related to the summary statistics for numeric covariates were revised to add information related to Q1 and Q3, as well as reordering outputs as follows: mean, standard deviation, minimum, Q1, median, Q3, maximum.

2.3.1 Descriptive Analysis

Descriptive statistics were reported in two tables for an overview of the overall population's characteristics, as well as for baseline characteristics stratified by gender, using the revised macro *%DESCRIPTIVE*.[14] **Table 3.1.1** contained demographic information like age, ethnicity, BMI field such as height, weight, obesity, as well as body composition. **Table 3.1.2** contained details for business travel frequency and lifestyle habits, including working hours, sleeping hours, alcohol use status, smoking status and exercise time per week. In each table, the frequency and proportion were listed for categorical covariates and summary statistics (mean, standard deviation, min, Q1, median, Q3, max) for numerical covariates, along with the frequency of missing values.

2.3.2 Univariate Associations with primary exposures/outcomes

Separate univariate associations were conducted between the primary exposure variables, travel frequencies, and all of other covariates. Here travel frequencies were assessed both as categorical and as continuous variables, including Summed Travel Frequency (STF), Domestic Travel Frequency (DTF) and International Travel Frequency (ITF) (**Tables 3.2.1 – 3.2.2.**). In addition, separate univariate associations were conducted between the outcomes of interest (continuous BMI, Binary Obesity, Multilevel Obesity) and each of the covariates (**Table 3.2.3**). When examining univariate associations for categorical variables, including categorical travel frequencies, Binary Obesity and Multilevel Obesity (normal weight, overweight, obese), a contingency table (frequency and row/column percentage, column percentage for travel frequencies, while row percentage for Binary Obesity and

Multilevel Obesity) along with the Chi-square test (parametric p-value) or Fisher's exact test (non-parametric p-value) were produced for categorical covariates, while the sample size, mean, standard deviation, minimum, Q1, median, Q3, maximum and missing frequency along with ANOVA test (parametric p-value) were produced for numerical covariates, using the revised macro *%UNI_CAT*.[14] When studying univariate associations for continuous variables, including continuous travel frequencies and BMI, the sample size, mean, standard deviation, minimum, Q1, median, Q3, maximum and missing frequency, along with ANOVA test (parametric p-value) or Kruskal-Wallis test (non-parametric p-value) were carried out for categorical covariates, while both Pearson correlation coefficient (parametric p-value) and Spearman correlation coefficient (non-parametric p-value) were calculated for numerical covariates, using the revised macro *%UNI_NUM*.[14] Considering the effect of gender, above univariate associations analysis were conducted for overall dataset as well as the datasets stratified by gender, in order to find possible differences in the associations between male and female.

Due to space limitations, the detailed result tables for all univariate associations were not included in this thesis. Instead the focus is univariate associations with categorical travel frequencies, continuous travel frequencies, as well as main health outcome-BMI/Binary Obesity/Multilevel Obesity, that are summarized in three tables (**Tables 3.2.1 – 3.2.3**). The criteria for p-value significance was set to an alpha 0.05. The criteria for absolute correlation coefficient r used to identify moderate to strong correlation was 0.15.

2.3.3 Univariate Analysis for Multilevel Obesity and Categorical Travel Frequency

According to the results from univariate associations with primary exposures/outcomes, categorical travel frequencies were more sensitive to the associations with other covariates as compared to continuous travel frequencies, while Multilevel Obesity was the most sensitive BMI outcome. Therefore, we moved forward with fitting univariate analysis (**UVA**) models with Multilevel Obesity as outcome and categorical travel frequencies as exposures separately.

Considering our outcome was with multiple levels and ordinal to some extent, we selected two modeling methods (multinomial logistic regression model and cumulative logit model) to study the exact associations between Travel Frequency and primary health outcome. Further, realizing the large difference between males and females, we decided to only study gender-stratified data for the following analysis. Also, as the results of univariate associations with Summed Travel Frequency (STF) and Domestic Travel Frequency (DTF) were quite similar, we decided not to study Domestic Travel Frequency (DTF).

Accordingly, we conducted univariate analysis between categorical STF/ITF and Multilevel Obesity with two modeling methods for males and females separately, generating overall 8 models. The Type 3 Analysis of Effects as well as odds ratios estimates calculated from each model are summarized in **Tables 3.3.1 – 3.3.8**. In addition, the assumptions for these models were checked.

The details for each modeling method are as belows:[15]

1) Multinomial Logistic Regression Model:

We generalized the model to the lowest category of Multilevel Obesity, with running index $j = 1,2,3$. Let p_{ij} be the probability that individual i falls into obesity category j. The model is then:

$$\log\left(\frac{p_{ij}}{p_{i1}}\right) = \boldsymbol{\beta}_{\boldsymbol{j}}\boldsymbol{x}_{\boldsymbol{j}}, j = 2,3$$

where $\boldsymbol{x_j}$ is the column vector of variables describing individual i, and $\boldsymbol{\beta_j}$ is a row vector of coefficients for category j. Note that each category is compared with the lowest category-"Normal". And these equations can be solved to yield

$$p_{ij} = \frac{e^{\boldsymbol{\beta_j x_j}}}{1 + \sum_2^3 e^{\boldsymbol{\beta_j x_j}}}, j = 2,3$$

As our primary exposure-STF/ITF contains 4 categories, there were 3 binary variables x_1, x_2, x_3 for travel frequency contained in Univariate Analysis models except for intercepts. With the lowest category for STF/ITF-"< 1" as the reference, x_1 identified whether in '1-6' category or not, x_2 identified whether in '7-13' category or not, while x_3 for '> 13' category.

There are many different statements about the main assumptions for the multinomial logistic regression models, and all of them emphasize the following points: the choice of or membership in one category is not related to the choice or membership of another category, the dependent variable cannot be perfectly predicted from the independent variables for any case, low collinearity, along with Independence of Irrelevant Alternatives (IIA) assumptions (adding or deleting alternative outcome categories does not affect the odds among the remaining outcomes).[16,17] However, these assumptions are not straightforward to diagnosis. When conducting multinomial regression models with SAS, what we generally do are: check for empty or small cells, check for the results of Deviance and Pearson Goodness of Fit Statistics.[18]

2) Cumulative Logit Model:

Still let p_{ij} be the probability that individual i falls into multilevel obesity category j. We assume that the categories are ordered in the sequence $j = 1,2,3$, "Normal", "Overweight", "Obese". Then we define cumulative probabilities as $F_{ij} = \sum_{m=1}^{j} P_{im}$. In words, F_{ij} is the probability that individual i is in the jth obesity category or lower. We then specify the model as a set of 3 equations,

$$\log\left(\frac{F_{ij}}{1-F_{ij}}\right) = \alpha_j + \boldsymbol{\beta x}_i, j = 1,2$$

where $\boldsymbol{\beta x_i} = \beta_1 x_{i1} + \cdots + \beta_3 x_{i3}$. And there is a different intercept for each equation. And similar to Multinomial Regression Model, 3 binary variables x_1, x_2, x_3 for travel frequency were contained in univariate analysis models except for intercepts.

The main assumption for cumulative logit model is Proportional Odds Assumption, which means the effects of any explanatory variables are consistent across the different thresholds.[19] In other words, the primary exposure – STF/ITF are supposed to have the same effect on the odds regardless of thresholds: '< 1', '1-6', '7-13', '>13' in this study.

2.3.4 Univariate Analysis for BMI and Continuous Travel Frequency

From the results for univariate associations with continuous travel frequencies, we found that none of the continuous STF/ITF/DTF was significantly associated with our primary outcomes (continuous BMI, Binary Obesity, Multilevel Obesity). However, previous tests were only conducted by Pearson (Spearman) correlation or ANOVA (Kruskal-Wallis) tests, which means only the linear/monotonic association between continuous travel frequencies and BMI (in three ways) were studied. Therefore, we decided to further study the association between business travel frequency and continuous BMI by fitting linear regression models with an additional squared term of travel frequency.[20] The univariate analysis (UVA) models were fitted to study the association between continuous STF/ITF and continuous BMI separately for males and females, generating overall 4 models. The coefficient estimates calculated from each model are summarized in **Tables 3.4.1 – 3.4.2**. Main assumptions for a linear regression model were checked, including linear relationship between independent and dependent variables, multivariate normality of all variables, no or little multicollinearity, no autocorrelation, homoscedasticity.

2.3.5 Multivariable Analysis for Multilevel Obesity and Categorical Travel Frequency

As the association between categorical travel frequency and Multilevel Obesity might be affected by other covariates, we added other covariates related to demographics, body composition, and lifestyle habits to the model for multivariable analysis (MVA). Here, we still used two modeling methods: multinomial logistic regression model and cumulative logit model. Considering the high correlation between covariates from the same category (e.g. BMI, body count composition, lifestyle habits), we selected only one covariate from each category to add into the full models, which would be used for model selection to decide the final multivariable analysis models. For body composition, the body fat percentage was selected due to its importance and representative, while current alcohol user, never smoked were selected for drink/smoke status since they were with large enough sample size in each level. And continuous working hours was selected instead of categorical working hours, because continuous working hours was more sensitive, and easier to interpret. In addition, covariates including age, ethnicity, sleeping hours, exercise time were also considered in the full model, and ethnicity was re-categorized to a binary variable with levels: ‘Caucasian’, ‘Not Caucasian’. The full models were with Multilevel Obesity as outcome, STF/ITF as main exposure (separate model for each), and other covariates including age, body fat percentage, working hours, exercise time, sleeping hours, binary ethnicity, current alcohol user, never smoked. Then model selection was conducted with backward method, using criteria p=0.05. For males and females, STF and ITF, model selection was conducted separately with two modeling methods, resulting in 8 final multivariable analysis models. The beta estimates as well as odds ratios calculated from each model are summarized in **Tables 3.5.1 – 3.5.8**. In addition, the assumptions for these models were checked.

The statistical model details for the Multivariable Analysis models were the same as those described in the Univariate Analysis models section, with the addition of covariates.

2.3.6 Multivariable Analysis for BMI and Continuous Travel Frequency

Multivariable analyses were also conducted to assess the association between continuous travel frequencies and continuous BMI. Covariates including age, body fat percentage, working hours, exercise time, sleeping hours, binary ethnicity, current alcohol user, never smoked were added to the full model for model selection. Backward selection method with criteria p=0.05 was carried out based on linear regression models. For males and females, STF and ITF, model selection was conducted separately and resulted in overall 4 final Multivariable Analysis (MVA) models. The coefficient estimates calculated from each model are summarized in **Tables 3.6.1 – 3.6.2**. In addition, the assumptions for these models were checked.

3. Results

High-frequent business travel may expose excess risks for poor health. This thesis investigated whether high frequent business travel was associated with chronic disease factors, such as: BMI, obesity, body composition, and lifestyle habits. In addition, we focused particularly on the association between BMI (continuous, Binary Obesity and Multilevel Obesity) and travel frequency when adjusting for other covariates. We conducted overall summary statistics and those stratified by gender, then univariate associations with categorical/continuous travel frequencies and BMI/ Binary Obesity/Multilevel Obesity separately. Finally, we conducted univariate analysis and multivariable analysis between categorical STF/ITF and Multilevel Obesity, along with those between continuous STF/ITF and BMI.

3.1 Descriptive Statistics

A total of 795 executives were included in this retrospective study, among which 651 (81.9%) were males and 144 (18.1%) females (**Table 3.1.1**). The average age of participants was 52, and the average age for males (52.41) was approximately two years older than the average age for females (50.18). Most of the participants (89.7%) were Caucasian, followed by African American, Other countries, or Southeast Africa. Compared with females, males had higher weight, height, BMI, fat free mass value, as well as all other covariates related to body composition except for body fat percentage. The averaged body fat percentage of overall participants was 29.15%, and average body fat percentage for males (27.56) was around 9 units smaller than those of females (36.35). Among the participants, 230 (28.9%) were normal weight (two underweight included in this category), 395 (49.7%) were overweight, and 170 (21.4%) were obese. The proportion of males being obese (22.6%) was higher than that of females (16%). While most of the males (53.9%) were overweight, most of females (53.5%) were with normal weight. Detailed demographics, BMI, and body composition information are shown in **Table 3.1.1**.

Table 3.1.1 Demographics, BMI fields and Body Composition for all covariates

Covariates	Statistics/Level	Overall	Male	Female
		N (%) = 795	N (%) = 651	N (%) = 144
age	Mean	52.01	52.41	50.18
	Std Dev	7.96	7.95	7.8
	Minimum	23	23	26
	Q1	47	47	46
	Median	52	53	50
	Q3	57	58	54.5
	Maximum	82	82	74
	Missing	0	0	0
ethnicity	African American	49 (6.2)	33 (5.1)	16 (11.1)
	Caucasian	713 (89.7)	594 (91.2)	119 (82.6)
	Other	24 (3.0)	20 (3.1)	4 (2.8)
	South East Asian	9 (1.1)	4 (0.6)	5 (3.5)
height value	Mean	5.86	5.94	5.47
	Std Dev	0.28	0.21	0.2
	Minimum	5	5.25	5
	Q1	5.67	5.83	5.33
	Median	5.92	5.92	5.5
	Q3	6	6.08	5.58
	Maximum	6.67	6.67	6.08
	Missing	0	0	0
weight value	Mean	193.05	200.86	157.76
	Std Dev	35.42	31.13	32.1
	Minimum	108.58	130.51	108.58
	Q1	169.54	178.24	135.25
	Median	191.8	197.42	151.9
	Q3	214.73	218.37	174.22
	Maximum	321.6	321.6	265.55
	Missing	0	0	0
BMI	Mean	27.39	27.76	25.76
	Std Dev	4.14	3.82	5.06
	Minimum	16.39	19.09	16.39
	Q1	24.62	25.15	22.41
	Median	26.85	27.21	24.58

Covariates	Statistics/Level	Overall N (%) = 795	Male N (%) = 651	Female N (%) = 144
	Q3	29.46	29.63	27.85
	Maximum	45.49	45.49	42.9
	Missing	0	0	0
Binary Obesity	No	625 (78.6)	504 (77.4)	121 (84.0)
	Yes	170 (21.4)	147 (22.6)	23 (16.0)
Multilevel Obesity	Normal	230 (28.9)	153 (23.5)	77 (53.5)
	Overweight	395 (49.7)	351 (53.9)	44 (30.6)
	Obesity	170 (21.4)	147 (22.6)	23 (16.1)
Absolute fat mass value	Mean	57.11	56.62	59.33
	Std Dev	20.51	19.50	24.51
	Minimum	17.14	17.14	17.86
	Q1	42.46	42.82	41.23
	Median	54.27	54.03	55.75
	Q3	67.84	67.25	70.07
	Maximum	139.48	128.90	139.48
	Missing	0.00	0.00	0.00
Fat free mass	Mean	135.95	144.24	98.44
	Std Dev	23.41	16.08	11.72
	Minimum	69.92	99.76	69.92
	Q1	124.7	132.96	89.81
	Median	138.48	142.91	98.31
	Q3	150.79	153.54	104.5
	Maximum	202.95	202.95	132.87
	Missing	0	0	0
body fat percentage (%)	Mean	29.15	27.56	36.35
	Std Dev	7.23	5.94	8.14
	Minimum	10.13	10.13	14.68
	Q1	24.19	23.35	30.53
	Median	28.97	27.67	37.28
	Q3	33.42	31.57	41.84
	Maximum	55.14	43.61	55.14
	Missing	0	0	0
Skeletal muscle mass	Mean	66.46	71.21	44.98
	Std Dev	13.45	9.24	7
	Minimum	26.21	43.73	26.21
	Q1	59.72	64.78	40.04
	Median	68.13	70.65	44.88

Covariates	Statistics/Level	Overall	Male	Female
		N (%) = 795	N (%) = 651	N (%) = 144
	Q3	74.74	76.81	48.38
	Maximum	104.83	104.83	67.17
	Missing	0	0	0
waist circumference	Mean	36.02	36.98	31.72
	Std Dev	4.77	4.21	4.81
	Minimum	0	0	24
	Q1	33	34	28
	Median	36	36	31
	Q3	39	39	35
	Maximum	51	51	49
	Missing	0	0	0
visceral adipose tissue	Mean	2.34	2.61	1.12
	Std Dev	1.41	1.37	0.79
	Minimum	0	0	0
	Q1	1.33	1.66	0.56
	Median	2.14	2.37	1
	Q3	3.1	3.32	1.53
	Maximum	8.98	8.98	5.08
	Missing	1	1	0
Resting energy expenditure	Mean	1786.05	1867.63	1417.22
	Std Dev	252.52	193.64	127.82
	Minimum	1116.66	1336.65	1116.66
	Q1	1619.36	1755.31	1334.19
	Median	1818.4	1871.18	1402.25
	Q3	1956.64	1983.32	1485.54
	Maximum	2546.61	2546.61	1824.4
	Missing	0	0	0
SMM LA	Mean	4.14	4.49	2.56
	Std Dev	1.01	0.72	0.44
	Minimum	1.66	2.34	1.66
	Q1	3.62	4.01	2.26
	Median	4.23	4.42	2.56
	Q3	4.79	4.9	2.81
	Maximum	7.41	7.41	4.02
	Missing	0	0	0
SMM LL	Mean	13.8	14.55	10.43
	Std Dev	2.52	2.01	1.72

Covariates	Statistics/Level	Overall	Male	Female
		N (%) = 795	N (%) = 651	N (%) = 144
	Minimum	6.46	9.77	6.46
	Q1	12.31	13.19	9.31
	Median	13.89	14.43	10.43
	Q3	15.3	15.61	11.31
	Maximum	22.42	22.42	15.89
	Missing	0	0	0
SMM RA	Mean	4.32	4.69	2.66
	Std Dev	1.05	0.74	0.46
	Minimum	1.66	2.37	1.66
	Q1	3.78	4.18	2.29
	Median	4.44	4.62	2.68
	Q3	4.98	5.11	2.96
	Maximum	7.51	7.51	3.94
	Missing	0	0	0
SMM RL	Mean	13.85	14.6	10.49
	Std Dev	2.53	2.04	1.69
	Minimum	6.76	9.62	6.76
	Q1	12.25	13.27	9.4
	Median	13.96	14.5	10.47
	Q3	15.41	15.67	11.38
	Maximum	22.3	22.3	15.54
	Missing	0	0	0
SMM torso	Mean	30.34	32.89	18.84
	Std Dev	6.79	4.26	3.33
	Minimum	9.54	19.56	9.54
	Q1	27.36	30.14	16.71
	Median	31.39	32.52	18.54
	Q3	34.78	35.49	20.47
	Maximum	47.72	47.72	30.94
	Missing	0	0	0

The averaged overall Summed Travel Frequency was 7.16 times per month, with 5.74 times for domestic travel and 1.42 for international travel. On average, 83% of Summed Travel Frequency came from Domestic Travel Frequency. Generally, males travelled more frequently than females, regardless of domestic or international travel. Participants slept

6.8 hours per night, worked 53 hours and exercised 222.5 minutes per week on average. Compared to females, males slept, worked, and exercised more (**Table 3.1.2**). Most of those studied were current alcohol users (87%), although not current smokers (98%), and the proportion of males who drank or smoked was larger than that of females. (**Table 3.1.2**)

Table 3.1.2 Travel Frequency and Lifestyle Habits of all covariates

Covariates	**Statistics/Level**	**Overall**	**Male**	**Female**
		N (%) = 795	**N (%) = 651**	**N (%) = 144**
Categorical STF (days/month)	< 1	49 (6.2)	28 (4.3)	21 (14.8)
	1-6	354 (45.0)	295 (45.8)	59 (41.5)
	7-13	290 (36.9)	241 (37.4)	49 (34.5)
	> 13	93 (11.8)	80 (12.4)	13 (9.2)
	Missing	9	7	2
Categorical DTF (days/month)	< 1	58 (7.4)	35 (5.4)	23 (16.2)
	1-6	453 (57.6)	376 (58.4)	77 (54.2)
	7-13	217 (27.6)	184 (28.6)	33 (23.2)
	> 13	58 (7.4)	49 (7.6)	9 (6.3)
	Missing	9	7	2
Categorical ITF (days/month)	< 1	437 (55.6)	351 (54.5)	86 (60.6)
	1-6	308 (39.2)	260 (40.4)	48 (33.8)
	7-13	36 (4.6)	29 (4.5)	7 (4.9)
	> 13	5 (0.6)	4 (0.6)	1 (0.7)
	Missing	9	7	2
Continuous STF (days/month)	Mean	7.16	7.36	6.27
	Std Dev	5.18	5.12	5.38
	Minimum	0	0	0
	Q1	3	3.75	2
	Median	6	6.17	5
	Q3	10	10	10
	Maximum	30	30	30
	Missing	9	7	2
Continuous DTF (days/month)	Mean	5.74	5.92	4.93
	Std Dev	4.51	4.52	4.4
	Minimum	0	0	0
	Q1	2	2	2
	Median	5	5	4
	Q3	8	8.42	8
	Maximum	25	25	20

Covariates	Statistics/Level	Overall	Male	Female
		N (%) = 795	N (%) = 651	N (%) = 144
	Missing	9	7	2
Continuous ITF (days/month)	Mean	1.42	1.44	1.34
	Std Dev	2.5	2.52	2.43
	Minimum	0	0	0
	Q1	0	0	0
	Median	0.07	0.13	0
	Q3	2	2	2
	Maximum	20	20	15
	Missing	9	7	2
Proportion of DTF in STF	Mean	0.83	0.83	0.82
	Std Dev	0.24	0.24	0.23
	Minimum	0.00	0.00	0.00
	Q1	0.71	0.71	0.69
	Median	0.94	0.94	0.94
	Q3	1.00	1.00	1.00
	Maximum	1.00	1.00	1.00
	Missing due to 0	42.00	24.00	18.00
	Missing	9.00	9.00	9.00
Categorical work hours (hours/week)	[30,35)	20 (2.5)	15 (2.3)	5 (3.5)
	[35,40)	9 (1.1)	5 (0.8)	4 (2.8)
	[40,45)	73 (9.2)	54 (8.3)	19 (13.4)
	[45,50)	78 (9.9)	63 (9.7)	15 (10.6)
	[50,55)	240 (30.3)	216 (33.3)	24 (16.9)
	[55,60)	116 (14.7)	84 (12.9)	32 (22.5)
	[60,65)	167 (21.1)	142 (21.9)	25 (17.6)
	[65,70)	31 (3.9)	28 (4.3)	3 (2.1)
	>= 70	57 (7.2)	42 (6.5)	15 (10.6)
	Missing	4	2	2
continuous work hours (hours/week)	Mean	53.21	53.23	53.15
	Std Dev	9.65	9.13	11.74
	Minimum	30	30	30
	Q1	50	50	45
	Median	50	50	55
	Q3	60	60	60
	Maximum	100	90	100
	Missing	4	2	2
exercise (hours/week)	Mean	3.66	3.70	3.46

Covariates	**Statistics/Level**	**Overall**	**Male**	**Female**
		N (%) = 795	**N (%) = 651**	**N (%) = 144**
	Std Dev	3.01	2.99	3.10
	Minimum	0	0	0
	Q1	1.50	1.50	1.50
	Median	3.00	3.00	2.50
	Q3	4.75	4.87	4.50
	Maximum	24.00	24.00	21.50
	Missing	89	75	14
sleeping hours (hours/night)	Mean	6.8	6.81	6.75
	Std Dev	0.88	0.83	1.05
	Minimum	4	4	4
	Q1	6	6	6
	Median	7	7	7
	Q3	7	7	7.75
	Maximum	10	10	9
	Missing	0	0	0
Current alcohol user	No	103 (13.0)	79 (12.1)	24 (16.7)
	Yes	692 (87.0)	572 (87.9)	120 (83.3)
Never Drank	No	737 (92.7)	612 (94.0)	125 (86.8)
	Yes	58 (7.3)	39 (6.0)	19 (13.2)
Drank in the past but quit	No	754 (94.8)	612 (94.0)	142 (98.6)
	Yes	41 (5.2)	39 (6.0)	2 (1.4)
average number of drinks per week	0	58 (7.7)	39 (6.3)	19 (14.1)
	1-2	198 (26.3)	151 (24.4)	47 (34.8)
	3-5	198 (26.3)	166 (26.9)	32 (23.7)
	6-10	170 (22.6)	149 (24.1)	21 (15.6)
	11-15	90 (12.0)	79 (12.8)	11 (8.1)
	16+	39 (5.2)	34 (5.5)	5 (3.7)
	Missing	42	33	9
Current Smoker	No	779 (98.0)	636 (97.7)	143 (99.3)
	Yes	16 (2.0)	15 (2.3)	1 (0.7)
Never Smoked	No	161 (20.3)	133 (20.4)	28 (19.4)
	Yes	634 (79.7)	518 (79.6)	116 (80.6)
Smokeless	No	744 (93.6)	600 (92.2)	144 (100.0)
	Yes	51 (6.4)	51 (7.8)	0 (0)

Covariates	Statistics/Level	Overall	Male	Female
		N (%) = 795	N (%) = 651	N (%) = 144
Pipe	No	793 (99.7)	649 (99.7)	144 (100.0)
	Yes	2 (0.3)	2 (0.3)	0 (0)
Cigar	No	750 (94.3)	607 (93.2)	143 (99.3)
	Yes	45 (5.7)	44 (6.8)	1 (0.7)

- Overall, there were only 2 males and no females who indicated yes for using a Pipe, 1 female indicated yes for smoking currently or using Cigars, no females indicated yes for smokeless (vaping), 2 females indicated yes for 'Drank in the past but quit'.

3.2 Univariate Associations with primary exposures/outcomes

3.2.1 Univariate Associations with categorical travel frequencies

Univariate associations between categorical travel frequencies and other covariates were conducted for overall data and gender-stratified data, using Chi-square and ANOVA tests (**Table 3.2.1**). Focusing on the main health outcomes, we found the categorical ITF was associated with Multilevel Obesity overall and for males, while categorical STF was associated with Binary Obesity for females. No significant association between categorical travel frequencies and BMI was found. Further, all types of categorical travel frequency were significantly associated with working hours. Additional significant associations between categorical travel frequencies and covariates related to body composition were found overall and for females. Generally, STF and DTF were significantly associated with similar covariates and additionally associated with smoke status, sleeping hours, as well as some covariates related to body composition.

Table 3.2.1 Univariate Associations with categorical travel frequencies

covariates	Overall			Male			Female		
	STF	DTF	ITF	STF	DTF	ITF	STF	DTF	ITF
Gender	**	**		/	/	/	/	/	/
age									
ethnicity									

covariates	**Overall**			**Male**			**Female**		
	STF	**DTF**	**ITF**	**STF**	**DTF**	**ITF**	**STF**	**DTF**	**ITF**
height value		**						**	
weight value									
BMI									
Binary Obesity							**		
Multilevel Obesity			**			**			
absolute fat mass value		**						**	
body fat percentage (%)	**	**	**						
fat free mass value	**	**							
skeletal muscle value	**	**							
waist circumference value							**	**	
visceral adipose tissue value									
resting energy expenditure	**	**							
SMM LA value	**	**							
SMM LL value							**		
SMM RA value	**	**					**		
SMM RL value							**		
SMM torso value	**	**							
categorical work hours (hours/week)	**	**	**	**	**				**
continuous work hours (hours/week)	**	**	**	**	**	**	**	**	
exercise (minuets/week)									**
sleeping hours (hours/night)	**	**		**	**	**			
current alcohol user									
Drank in the past but quit							/	/	/
Never Drank									
average number of drinks per week		**				**			
Current Smoker		**		**	**		/	/	/
Never Smoked	**	**							
Smokeless							/	/	/
Pipe	/	/	/	/	/	/	/	/	/
Cigar							/	/	/

- '**'-significant association (p-value<0.05), significant moderate to strong correlation (r>=0.15 & p-value<0.05);
  '*'-significant correlation with small r (r<0.15 & p-value<0.05);
  '/'-unable to study the association due to sparse data/small sample sizes;
  blank-no significant association or correlation.

3.2.2 Univariate Associations with continuous travel frequencies

Univariate associations between continuous travel frequencies and other covariates were conducted for overall data and stratified by gender, using ANOVA (Kruskal-Wallis) test and Pearson (Spearman) correlations (**Table 3.2.2**). Focusing on the main health outcomes, we found the continuous travel frequencies were not significantly associated with all types of BMI (continuous BMI, Binary Obesity, Multilevel Obesity). Similar to categorical travel frequencies, all types of continuous travel frequencies were significantly associated with working hours. Additional significant associations between continuous travel frequencies and smoke status, sleeping hours were found overall and for males. It was interesting that DTF had small but significant correlations with many covariates related to body composition for overall data. The same trend was found with ITF for males.

Table 3.2.2 Univariate Associations with continuous travel frequencies

covariates	**overall**			**Male**			**Female**		
	STF	**DTF**	**ITF**	**STF**	**DTF**	**ITF**	**STF**	**DTF**	**ITF**
Gender	**	**		/	/	/	/	/	/
age									
ethnicity									
height value		*				*			
weight value						*			
BMI									
Binary Obesity									
Multilevel Obesity									
absolute fat mass value		*							
fat free mass value		*				*			
body fat percentage (%)									
skeletal muscle value		*				*			
waist circumference value		*							
visceral adipose tissue value		*							
resting energy expenditure		*			*	*			
SMM LA value		*				*			
SMM LL value		*							
SMM RA value		*				*			

covariates	overall			Male			Female		
	STF	DTF	ITF	STF	DTF	ITF	STF	DTF	ITF
SMM RL value		*							
SMM torso value		*				*			
categorical work hours (hours/week)	**	**	**	**	**	**	**	**	
continuous work hours (hours/week)	**	**	*	**	**	*	**	**	
exercise (minuets/week)									**
sleeping hours (hours/night)	*	*		*	*	*			
current alcohol user									
Drank in the past but quit							/	/	/
Never Drank									
average number of drinks per week		**							
Current Smoker	**	**		**	**		/	/	/
Never Smoked	**	**					*		
Smokeless							/	/	/
Pipe	/	/	/	/	/	/	/	/	/
Cigar							/	/	/

- ‘**’-significant association (p-value<0.05), significant moderate to strong correlation (r>=0.15 & p-value<0.05);
  ‘*’-significant correlation with small r (r<0.15 & p-value<0.05);
  ‘/’-unable to study the association due to sparse data/small sample sizes;
  blank-no significant association or correlation.

3.2.3 Univariate Associations with BMI (continuous, Binary Obesity, Multilevel Obesity)

Univariate associations between BMI (continuous, Binary Obesity, Multilevel Obesity) and other covariates were conducted for overall data and gender-stratified data, using Chi-square (Fisher’s exact), ANOVA (Kruskal-Wallis) tests and Pearson (Spearman) correlations (**Table 3.2.3**). Focusing on the main exposure-travel frequency, the associations between travel frequencies and the main health outcomes were consistent with the results from above two association studies. Specifically, none of the main health outcomes (continuous BMI, Binary Obesity, Multilevel Obesity) were significantly associated with continuous travel frequencies. Further, Binary Obesity was only significantly associated with categorical STF for females and Multilevel Obesity was only significantly associated with categorical ITF overall and for males. Moreover, as we expected, all types of main outcome were significantly associated with most of covariates related to BMI (e.g. height, weight) and body composition. Additional significant associations between health outcomes and drink/smoke status, sleeping hours were found overall and for males. Interestingly, all types of BMI were significantly associated with age overall and for males. Associations between ethnicity and BMI/Binary Obesity were found overall, and significance between ethnicity and BMI/Multilevel Obesity was found for males.

Table 3.2.3 Univariate Associations with BMI (in three ways)

covariates	**overall**			**Male**			**Female**		
	BMI	**Binary Obesity**	**Multi-Obesity**	**BMI**	**Binary Obesity**	**Multi-Obesity**	**BMI**	**Binary Obesity**	**Multi-Obesity**
Gender	**		**	/	/	/	/	/	/
age	*	**	**	*	**	**			
ethnicity	**	**		**		**			
height	*		**						
weight	**	**	**	**	**	**	**	**	**
AFM	**	**	**	**	**	**	**	**	**
FFM	**	**	**	**	**	**	**	**	**
body fat (%)	**	**	**	**	**	**	**	**	**
SM	**	**	**	**	**	**	**	**	**
WC	**	**	**	**	**	**	**	**	**
VA	**	**	**	**	**	**	**	**	**
REE	**	**	**	**	**	**	**	**	**
SMM LA	**	**	**	**	**	**	**	**	**
SMM LL	**	**	**	**	**	**	**	**	**
SMM RA	**	**	**	**	**	**	**	**	**
SMM RL	**	**	**	**	**	**	**	**	**
SMM torso	**	**	**	**	**	**	**	**	**
cat STF								**	
cat DTF									
cat ITF			**			**			
con STF									

covariates	overall			Male			Female		
	BMI	Binary Obesity	Multi-Obesity	BMI	Binary Obesity	Multi-Obesity	BMI	Binary Obesity	Multi-Obesity
con DTF									
con ITF									
cat work									
con work									
exercise		**							**
sleep hours	*	**	**	*	**	**	**		**
Alcohol user	**	**	**	**	**	**			
Drank quit	**	**	**	**	**	**	/	/	/
Never Drank					**				
# of drinks							**		
Smoker							/	/	/
N-Smoked									
Smokeless	**	**	**	**	**	**	/	/	/
Pipe	/	/	/	/	/	/	/	/	/
Cigar	**	**	**	**		**	/	/	/

- '**'-significant association (p-value<0.05), significant moderate to strong correlation (r>=0.15 & p-value<0.05);
  '*'-significant correlation with small r (r<0.15 & p-value<0.05);
  '/'-unable to study the association due to sparse data/small sample sizes;
  blank-no significant association or correlation.
- Abbreviations: AFM=absolute fat mass, FFM=fat free mass, SM=skeletal muscle, WC=waist circumference, VA=visceral adipose, REE=resting energy expenditure, N-Smoked=Never Smoked, con=continuous, cat=categorical.

3.3 Univariate Analysis for Multilevel Obesity and Categorical Travel Frequency

The specific univariate analysis (UVA) between travel frequency and health outcome was studied with categorical STF/ITF as the exposure and Multilevel Obesity as the outcome for males and females, using Multinomial Logistic Regression Model and Cumulative Logit Model. The results for Type 3 Analysis of Effect as well as odds ratio estimates are summarized in **Table 3.3.1 - 3.3.8**.

3.3.1 Multinomial Regression Model

The Type 3 Analysis of Effect as well as odds ratio estimates for univariate analysis between Multilevel Obesity and categorical STF/ITF for males and females under multinomial logistic regression modeling method are as shown in **Table 3.3.1 - 3.3.4**. Only international travel frequency was found to have a significant effect on Multilevel Obesity for males (Type 3 p-value = 0.003).

Specifically, the odds that a male who traveled 1-6 times internationally per month would be obese (conditional on being normal weight) were 0.528 times the odds for a male who traveled less than 1 time internationally per month. The odds that a male who traveled 1-6 times internationally per month would be overweight (conditional on being normal weight) were 0.424 times the odds for a male who traveled less than 1 time internationally per month. The odds that a male who traveled 7-13 times internationally per month would be obese (conditional on being normal weight) were 0.955 times the odds for a male who traveled less than 1 time internationally per month. The odds that a male who traveled 7-13 times internationally per month would be overweight (conditional on being normal weight) were 0.796 times the odds for a male who traveled less than 1 time internationally per month. The odds that a male who traveled more than 13 times internationally per month would be obese (conditional on being normal weight) were 1.632 times the odds for a male who traveled less than 1 time internationally per month. The odds that a male who traveled more than 13 times internationally per month would be overweight (conditional on being normal weight) were 0.298 times the odds for a male who traveled less than 1 time internationally per month. **(Table 3.3.4 - model 3)**

The specific univariate effect of summed travel frequency on Multilevel Obesity for males (model 1), as well as the effect of summed travel frequency or international travel frequency on Multilevel Obesity for females (model2, model 4) under multinomial logistic regression model were described in the same way as model 3. However, these effects were not significant under 0.05 level.

In conclusion, for international travel among males (model 3), compared with people who traveled less than 1 times internationally per month, those who traveled 1-13 times were more likely to be normal weight, while those who traveled more than 13 times were more likely to be obese and much less likely to be overweight. For summed (combination of domestic and international) travel among males (model 1), those who traveled more than 1 times in total per month were more likely to be overweight rather than being normal weight, compared with those who travelled less than 1 time per month. For summed travel among females (model 2), those who traveled 1-6 or more than 13 times in total per month were more likely to be obese and less likely to be overweight (rather than being normal weight), compared with those who travelled less than 1 time per month. In contrast, those who traveled 7-13 times totally per month were much less likely to be obese or overweight than those who traveled less than 1 time per month. For international travel among females (model 4), those who traveled more than 1 time per month were less likely to be obese or overweight (rather than being normal weight), compared with those who travelled less than 1 time per month. Here, we found some odds ratio estimates were not finite when evaluating the effect of ITF for females. This will be discussed in the following section.

Table 3.3.1 Type 3 Analysis of Effects of UVA for STF under Multinomial Regression Model

		Male (model 1)		Female (model 2)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical STF	6	5.1549	0.5241	7.9208	0.2440

Table 3.3.2 Odds Ratio Estimates of UVA for STF under Multinomial Regression Model

Effect	Multilevel Obesity	Male (model 1)			Female (model 2)		
		Estimate	95% CI		Estimate	95% CI	
STF 1-6 vs < 1	Obese vs Normal	0.590	0.211	1.646	1.121	0.296	4.243
STF 1-6 vs < 1	Overweight	1.509	0.562	4.053	0.837	0.269	2.609
STF 7-13 vs < 1	Obese	0.653	0.232	1.834	0.161	0.026	1.017
STF 7-13 vs < 1	Overweight	1.370	0.506	3.711	0.737	0.236	2.303
STF > 13 vs < 1	Obese	0.865	0.272	2.754	1.429	0.264	7.737
STF > 13 vs < 1	Overweight	1.572	0.522	4.735	0.408	0.065	2.582

Table 3.3.3 Type 3 Analysis of Effects of UVA for ITF under Multinomial Regression Model

Effect	DF	Male (model 3) Wald Chi-Square	Male (model 3) Pr > ChiSq	Female (model 4) Wald Chi-Square	Female (model 4) Pr > ChiSq
categorical ITF	6	19.8399	0.0030	2.4791	0.8708

Table 3.3.4 Odds Ratio Estimates of UVA for ITF under Multinomial Regression Model

Effect	Multilevel Obesity	Male (model 3)			Female (model 4)		
		Estimate	95% CI		Estimate	95% CI	
ITF 1-6 vs < 1	Obese vs Normal	0.582	0.363	0.932	0.562	0.197	1.603
ITF 1-6 vs < 1	Overweight	0.424	0.285	0.632	0.919	0.414	2.039
ITF 7-13 vs < 1	Obese	0.955	0.305	2.985	<0.001	<0.001	>999.999
ITF 7-13 vs < 1	Overweight	0.796	0.299	2.121	0.276	0.031	2.420
ITF > 13 vs < 1	Obese	1.636	0.145	18.456	<0.001	<0.001	>999.999
ITF > 13 vs < 1	Overweight	0.298	0.018	4.840	<0.001	<0.001	>999.999

Considering the difficulties with checking assumptions for multinomial logistic regression model, we checked small or empty cells, as well as the results for goodness of fit (GOF) statistics (Deviance and Pearson test) instead. When checking the distribution of international travel frequency among Multilevel Obesity for females, we found there were many cells were empty or with small values (**Table 3.3.5**). In addition, as shown in **Table 3.3.6**, the goodness of fit (GOF) statistics (Deviance and Pearson test) for all models were with 0 degree of freedom and missing statistics values or p-values. Further, there was a green error warning saying ‘there is possibly a quasi-complete separation of data points’ shown in Log Window when fitting model for ITF among females.

Table 3.3.5 Univariate Association between Multilevel Obesity and ITF for females

			Multilevel Obesity			
Covariate	Statistics	Level	Normal N=77	Overweight N=44	Obese N=23	P-value
ITF (days/month)	N (Row %)	< 1	43 (50)	26 (30.23)	17 (19.77)	0.510
	N (Row %)	1-6	27 (56.25)	15 (31.25)	6 (12.49)	
	N (Row %)	7-13	6 (85.71)	1 (14.29)	0 (0)	
	N (Row %)	> 13	1 (100)	0 (0)	0 (0)	

Table 3.3.6 Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	0.0000	0	.	.
Pearson	0.0000	0	.	.

3.3.2 Cumulative Logit Model

The Type 3 Analysis of Effect as well as the odds ratio estimates for univariate analysis between Multilevel Obesity and categorical STF/ITF for males and females under cumulative logit model are as shown in **Table 3.3.7 - 3.3.10**. Neither of summed travel frequency nor international travel frequency had a significant effect on Multilevel Obesity for males and females (Type 3 p-value > 0.05) (**Table 3.3.7, Table 3.3.9).** In addition, the proportional odds assumption was violated when evaluating effect of ITF on Multilevel Obesity among males and females.

As shown in **Table 3.3.8 – model 5**, the odds of males who traveled 1-6 times per month being in a lower category of Multilevel Obesity were 1.554 times the odds for those who traveled less than 1 time per month. The odds of males who traveled 7-13 times per month being in a lower category of Multilevel Obesity were 1.460 times the odds for those who traveled less than 1 time per month. The odds of males who traveled more than 13 times per month being in a lower category of Multilevel Obesity were 1.208 times the odds for those who traveled less than 1 time per month.

The specific univariate effect of international travel frequency on Multilevel Obesity for males (model 7), as well as the effect of summed travel frequency or international travel frequency on Multilevel Obesity for females (model 6, model 8) under cumulative logit model were described in the same way as model 5.

In conclusion, males who traveled more than 1 time in total per month were more likely to protect themselves from being obese or overweight, compare with those who traveled less than 1 time per month (model 5). Females who traveled 1-6 or more than 13 times in total per month were less likely to be in a lower category of Multilevel Obesity compare with those who traveled less than 1 time per month. In contrast, those who traveled 7-13 times in total per month were more likely to be in a lower category of Multilevel Obesity (model 6). Males who traveled 1-6 times internationally per month were more likely to be in a lower category of Multilevel Obesity compare with those who traveled less than 1 time per month. In contrast, those who traveled 7-13 times internationally per month has a similar probability of being in a lower category of Multilevel Obesity. In addition, those who traveled more than 13 times per month were less likely to be in a lower category of Multilevel Obesity (model 7). Females who traveled more than 1 time in total per month were more likely to be in a lower category of Multilevel Obesity compare with those who traveled less than 1 time per month (model 8). Here, we still found some odds ratio estimates were not finite when evaluating the effect of ITF for females.

Table 3.3.7 Type 3 Analysis of Effects of UVA for STF under Cumulative Logit Model

		Male (model 5)		Female (model 6)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical STF	3	2.0903	0.5539	4.4362	0.2181

Table 3.3.8 Odds Ratio Estimates of UVA for STF under Cumulative Logit Model

	Male (model 5)			Female (model 6)		
Effect	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
categorical STF 1-6 vs < 1	1.554	0.736	3.242	0.988	0.387	2.521
categorical STF 7-13 vs < 1	1.460	0.691	3.086	2.084	0.772	5.627
categorical STF > 13 vs < 1	1.208	0.531	2.751	0.957	0.261	3.506

Table 3.3.9 Type 3 Analysis of Effects of UVA for ITF under Cumulative Logit Model

		Male (model 7)		Female (model 8)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical ITF	3	6.9642	0.0730	3.3296	0.3435

Table 3.3.10 Odds Ratio Estimates of UVA for ITF under Cumulative Logit Model

	Male (model 7)			Female (model 8)		
Effect	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
categorical ITF 1-6 vs < 1	1.468	1.078	1.999	1.372	0.694	2.715
categorical ITF 7-13 vs < 1	1.012	0.490	2.090	6.434	0.725	57.074
categorical ITF > 13 vs < 1	0.480	0.073	3.158	>999.999	<0.001	>999.999

3.4 Univariate Analysis for BMI and Continuous Travel Frequency

The specific univariate associations between travel frequencies and health outcome were also studied with continuous STF/ITF as the exposure and continuous BMI as the outcome for males and females, using linear regression models with squared term for travel frequency. The results for coefficient estimates are summarized in **Table 3.4.1 - 3.4.2**. Similar to what was found with multinomial logistic regression model, international travel frequency had a significant effect on BMI for males (p-value for ITF = 0.0125, p-value for ITF^2 = 0.0325). In addition, the assumptions for these models were met without any violations.

Specifically, as shown in **Table 3.4.2 – model 3**, ITF had a slightly negative effect on BMI, and the positive estimate of squared ITF means that the effect of ITF lessened as travel frequency increased. The same was observed for females with STF, as well as ITF/STF for males and females, though the effects were not statistically significant (p-value > 0.05).

Table 3.4.1 Coefficient Estimates of UVA for STF under Linear Regression Model

		Male (model 1)				Female (model 2)			
Parameter	**DF**	**Estimate**	**95% CI**		**Pr > ChiSq**	**Estimate**	**95% CI**		**Pr > ChiSq**
Intercept	1	27.996	27.279	28.713	<0.001	26.914	25.356	28.472	<0.001
STF	1	-0.102	-0.268	0.065	0.230	-0.293	-0.662	0.075	0.118
STF^2	1	0.006	-0.002	0.014	0.132	0.020	-0.008	0.027	0.275
Scale	1	3.821	3.618	4.036		5.024	4.472	5.643	

Table 3.4.2 Coefficient Estimates of UVA for ITF under Linear Regression Model

		Male (model 3)				Female (model 4)			
Parameter	**DF**	**Estimate**	**95% CI**		**Pr > ChiSq**	**Estimate**	**95% CI**		**Pr > ChiSq**
Intercept	1	28.012	27.645	28.378	<0.001	26.027	24.998	27.056	<0.001
ITF	1	-0.308	-0.549	-0.066	0.013	-0.297	-1.109	0.516	0.474
ITF^2	1	0.021	0.002	0.041	0.033	0.014	-0.063	0.092	0.716
Scale	1	3.811	3.608	4.025		5.060	4.505	5.684	

3.5 Multivariable analysis for Multilevel Obesity and Categorical Travel Frequency

Backward model selection with p=0.05 as criteria was conducted to decide the final models for multivariable analysis between categorical STF/ITF and Multilevel Obesity for males and females, using two different modeling methods. The full models contained additional covariates including Age, body fat percentage, working hours, exercise time, sleeping hours, binary ethnicity, current alcohol user, never smoked. Results for Type 3 Analysis of Effects and odds ratios calculated from final multivariable analysis (MVA) models are summarized in **Tables 3.5.1 – 3.5.8**.

3.5.1 Multinomial Regression Model

When examining the effect of categorical STF on Multilevel Obesity under multivariable analysis for males, the final multinomial regression models contained covariates: binary ethnicity, current alcohol user, age, sleeping hours, and body fat percentage except for travel frequency, while age was not contained in the final model when studying the effect of categorical ITF for males. For females, covariates including age, and body fat percentage were contained in the final multinomial logistic regression models for studying both categorical STF and ITF. **Table 3.5.1 - 3.5.4** summarize the Type 3 Analysis of Effect and odds ratios calculated from these models. Only ITF had a significant effect on Multilevel Obesity for males (Type 3 p-value=0.0197).

Specifically, as shown in **Table 3.5.4 – model 3**, when adjusting for all other covariates, the odds that a male who traveled 1-6 times per month would be obese (conditional on being normal weight) were 0.775 times the odds for a male who traveled less than 1 time per month. The odds that a male who traveled 1-6 times per month would be overweight (conditional on being normal weight) were 0.435 times the odds for a male who traveled less than 1 time per month. The odds that a male who traveled 7-13 times per month would be obese (conditional on being normal weight) were quite similar to (0.940 times) the odds for a male who traveled less than 1 time per month. The odds that a male who traveled 7-13 times per month would be overweight (conditional on being normal weight) were 0.744 times the odds for a male who traveled less than 1 time per month. The odds that a male who traveled more than 13 times per month would be obese (conditional on being normal weight) were 1.182 times the odds for a male who traveled less than 1 time per month. The odds that a male who traveled more than 13 times per month would be overweight (conditional on being normal weight) were 0.345 times the odds for a male who traveled less than 1 time per month. In addition, odds that a male Caucasian would not be normal weight conditional on being normal weight were around 0.25 times the odds for a male who is not Caucasian. Further, the odds that a male drinker would be overweight conditional on being normal weight were similar to the odds for a male who is not a current alcohol user, while the odds that a male drinker would be obese conditional on being normal weight were 0.375 times the odds for a male who is not a current alcohol user. Moreover, each hour increased in sleeping hours for males would 0.4-0.5 times the odds of being overweight or obese, rather than being normal weight. In addition, one unit increased in body fat percentage for males would 2.079 times the odds of being obese, or 1.393 times the odds of being overweight, rather than being normal weight.

The specific effect of summed travel frequency and other covariates on Multilevel Obesity for males and females (model 1-2), as well as the effect of international travel frequency and other covariates on Multilevel Obesity for females (model 4) under multinomial logistic regression model were described in the same way as above model 3.

In conclusion, for international travel among males (model 3), compared with people who traveled less than 1 times internationally per month, those who traveled 1-13 times were more likely to be normal weight, while those who traveled more than 13 times were more likely to be obese and much less likely to be overweight. For summed travel among males (model 1), those traveled 1-13 times in total per month were less likely to be obese and had a same probability of being overweight rather than being normal weight, compared with those who travelled less than 1 time per month. In contrast, those who traveled more than 13 times in total per month were more likely to be normal weight. For summed travel among females (model 2), compared with those who travelled less than 1 time per month, those who traveled 1-6 times in total per month were more likely to be obese and had a same probability of being overweight (rather than being normal weight). In contrast, those who traveled 7-13 times totally per month were much less likely to be obese or overweight, and those who travel more than 13 times per month were much more likely to be obese and less likely to be overweight. For international travel among females (model 4), those who traveled more than 1 time per month were less likely to be obese or overweight (rather than being normal weight), compared with those who travelled less than 1 time per month. Here, we still found some odds ratio estimates were not finite when examining the effect of ITF for females.

Additionally, other covariates including ethnicity, current alcohol/smoke status, sleeping hours had consistent effects on Multilevel obesity among males and females when assessing effect of STF/ITF. Specifically, Caucasian who were current alcohol users with more sleeping hours per night and smaller body fat percentage were more likely to be normal weight. Interestedly, increase in age would slightly decrease the probability of being obese or overweight.

Table 3.5.1 Type 3 Analysis of Effects of MVA for STF under Multinomial Regression Model

		Male (Model 1)		**Female (model 2)**	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical STF	6	3.4961	0.7445	6.0667	0.4158
Ethnicity	2	11.0453	0.0040	----	----
Current alcohol user	2	7.8126	0.0201	----	----
Age	2	6.9886	0.0304	12.6964	0.0017
Sleeping hours	2	18.1061	0.0001	----	----
Body fat percentage	2	203.3892	<0.0001	38.4880	<0.0001

Table 3.5.2 Odds Ratio Estimates of MVA for STF under Multinomial Regression Model

		Male (model 1)			Female (model 2)		
Effect	**Multi- Obesity**	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
STF 1-6 vs < 1	Obese vs Normal	0.432	0.077	2.419	5.779	0.343	97.397
STF 1-6 vs < 1	Overweight	1.011	0.291	3.510	1.171	0.195	7.037
STF 7-13 vs < 1	Obese	0.601	0.107	3.387	0.643	0.023	18.320
STF 7-13 vs < 1	Overweight	0.973	0.279	3.397	0.666	0.116	3.830
STF > 13 vs < 1	Obese	0.339	0.051	2.257	19.851	0.296	>999.999
STF > 13 vs < 1	Overweight	0.758	0.188	3.052	0.639	0.054	7.547
Others vs Caucasian	Obese	0.270	0.083	0.878	----	----	----
Others vs Caucasian	Overweight	0.241	0.104	0.558	----	----	----
Current alcohol user	Obese	0.408	0.147	1.131	----	----	----
Current alcohol user	Overweight	1.078	0.473	2.457	----	----	----
Age	Obese	0.948	0.907	0.991	0.771	0.647	0.920
Age	Overweight	0.964	0.935	0.993	0.876	0.807	0.952
Sleeping hours	Obese	0.414	0.270	0.635	----	----	----
Sleeping hours	Overweight	0.533	0.386	0.736	----	----	----
Body fat percentage	Obese	2.139	1.926	2.376	2.867	1.997	4.115
Body fat percentage	Overweight	1.425	1.332	1.525	1.642	1.358	1.984

Table 3.5.3 Type 3 Analysis of Effects of MVA for ITF under Multinomial Regression Model

		Male (model 3)		Female (model 4)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical ITF	6	15.0684	0.0197	3.3013	0.7702
Ethnicity	2	11.3997	0.0033	----	----
Current alcohol user	2	8.0159	0.0182	----	----
Age	2	----	----	9.6087	0.0082
Sleeping hours	2	17.1653	0.0002	----	----
Body fat percentage	2	201.6541	<0.0001	40.2491	<0.0001

Table 3.5.4 Odds Ratio Estimates of MVA with ITF under Multinomial Regression Model

		Male (model 3)			Female (model 4)		
Effect	**Multilevel Obesity**	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
ITF 1-6 vs < 1	Obese	0.775	0.386	1.555	0.987	0.139	7.007
ITF 1-6 vs < 1	Overweight	0.435	0.266	0.712	0.847	0.251	2.862
ITF 7-13 vs < 1	Obese	0.940	0.182	4.849	<0.01	<0.01	>999
ITF 7-13 vs < 1	Overweight	0.744	0.218	2.541	0.102	0.009	1.218
ITF > 13 vs < 1	Obese	1.182	0.010	133.161	<0.01	<0.01	>999

ITF > 13 vs < 1	Overweight	0.345	0.018	6.582	<0.01	<0.01	>999
Others vs Caucasian	Obese	0.291	0.091	0.933	----	----	----
Others vs Caucasian	Overweight	0.232	0.099	0.542	----	----	----
Current alcohol user	Obese	0.375	0.133	1.061	----	----	----
Current alcohol user	Overweight	1.005	0.429	2.353	----	----	----
Age	Obese	----	----	----	0.828	0.711	0.963
Age	Overweight	----	----	----	0.886	0.815	0.962
Sleeping hours	Obese	0.417	0.270	0.643	----	----	----
Sleeping hours	Overweight	0.536	0.385	0.746	----	----	----
Body fat percentage	Obese	2.079	1.878	2.301	2.813	2.025	3.908
Body fat percentage	Overweight	1.393	1.306	1.485	1.695	1.383	2.077

As conducted for univariate analysis, we checked small or empty cells, as well as the results for goodness of fit (GOF) statistics (Deviance and Pearson test) for these multivariable analysis models. In ITF for females, there were many small or empty cells as reported in **Table 3.3.5**. In addition, the green error warning saying 'there is possibly a quasi-complete separation of data points' was shown in Log Window again when fitting model for ITF among females. The quasi-complete separation indicates our outcome variable (Multilevel Obesity) separates a predictor variable or a combination of predictor variables to certain degree. The goodness of fit (GOF) statistics were not rejected for all multivariable analysis models (p-value > 0.05).

3.5.2 Cumulative Logit Model

When evaluating the effect of categorical STF/ITF on Multilevel Obesity under multivariable analysis for males, the final cumulative logit models contained covariates: binary ethnicity, Current alcohol user, age, sleeping hours, and body fat percentage except for travel frequency. For females, covariates including age, sleeping hours and body fat percentage were contained in the final cumulative logit models for studying both STF and ITF. **Table 3.5.5 - 3.5.8** summarize the results for Type 3 Analysis of Effect and odds ratios calculated from above models. Neither STF nor ITF had a significant effect on Multilevel Obesity (Type 3 P-value > 0.05). In addition, the proportional odds assumption was violated when evaluating the effect of STF/ITF on Multilevel Obesity among males.

Considering the specific effect of STF on Multilevel Obesity for males, as shown in **Table 3.5.6 - model 5**, the odds of a male who traveled 1-6 times per month being in a lower category of Multilevel Obesity were 1.539 times the odds for a male who traveled less than 1 time per month. The odds of a male who traveled 7-13 times per month being in a lower category of Multilevel Obesity were 1.315 times the odds for a male who traveled less than 1 time per month. In contrast, the odds of a male who traveled more than 13 times per month being in a lower category of Multilevel Obesity were 1.725 times the odds for a male who traveled less than 1 time per month. Further, odds of a Caucasian being in a lower category of Multilevel Obesity were 2.206 times the odds for not Caucasian among males. In addition, the odds of a current alcohol user being in a lower category of Multilevel Obesity were 1.743 times the odds of a not current alcohol user among males. Moreover, each one year increased in age for males would increase slightly (1.029 times) the odds of being in a lower category of Multilevel Obesity. One hour increased in sleeping hours for males would 1.626 times the odds of being in a lower category of Multilevel Obesity. In addition, one unit increased in body fat percentage for males would decrease (0.673 times) the odds of being in a lower category of Multilevel Obesity.

The specific effect of summed travel frequency and other covariates on Multilevel Obesity for females (model 2), as well as the effect of international travel frequency and other covariates on Multilevel Obesity for both males and females (model 3-4) under cumulative logit model were described in the same way as above model 5.

In conclusion, males who traveled more than 1 time totally per month were more likely to protect themselves from being obese, compared with those who traveled less than 1 time per month (model 5). Females who traveled 1-6 or more than 13 times in total per month were less likely to be in a lower category of Multilevel Obesity compared with those who traveled less than 1 time per month, while those who traveled 7-13 times in total per month were more likely to be normal weight (model 6). Males who traveled 1-6 or more than 13 times internationally per month were more likely to be in a lower category of Multilevel Obesity compared with those who traveled less than 1 time per month, while those who traveled 7-13 times internationally per month were with less probability to be in a lower category of Multilevel Obesity (model 7). Females who traveled more than 1 time in total per month were more likely to be in a lower category of Multilevel Obesity compared with those who traveled less than 1 time per month (model 8). Here, we still found some odds ratio estimates were not finite when evaluating the effect of ITF for females.

Additionally, other covariates including ethnicity, current alcohol/smoke status, sleeping hours had consistent effects on Multilevel obesity among males and females, and the effects were quite similar with what were found with multinomial regression models, except the effect of sleeping hours for females when studying in model with ITF. Generally, Caucasian who were current alcohol users with more sleeping hours per night and smaller body fat percentage were more likely to be normal weight. Age did not affect the probability of being overweight or obese much. Special case occurred when examining the effect of ITF for females. In this model, a one hour increase in sleeping hours would slightly decrease (0.952 times) the odds of being in a lower category of Multilevel Obesity.

Table 3.5.5 Type 3 Analysis of Effects of MVA for STF under Cumulative Logit Model

		Male (Model 5)		Female (model 6)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical STF	3	1.8084	0.6131	2.4793	0.4790
Ethnicity	1	5.9380	0.0148	----	----
Current alcohol user	1	4.0302	0.0447	----	----
Age	1	5.7430	0.0166	11.8750	0.0006
Sleeping hours	1	18.2245	<.0001	5.3001	0.0213
Body fat percentage	1	247.5373	<.0001	48.8506	<0.0001

Table 3.5.6 Odds Ratio Estimates of MVA for STF under Cumulative Logit Model

	Male (model 5)			Female (model 6)		
Effect	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
categorical STF 1-6 vs < 1	1.539	0.625	3.791	0.647	0.151	2.767
categorical STF 7-13 vs < 1	1.315	0.529	3.266	1.508	0.336	6.776
categorical STF > 13 vs < 1	1.725	0.634	4.694	0.638	0.085	4.807
Not Caucasian vs Caucasian	2.206	1.167	4.170	---	---	---
Current alcohol user	1.743	1.013	3.000	---	---	---
Age	1.029	1.005	1.053	1.141	1.058	1.229
Sleeping hours (hours/night)	1.626	1.301	2.032	1.650	1.077	2.528
Body fat percentage	0.673	0.641	0.707	0.577	0.494	0.673

Table 3.5.7 Type 3 Analysis of Effects of MVA for ITF under Cumulative Logit Model

		Male (Model 7)		Female (model 8)	
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**	**Wald Chi-Square**	**Pr > ChiSq**
categorical ITF	3	0.9742	0.8075	5.2306	0.1557
Ethnicity	1	5.7441	0.0165	----	----
Current alcohol user	1	4.0779	0.0434	----	----
Age	1	5.1408	0.0234	10.8312	0.0010
Sleeping hours	1	17.7046	<0.0001	4.5738	0.0325
Body fat percentage	1	245.1495	<0.0001	47.0371	<0.0001

Table 3.5.8 Odds Ratio Estimates of MVA for ITF under Cumulative Logit Model

	Male (model 7)			Female (model 8)		
Effect	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
categorical ITF 1-6 vs < 1	1.163	0.805	1.680	1.342	0.486	3.704
categorical ITF 7-13 vs < 1	0.853	0.362	2.013	17.564	1.503	205.292
categorical ITF > 13 vs < 1	1.401	0.132	14.889	>999.999	<0.001	>999.999

Not Caucasian vs Caucasian	2.175	1.152	4.107	---	---	---
Current alcohol user	1.746	1.017	3.000	---	---	---
Age	1.027	1.004	1.051	1.133	1.052	1.220
Sleeping hours (hours/night)	1.612	1.291	2.013	0.952	0.910	0.996
Body fat percentage	0.675	0.642	0.709	0.560	0.475	0.661

3.6 Multivariable analysis for BMI and Continuous Travel Frequency

Backward model selection with p=0.05 as criteria was also conducted to decide the final models for multivariable analysis between continuous STF/ITF and continuous BMI for males and females, using linear regression modeling method with squared term for travel frequency. The full models contained additional covariates including Age, body fat percentage, working hours, exercise time, sleeping hours, binary ethnicity, current alcohol user, never smoked. The final linear regression models for males contained additional covariates including current alcohol user, sleeping hours, and body fat percentage. For females, covariates including age, sleeping hours and body fat percentage were contained in the final models. Results for coefficient estimates calculated from above final multivariable analysis (MVA) models are summarized in **Tables 3.6.1 – 3.6.2**. Neither of STF/ITF nor their squared terms had a significant effect on Multilevel Obesity (p-value > 0.05). In addition, the assumptions for these multiple linear regression models were satisfied.

As shown in **Table 3.6.1 – model 1**, when controlling other covariates as constant, STF had a slightly positive effect on BMI for males, and the negative estimate of squared STF means that the effect of STF lessened as travel frequency increased. STF had a slightly negative effect on BMI for females, and the effect of STF lessened as travel frequency increased (model 2). Further, ITF had a slightly negative effect on BMI, and the effect of STF strengthened as travel frequency increased for males and females (model 3-4).

Additionally, other covariates including age, current alcohol/smoke status, sleeping hours had consistent effects on BMI among males and females. Generally, BMI decreased as age/sleeping hours increased or body fat percentage decreased. Further, current alcohol users were with higher BMI compared with non-alcohol users.

Table 3.6.1 Coefficient Estimates of MVA for STF under Linear Regression Model

		Male (model 1)				Female (model 2)			
Parameter	**DF**	**Estimate**	**95% CI**		**Pr > ChiSq**	**Estimate**	**95% CI**		**Pr > ChiSq**
Intercept	1	17.455	15.501	19.410	<0.001	16.586	12.390	20.781	<0.001
STF	1	0.014	-0.093	0.120	0.800	-0.106	-0.296	0.084	0.273
STF^2	1	-0.001	-0.006	0.004	0.742	0.004	-0.005	0.013	0.368
Age	--	----	----	----	----	-0.118	-0.177	-0.060	<0.001
Sleep hours	1	-0.393	-0.621	-0.164	0.001	-0.624	-1.034	-0.213	0.003
Body fat %	1	0.490	0.458	0.521	<0.001	0.541	0.486	0.597	<0.001
Alcohol user	1	-0.612	-1.192	-0.033	0.038	----	----	----	----
Scale	1	2.433	2.304	2.569		2.565	2.283	2.881	

Table 3.6.2 Coefficient Estimates of MVA for ITF under Linear Regression Model

		Male (model 3)				Female (model 4)			
Parameter	**DF**	**Estimate**	**95% CI**		**Pr > ChiSq**	**Estimate**	**95% CI**		**Pr > ChiSq**
Intercept	1	17.681	15.779	19.584	<0.001	15.610	11.510	19.711	<0.001
STF	1	-0.054	-0.208	0.101	0.497	-0.003	-0.416	0.409	0.988
STF^2	1	-0.001	-0.013	0.012	0.890	-0.013	-0.052	0.027	0.535
Age	--	----	----	----	----	-0.112	-0.170	-0.054	<0.001
Sleep hours	1	-0.408	-0.634	-0.181	<0.001	-0.588	-0.996	-0.180	0.005
Body fat %	1	0.489	0.457	0.521	<0.001	0.546	0.491	0.601	<0.001
Alcohol user	1	-0.596	-1.172	-0.019	0.043	----	----	----	----
Scale	1	2.428	2.299	2.564		2.556	2.275	2.871	

4. Discussion

The purpose of this study was to identify the health risks associated with business travel and to assess the specific association between business travel frequency and BMI (continuous, Binary Obesity, Multilevel Obesity). Our hypotheses were that high-frequent business travel would result in excess risks for poor health (e.g. higher BMI, higher body fat percentage, more likely to be obese).

In general, all types of travel frequency were significantly associated with average number of working hours and hours sleeping. Compared with ITF, STF and DTF were additionally associated with some of the covariates related to smoke status, body composition (e.g. fat free mass value). As we expected, STF and DTF were associated with similar covariates, because the average proportion of STF comes from DTF was 83% for overall data and gender-stratified data (**Table 3.1**). To our knowledge, this is the first time that the effects of three types of business travel have been compared. While other studies focused on either domestic travel or international travel[10,11,12], we examined domestic travel, international travel, as well as the total travel (summation of above two types of business travel) and compare the specific effects of them on BMI. Our results are also consistent with a prior study which found international business travel was significantly associated with sleep deprivation.[10]

Focusing on the specific association between travel frequency and BMI in three ways, international travel frequency (ITF) had a significant effect on BMI (continuous or Multilevel Obesity) among males, in the univariate and multivariable analysis per the multinomial logistic regression model and linear regression models. Specifically, we found that moderate business travel seemed to be healthier with most of these travelers being classified as normal weight, while high-frequent business travel had higher BMI. In other words, the BMI displayed a U-shaped pattern of associations with the international travel frequency. These results are consistent with prior study conducted by Dr. Andrew, which found high-frequent business travelers (traveled 14+ nights away from home per month) had significantly higher body mass index, obesity index, compared with those who spent 1-6 nights a month away from home.[11,12] While Dr. Andrew's group examined the association between travel frequency and BMI/Binary Obesity, we also evaluated the association between travel frequency and Multilevel Obesity, and found high-frequent business travelers had significantly higher prevalence of multilevel obesity (either overweight or obese).

4.1 Limitations

There are several limitations associated with this study, particularly from data source and regression models.

4.1.1 Data Source

Our project conducted a retrospective study on 795 executives that were seen at Emory Executive Health between January 2017 and June 2018. Among these executives, 651 (81.9%) were males and 144 (18.1%) females (**Table 3.1.1**). Due to the relatively small sample size for females, the distribution of some covariates (e.g. travel frequency, smoke status, drink status) were more likely to be sparse. Specifically, there was only 1 female who traveled more than 13 times internationally per month, 1 female indicated yes for smoking currently or using Cigars, no females indicated yes for smokeless (vaping) or using a Pipe, and 2 females indicated yes for 'Drank in the past but quit'. Accordingly, modeling difficulties were encountered when studying the risk factors associated with travel frequency, as well as the specific association between travel frequency and BMI for females. Additionally, the average international travel frequency was only around 1.4 times per month for males and females, again resulting in limitations in examining the effect of high-frequent international travel frequency on BMI.

4.1.2 Regression Models

1) Multinomial logistic regression model

In our project, as mentioned in **Results Part 3.3.1 & 3.5.1**, we found that due to sparsity of data, there were some infinite estimated when studying the effect of ITF among females. In addition, when fitting univariate and multivariable models for ITF among females, there were green error warnings showed in **Log Window**, saying 'there is possibly a quasi-complete separation of data points'. As shown in **Table 3.3.5**, there were many cells were empty or with small values when drawing contingency tables for ITF and Multilevel Obesity among females. Such large proportion of empty or small cells might be the reason for quasi-complete separation, thus making maximum likelihood estimations impossible (infinite estimates).

Further, we found the goodness of fit (GOF) statistics were always with 0 degree of freedom and missing statistics values or p-values for univariate analysis models. Considering the degree of freedom $\boldsymbol{df} = (\boldsymbol{N} - \boldsymbol{p})(\boldsymbol{r} - \mathbf{1})$,[21] where r is the number of categories for dependent variable, p is the number of parameters (including intercept), and N is the number of categories for travel frequency, the degree of freedom should be $\boldsymbol{df} = (\mathbf{4} - \mathbf{4})(\mathbf{3} - \mathbf{1}) = \mathbf{0}$ for our univariate models, thus leading to missing GOF statistics values. In other words, our univariate models were saturated models, as a result not able to conduct overall model fit assessment. In addition, ignoring the degree of freedom, there are two requirements for assessing overall fit of models with deviance and Pearson statistics: no more than 20% of the estimated expected counts are smaller than 5, and none of the expected counts are below 1.[21] From previous chi-square tests, we knew that more than 20% of the cells have expected counts smaller than 5 when studying ITF for males and STF/ITF for females, except STF for males. Such large proportion of small expected counts would also affect the statistical results.

2) Cumulative logit model

In our project, the proportional odds assumption was violated when conducting univariate analysis for ITF and multivariable analysis for STF/ITF among males. However, the test of proportional odds assumption has been described as anti-conservative and frequently violated, particularly under following situations:[22,23]

(a) the number of explanatory variables is large;
(b) the sample size is large;
(c) existence of continuous predictors in the model, resulting in very large tables which are often of limited value in evaluating the model because they are so extensive;
(d) special case: small samples with artificially small p-value.

Under multinomial logistic regression model and linear regression model, we found the BMI (Multilevel Obesity) displayed a U-shaped pattern of associations with the international travel frequency, which explains why the proportion odds assumption was not satisfied actually. Therefore, cumulative logit model is not the most appropriate method for studying the specific association between travel frequency and Multilevel Obesity.

3) Linear regression model

According to diagnostic plots from univariate and multivariable models, main assumptions for linear regression models were satisfied. We further studied the possible quadratic association between travel frequency and BMI, and such quadratic linear was less restrictive than simple linear specification. However, it was harder to interpret the results with the squared term of travel frequency.[20]

4) Comparation among three modeling methods

Compared with cumulative logit model as well linear regression model, multinomial logistic regression model would help examine the specific association between business travel frequency and BMI (Multilevel Obesity) better. Specifically, the multinomial logistic regression model had two coefficients for every explanatory variable while the cumulative logit model had only one, making the cumulative logit model much simpler to present and interpret. In addition, the p-value results calculated from multinomial logistic regression models were consistent with the ones from chi-square or fisher's exact test, while results from cumulative logit models were not. Also, the proportional odds assumptions for cumulative logit models were sometimes violated. Additionally, linear regression models with squared travel frequency terms better fit the U shape observed in the association between international travel frequency and BMI among males, but it was hard to interpret the squared term of travel frequency.

4.2 Future Research

Further studies examining the effect of travel frequency on BMI, body composition and lifestyle habits could be improved by collecting data that contains a large sample size for both males and females, high-frequent international travel frequency, as well as younger participants. Further, compared with the cumulative logit model and linear regression model, multinomial logistic regression performed better when studying the specific association between travel frequency and BMI (Multilevel Obesity). In addition, the specific association between business travel and other chronic disease factors, including SBP, DBP (for hypertension), HDL, LDL (for hyperlipidemia) are worthwhile assessing in future researches.

This work has implications for policies at the corporate and governmental levels in order to promote healthy lifestyles for employees who are most at risk for travel-related obesity and for being overweight. Specifically, employees who travel extensively for work should be provided with frequent physical examinations to monitor their health and appropriate occupational health prevention programs would be developed.

4.3 Conclusions

The purpose of this study was to identify the health risks associated with business travel and to assess the specific association between business travel frequency and BMI (continuous BMI, Binary Obesity, Multilevel Obesity). In particular, we found all types of travel frequency were significantly associated with the excess hours working and sleep deprivation. International travel frequency had a significant effect on BMI (continuous or Multilevel Obesity) among males. Specifically, the BMI and prevalence of multilevel obesity displayed a U-shaped pattern of associations with the business travel frequency. This work recommends that future studies examine the specific association between travel frequency and BMI (Multilevel Obesity) using multinomial logistic regression models. In addition, this work will provide a framework for future studies to include associations between travel frequency and other chronic disease factors (e.g. HDL, LDL). Importantly, this works has implications for policies at corporate and governmental levels in order to promote healthy lifestyles for employees who are most at risk for travel-related obesity and for being overweight.

5. References/Bibliography

1. Global Business Travel. https://www.amexglobalbusinesstravel.com/press-room/business-travel-responsible-547-billion-gdp-2016-creates-74-million-jobs/. Published 2016. Accessed 11.16, 2019.
2. 2016 U.S. to overseas for business and convention travel. https://travel.trade.gov/outreachpages/outbound.general_information.outbound_overview.asp. Published 2017. Accessed 1.16, 2020.
3. FERN A. 14 Scientific Effects Traveling Has On Your Mind, Body And Soul. elite daily. https://www.elitedaily.com/life/ways-travel-affects-body/1029885. Published 2015. Updated 5.12. Accessed 3.22, 2020.
4. Jay Keystone, Phyllis Kozarsky, Bradley A. Connor, Hans Nothdurft, Marc Mendelson, Leder K. Travel Medicine E-Book. In: 4th ed.: Elsevier; 2018. Accessed 3.22, 2020.
5. Bunn W. Business travel-associated illness: a geosentinel analysis. *J Travel Med.* 2018;25(1).
6. Druckman M, Harber P, Liu Y, Quigley RL. Assessing the risk of work-related international travel. *J Occup Environ Med.* 2014;56(11):1161-1166.
7. Evans MR, Sarvotham T, Thomas DR, Howard AJ. Domestic and travel-related foodborne gastrointestinal illness in a population health survey. *Epidemiol Infect.* 2006;134(4):686-693.
8. McFee RB. Travel-related illness. *Dis Mon.* 2013;59(12):434-436.
9. Ziegler CC. Travel-related illness. *Crit Care Nurs Clin North Am.* 2013;25(2):333-340.
10. Rogers HL, Reilly SM. Health problems associated with international business travel. A critical review of the literature. *AAOHN J.* 2000;48(8):376-384.
11. Richards CA, Rundle AG. Business travel and self-rated health, obesity, and cardiovascular disease risk factors. *J Occup Environ Med.* 2011;53(4):358-363.
12. Rundle AG, Revenson TA, Friedman M. Business Travel and Behavioral and Mental Health. *J Occup Environ Med.* 2018;60(7):612-616.
13. Carey I. Stress-Related Mental Health Risks Are Rising for Business Travelers. https://skift.com/2019/11/14/stress-related-mental-health-risks-are-rising-for-business-travelers/?utm_source=Newsletter_AH&utm_medium=Thrive. Published 2019. Accessed 1.16, 2020.
14. Liu Y, Nickleach DC, Zhang C, Switchenko JM, Kowalski J. Carrying out streamlined routine data analyses with reports for observational studies: introduction to a series of generic SAS ((R)) macros. *F1000Res.* 2018;7:1955.
15. Allison PD. Logistic Regression Using SAS: Theory and Application. In: Cary, NC: SAS Institute Inc.; 2012.
16. Multinomial logistic regression. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Multinomial_logistic_regression&oldid=936282105. Published 2020. Updated January 17. Accessed March 14, 2020
17. Starkweather J, Moske AK. Multinomial logistic regression. *https://ituntedu/sites/default/files/mlr_jds_aug2011pdf.* 2011;29:2825-2830.
18. MULTINOMIAL LOGISTIC REGRESSION | SAS ANNOTATED OUTPUT. UCLA: Statistical Consulting Group. https://stats.idre.ucla.edu/sas/output/multinomial-logistic-regression/. Accessed 3.4, 2020.
19. Using Statistical Gregression Methods in Education Research: Glossary. National Centre for Research Methods. http://www.restore.ac.uk/srme/www/fac/soc/wie/research-

new/srme/glossary/index52d5.html?selectedLetter=P#proportional-odds-assumption. Updated 7.28, 2011. Accessed 3.4, 2020.

20. Example of Including Nonlinear Components in Regression https://psych.unl.edu/psycrs/statpage/nonlin_eg.pdf. Accessed 3.14, 2020.
21. Lesson 8: Multinomial Logistic Regression Models. In: PennState Statistics: https://dev.stat.vmhost.psu.edu/stat504/book/export/html/788. Accessed 3.4.
22. 5.4 Example 1 - Running an Ordinal Regression on SPSS. National Centre for Research Methods. http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod5/4/index.html. Updated 7.25,2011. Accessed 3.4, 2020.
23. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics.* 1990;46(4):1171-1178.

6. Appendix

6.1 SAS code for Data Organization

```
libname x "h:\ape\data";

**** create all required format;
proc format;
value tfn   1 = '< 1'
          2 = '1-6'
          3 = '7-13'
          4 = '> 13'
           . = 'Missing';
value drink 1 = '0'
            2 = '1-2'
          3 = '3-5'
          4 = '6-10'
          5 = '11-15'
          6 = '16+'
           . = 'Missing';
value drinkm 1 = '0'
            2 = '1-10'
           3 = '11+'
           . = 'Missing';
value drinkf 1 = '0'
            2 = '1-5'
           3 = '6+'
           . = 'Missing';
value working 1 = 'Yes'
             0 = 'No';
value workcat 1 = '[30,35)'
            2 = '[35,40)'
            3 = '[40,45)'
             4 = '[45,50)'
             5 = '[50,55)'
             6 = '[55,60)'
             7 = '[60,65)'
             8 = '[65,70)'
             9 = '>= 70'
             . = 'Missing';
value obcatn  1 = 'Normal'
```

```
        2 = 'Overweight'
        3 = 'Obese'
         . = 'Missing';
value bin    0 = "No"
         1 = "Yes";
run;

**** data cleaning*******;
***1. HHQ lifestyle table;
***1.1 read the data from given file;
proc import out=file DATAFILE="h:\ape\Data\lifestyle.xlsx"
DBMS=xlsx;
run;

***1.2 find and delete the duplicated rows (N change from 1278 to 1182);
proc sort data=file NODUP;
by _all_;
run;

***1.3 create variables to reformat the travel frequency;
data reformat;
set file;
DTF=Domestic_Travel__Avg___of_days_p;
ITF=VAR7;
TTF=Total_Traveled;
label DTF="Domestic Travel Frequency (per month)"
         ITF="Int'l Travel Frequency (per month)"
         TTF="Total Travel Frequency (per month)"
   ;

        * 1)DTF;
        * equivalent to manually reformat;
        if MRN__=51949895 then DTF="12/year"; *special case with specific mrn;
        if DTF="Every three months" then DTF=0.33; ***;
        else if DTF="3 to 4 days a week" then DTF = "12-16";
        else if DTF="60-70 per year" then DTF = "5-6";
        else if DTF="every other month" then DTF=0.5; ***;
        else if DTF="one to twice monthly" then DTF= "1-2";  ***;
        else if DTF="Two" then DTF=2;
        else if DTF="one wk" then DTF=7;
        else if DTF="One" then DTF=1;
  else if (DTF in ("N/A","Non","n/a","NA","None","none","O",'minimal','Rare','rarely')) then DTF = 0;
        else if (DTF in ('a','yes','??',' ')) then DTF = '.';

        ** add some special case;
        else if DTF="15/30" then DTF=15;
        else if DTF="1/2 per month" then DTF="1/2"; ***;
        else if (index(DTF,"months") ne 0) and (compress(DTF,'/','A') ne ' ')
        then DTF = compress(DTF,'/','A')*30/12; /* here is a 10 months, according to following steps, we
see it as 10 months per year*/

        * programming reformat;
        else if (index(DTF,"month") ne 0 or index(DTF,"mos") ne 0) and (compress(DTF,'/','A') ne ' ')
        then DTF = compress(DTF,'/','A');
        else if (index(DTF,"year") ne 0 or index(DTF,"vacation") ne 0) and (compress(DTF,'/','A') ne ' ')
        then DTF = compress(DTF,'/','A')/12;
```

```
        else if (index(DTF,"week") ne 0) and (compress(DTF,'/','A') ne ' ')
        then DTF = compress(DTF,'/','A')*4;
        else if (index(DTF,"day") ne 0 or index(DTF,"Day") ne 0 or index(DTF,"night") ne 0) and
(compress(DTF,'/','A') ne ' ')
        then DTF = compress(DTF,'/','A');
        else if (index(DTF,"hour") ne 0)
        then DTF = 1;
        DTF = translate(DTF,'-','to');
        DTF = translate(DTF,'-','=');
        DTF = compress(DTF);

        * Define upper and lower bounds;
        DTFint = DTF;
        if substr(DTFint,notdigit(DTFint),1) in ('/','-') then do;
                DTF_lower = scan(compress(DTFint,'','kdp'),1);
                DTF_upper = scan(compress(DTFint,'','kdp'),2);
        end;
        else do;
                DTF_lower = DTFint;
                DTF_upper = DTF_lower;
        end;
        DTF_upper = DTF_upper * 1;
        DTF_lower = DTF_lower * 1;
        DTF_mid = (DTF_upper + DTF_lower)/2;

        * Categorize according to the midpoints;
        if '.' ne DTF_mid < 1 then DTFcat2 = 1;
        else if 1 le DTF_mid < 6.5 then DTFcat2 = 2;
        else if 6.5 le DTF_mid < 13.5 then DTFcat2 = 3;
        else if 13.5 le DTF_mid le 30 and DTF_upper le 30 then DTFcat2 = 4; *(N = 1178);


    * 2)ITF;
        * equivalent to manually reformat;
        if MRN__=13344213 then ITF=3; *special case for specific mrn;
        if ITF="0 (1-3 weeks/year)" then ITF="0-2";
        else if ITF="2 - 3 weeks per year" then ITF = "1-2";
        else if ITF="1-2 weeks per year" then ITF="0-1";
        else if ITF="2-4 weeks per year" then ITF = "1-2";
        else if ITF="3 days - once a year" then ITF="0-1";   ***;
        else if ITF="2 to 3 x per year" then ITF="0-1"; ***;
        else if ITF="3 to 4 times a year" then ITF="0-1";
        else if ITF="6-8 trips year" then ITF="0-1"; ***;
        else if ITF="3-5 trips/year" then ITF="0-1"; ***;
        else if ITF="5-7 per year" then ITF="0-1";
        else if ITF="3-4 weeks per year" then ITF= "1-3";
        else if ITF="8-12 per year" then ITF="0-1";
        else if ITF="flights bi-weekly" then ITF=4;
        else if ITF="one trip per quarter" then ITF=0.25;  ***;
        else if ITF="one" then ITF=1;
        else if ITF="Yearly 10 days" then ITF=0.83;
        else if (ITF in ("Once a year","Annual","once a year","usually yearly","vacation only","1/yr,"))
then ITF = 0.08;
        else if (ITF in ("twice a year","twice per year","Twice a year")) then ITF = 0.17;
        else if (ITF in("once or twice a year","1-2 times per year")) then ITF="0-1";
```

```
   else if (ITF in
("N/A","NA","n/a","0/month","None","none","O","o","0/0","00/00","Minimal","Mimimal","miminal","rar
e","rarely","<1","minimal","Seldom now","infrequent","very little","Occasional","occaisional","couple
times a year","Occassionally"))
            then ITF = 0;
            else if (ITF in ("varies","depends on month")) then ITF=0.5;
            else if (ITF in ('a','yes','??',' ')) then ITF='.';
            else if ITF="2 (Canada)" then ITF=2;
   else if ITF="2/30" then ITF=2;

            * programming reformat;
            else if (index(ITF,"months") ne 0) and (compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A')*30/12;
            else if (index(ITF,"month") ne 0 or index(ITF,"mos") ne 0) and (compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A');
            else if (index(ITF,"year") ne 0 and index(ITF,"week") ne 0) and (compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A')/12*7;
            else if (index(ITF,"year") ne 0 or index(ITF,"vacation") ne 0 or index(ITF,"yr") ne 0 ) and
(compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A')/12; ***;
            else if (index(ITF,"week") ne 0) and (compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A')*4;
            else if (index(ITF,"day") ne 0 or index(ITF,"Day") ne 0 or index(ITF,"night") ne 0) and
(compress(ITF,'/','A') ne ' ')
            then ITF = compress(ITF,'/','A');
            else if (index(ITF,"hour") ne 0)
            then ITF = 1;
            ITF = translate(ITF,'-','to');
            *ITF = translate(ITF,'-','=');
            ITF = compress(ITF);

            * Define upper and lower bounds;
            ITFint = ITF;
            if substr(ITFint,notdigit(ITFint),1) in ('/','-') then do;
                        ITF_lower = scan(compress(ITFint,'','kdp'),1);
                        ITF_upper = scan(compress(ITFint,'','kdp'),2);
            end;
            else do;
                        ITF_lower = ITFint;
               ITF_upper = ITF_lower;
            end;
            ITF_upper = ITF_upper * 1;
            ITF_lower = ITF_lower * 1;
            ITF_mid = (ITF_upper + ITF_lower)/2;

            * Categorize according to the midpoints;
            if '.' ne ITF_mid < 1 then ITFcat2 = 1;
            else if 1 le ITF_mid < 6.5 then ITFcat2 = 2;
            else if 6.5 le ITF_mid < 13.5 then ITFcat2 = 3;
            else if 13.5 le ITF_mid le 30 and ITF_upper le 30 then ITFcat2 = 4;


            * 3)TTF;
            * equivalent to manually reformat;
            if TTF="10-15/year" then TTF="0-1";
            else if TTF="maybe 3-4 weeks/year" then TTF = "2-3";
```

```
        else if TTF="96-100/12 months" then TTF = "8-9";
        else if TTF="75-80 per year" then TTF = "6-7";
        else if TTF="every other month" then TTF=0.5; ***;
        else if TTF="one to twice monthly" then TTF= "1-2"; ***;
        else if TTF="Two/month" then TTF=2;
        else if TTF="four" then TTF=4;
        else if TTF="One" then TTF=1;
  else if (TTF in ("N/A","Non","n/a","NA","None","Nonr","none","O",'minimal','Rare','rarely')) then TTF
= 0;
        else if (TTF in ('depends','varies')) then TTF = 0.5; *just let them will be in '<1' cat;
        else if (TTF in ('a','yes','??',' ','Don?t know','Not sure','alot','frequent')) then TTF='.';
        else if TTF="2016 - 50% Travel" then TTF = 15;

        ** add some special case;
        else if TTF="17/30" then TTF=17;
        else if TTF="5/12 ?" then TTF='5/12';
        else if TTF="15/30" then TTF=15;
        else if TTF="1/2 day a month" then TTF="1/2"; ***;
        else if TTF="16 a months" then TTF=16; /*otherwise will be 40 during the next steps*/

        * programming reformat;
        else if (index(TTF,"months") ne 0) and (compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/.','A')*30/12;
        else if (index(TTF,"month") ne 0 or index(TTF,"mos") ne 0) and (compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/.','A');
        else if (index(TTF,"year") ne 0 and index(TTF,"week") ne 0) and (compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/','A')/12*7;
        else if (index(TTF,"year") ne 0 or index(TTF,"vacation") ne 0 or index(TTF,"yr") ne 0 ) and
(compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/','A')/12;
        else if (index(TTF,"week") ne 0) and (compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/','A')*4;
        else if (index(TTF,"day") ne 0 or index(TTF,"Day") ne 0 or index(TTF,"night") ne 0) and
(compress(TTF,'/','A') ne ' ')
        then TTF = compress(TTF,'/','A');
        else if (index(TTF,"hour") ne 0)
        then TTF = 1;
        TTF = translate(TTF,'-','to');
        *TTF = translate(TTF,'-','=');
        TTF = compress(TTF);

        * Define upper and lower bounds;
        TTFint = TTF;
        if substr(TTFint,notdigit(TTFint),1) in('-','/') then do;
                TTF_lower = scan(compress(TTFint,'','kdp'),1);
                TTF_upper = scan(compress(TTFint,'','kdp'),2);
        end;
        else do;
                TTF_lower = TTFint;
                TTF_upper = TTF_lower;
        end;
        TTF_upper = TTF_upper * 1;
        TTF_lower = TTF_lower * 1;
        TTF_mid = (TTF_upper + TTF_lower)/2;

        * Categorize according to the midpoints;
```

```
    if '.' ne TTF_mid < 1 then TTFcat2 = 1;
    else if 1 le TTF_mid < 6.5 then TTFcat2 = 2;
    else if 6.5 le TTF_mid < 13.5 then TTFcat2 = 3;
    else if 13.5 le TTF_mid le 30 and TTF_upper le 30 then TTFcat2 = 4;

    format TTFcat2 tfn.
             DTFcat2 tfn.
   ITFcat2 tfn.;

drop DTFint DTF_upper DTF_lower ITFint ITF_upper ITF_lower TTFint TTF_upper TTF_lower;
run;

***1.4 sum up domestic and international travel frequency and compare with original data;
data reformat_sum;
set reformat;
STF_mid = DTF_mid + ITF_mid;
    if '.' ne STF_mid < 1 then STFcat2 = 1;
    else if 1 le STF_mid < 6.5 then STFcat2 = 2;
    else if 6.5 le STF_mid < 13.5 then STFcat2 = 3;
    else if 13.5 le STF_mid le 30 then STFcat2 = 4;

    if STF_mid > 30 then do;  *add the new rule;
        DTFcat2=.;
        ITFcat2=.;
        STFcat2=.;
    end;

format TTFcat2 tfn.
     DTFcat2 tfn.
     STFcat2 tfn.;

if DTFcat2 =. then DTF_mid = .;
if ITFcat2 =. then ITF_mid = .;
if TTFcat2 =. then TTF_mid = .;
if STFcat2 =. then STF_mid = .;

PDIS = round(DTF_mid/STF_mid,0.001);
* compare new variable with original ttf (match=0 - unmatched, match=1 - matched);
match=0;
if STFcat2 = TTFcat2 then match = 1;

label   PDIS = 'Proportion of DTF in STF'
        DTFcat2="categorical DTF (days/month)"
     ITFcat2="categorical ITF (days/month)"
     STFcat2="categorical STF (days/month)"
     DTF_mid="continuous DTF (days/month)"
     ITF_mid="continuous ITF (days/month)"
     STF_mid="continuous STF (days/month)"
;
run;

***1.5 reformating working and sleeping entries;
data reformat_wns;
set reformat_sum;
sleep = Sleep__Avg___of_hrs_per_night;
work = Work__Avg___of_hrs_per_week;
```

```
worker = 1;

** 1)sleep;
sleep = tranwrd(sleep,"+/-","");
sleep = tranwrd(sleep,"hrs","");
sleep = compress(sleep);
if substr(sleep,notdigit(sleep),1) in('-','/',' -') then do;
                sleep_mid = (scan(compress(sleep,'','kdp'),1)*1 + scan(compress(sleep,'','kdp'),2)*1) / 2;
        end;
        else do;
        sleep_mid = sleep*1;
        end;
if sleep_mid le 1 then sleep_mid=.;

** 2)working;
*  equivalent to manually reformat;
        if work = "20 - volunteer" then work = 20;
        else if work = "60 (w/commute)" then work = 60;
        else if work = "Fifty" then work = 50;
        else if work = "maybe 6" then work = 6;
        else if work = "now 3" then work = 3;
        else if work = "5 - retired" then work = 5;
        else if work = "Retired, but 50-70" then work = '50-70';
        else if work = "24/7" then work = 24;
        else if work in ('O','0/week','None','NA','n/a','none') then work = 0;

*  programming reformat;
        * 2-1) deal with those view as missing;
        else if work in ('work at home','varies','a') then do
                work = '';*(n=4);
                worker = 1;
        end;
        else if (index(work,"mom") ne 0 or index(work,"wife") ne 0 or index(work,'home') ne 0) then
work = 0;
   else if (index(work,"tired") ne 0 or index(work,"RETIRED") ne 0 or index(work,"Unemployed") ne 0)
then work = 0;

        * 2-2) with hrs, week, plus ... characters;
        else if (index(work,"week") ne 0 or index(work,"hour") ne 0 or index(work,"hrs") ne 0 or
index(work,"wk") ne 0 or index(work,"plus") ne 0)
        then work = compress(work,'/.','A');

        work = tranwrd(work,'+','');
        work = tranwrd(work,'to','-');
        work = compress(work);

* create new variable - work_mid;
        if substr(work,notdigit(work),1) = '-' then do;
                work_mid = (scan(compress(work,'','kdp'),1)*1 + scan(compress(work,'','kdp'),2)*1) / 2;
        end;
        else do;
                work_mid = work*1;
        end;

* mark non-workers;
   if 0 le work_mid < 30 or index(Work__Avg___of_hrs_per_week,"tired") ne 0 then worker = 0;
```

```
* view entires bigger than 100 as missing;
        if work_mid > 100 then work_mid =.;

* Categorize according to the midpoints;
        *if '.' ne work_mid < 25 then workcat = 1;
        *else if  25 le work_mid < 30 then workcat = 2;
        else if  30 le work_mid < 35 then workcat = 1;
        else if  35 le work_mid < 40 then workcat = 2;
        else if  40 le work_mid < 45 then workcat = 3;
        else if  45 le work_mid < 50 then workcat = 4;
        else if  50 le work_mid < 55 then workcat = 5;
        else if  55 le work_mid < 60 then workcat = 6;
        else if  60 le work_mid < 65 then workcat = 7;
        else if  65 le work_mid < 70 then workcat = 8;
        else if  70 le work_mid then workcat = 9;
        else workcat = .;

label worker = "Current worker or not"
         sleep_mid = "sleeping hour (hrs/night)"
         work_mid = "continuous work hour (hrs/week)"
         workcat = "categorical work hour (hrs/week)"
    ;
format worker working.
    workcat workcat.;
drop match;
run;

***1.6 further clean drink;
data reformat_drink;
set reformat_wns;
drink = .;
if Never_Drank=1 then drink=1;
if Avg___of_drinks_per_week="0-2" & Never_Drank=0 then drink=2;
if Avg___of_drinks_per_week="3-5" then drink=3;
if Avg___of_drinks_per_week="6-10" then drink=4;
if Avg___of_drinks_per_week="11-15" then drink=5;
if Avg___of_drinks_per_week in ('16-20', '21-25', '26+') then drink=6;
format drink drink.;
label drink='average number of drinks per week'
         current_drinker="current alcohol user";
run;

***1.7 output the cleaned datadaet-lifestyle;
data x.lifestyle;
set reformat_drink;
run;


***2. HHQ exercise table;
***2.1 read the data from given file;
proc import out=fexercise DATAFILE="h:\ape\data\exercise.xlsx"
DBMS=xlsx;
run;

***2.2 rename variables;
```

```
data fexercise (rename=(MRN__ = MRN Exercise_and_Recreation__Exercis = Exercise_Code
HHQ_Executive_Health__Last_Modif=DOS));
        set fexercise;
run;

***2.3 delete the duplicated activities;
proc sort data=fexercise out=exercise nodupkey;
by mrn activity minutes_per_session times_per_week dos;
run;

***2.4 keep all variables and records for data checking;
data file1;
        set exercise;
        by MRN;
        session=minutes_per_session;
        times=times_per_week;
        if session='3019' then session='.';
        if times ge 21 then times='.';
        minutes_per_week = session * times;
        if MRN ne . then do;
                if (first.MRN and last.MRN) then total_per_week=minutes_per_week;
                else do;
                        if (first.MRN) then total_per_week = 0;
                        total_per_week + minutes_per_week;
                end;
                if (last.MRN) then output;
        end;
        else do;
        total_per_week = minutes_per_week;
        output;
        end;
        drop minutes_per_week times_per_week minutes_per_session activity exercise_code session
times;
run;


***2.5 output the dataset for fuuther merging step;
data x.exercise;
set file1;
        exercise=round(total_per_week/60,0.01);
  if exercise > 25 then exercise=.;
        label exercise='exercise (hours/week)';
run;


*** data merging part****************************************;
***1.read tables;
***1.1 read SECA table;
data seca;
        set x.seca(rename=(date_of_birth=dob));
        drop date_of_service; ****;
        label dos='date_of_service';
run;

***1.2 read reference table;
data dobmrn;
```

```
        set x.dobmrn(rename=(patient_birth_date=dob arrival_day=dos)); *mrn=original_mrn));
        label dob='date_of_birth' dos = 'date_of_service';
run;

***1.3 read lifestyle table;
data lifestyle;
        set x.lifestyle(rename=(MRN__ = MRN));
run;

* sort before merge;
proc sort data=lifestyle;
        by MRN;
run;
proc sort data=dobmrn;
        by MRN;
run;

*** 2. merge lifestyle table with reference table by mrn;
data lifestyleDOB;
        merge lifestyle(in=a) dobmrn(in=b);
        by MRN;
        if a and b;
run;

data lifestyleDOB;
        set lifestyleDOB(rename=(mrn=original_mrn));
run;

* sort before merge;
proc sort data=lifestyleDOB;
        by dob dos;
run;
proc sort data=seca;
        by dob dos;
run;

***3. merge SECA with the already merged lifestyle and reference tables and check whether MRN
matches;
data secalifestyle;
        merge lifestyleDOB(in=a) seca(in=b);
        by dob dos;
        if a and b;
        match = 0;
        if mrn=original_mrn then match=1;
run;

***4. data cleaning for SECA-Lifestyle dataset;
***4.1 calculate the average of records;
proc sql;
create table want as
select *, mean(BMI_value) as avg_BMI,
      mean(Absolute_fat_mass_value) as avg_Absolute_fat_mass_value,
      mean(Fat_free_mass_value) as avg_Fat_free_mass_value,
      mean(Skeletal_muscle_mass_value) as avg_Skeletal_muscle_mass_value,
      mean(SMM__torso__value) as avg_SMM__torso__value,
      mean(SMM__RL__value) as avg_SMM__RL__value,
```

```
            mean(SMM__LL__value) as avg_SMM__LL__value,
            mean(SMM__LA__value) as avg_SMM__LA__value,
            mean(SMM__RA__value) as avg_SMM__RA__value,
            mean(waist_circumference_value) as avg_waist_circumference_value,
            mean(Weight_value) as avg_Weight_value,
            mean(Height_value) as avg_Height_value,
            mean(Resting_energy_expenditure_value) as avg_Resting_energy_expenditure,
            mean(visceral_adipose_tissue_value) as avgvisceral_adipose_tissue_value
from secalifestyle
group by original_mrn;
quit;

***4.2 delete duplicated MRNs,create variables;
data secalifestyle1;
length ethnic $ 18;
set want;
by original_mrn;
if ethnic="African" then ethnic="African American";
age=floor((dos-dob)/365.25); * capture the age according to DOB;
body_fat=round((avg_Absolute_fat_mass_value/avg_Weight_value)*100,.01);
if (first.original_mrn);

obesity = 0;
if avg_bmi >= 30 then obesity =1;
if '.' ne avg_BMI < 25 then obesity3 = 1;
else if 25 <= avg_bmi < 30 then obesity3 = 2;
else if 30 <= avg_bmi then obesity3 = 3;
else obesity3 = .;

format obesity3 obcatn. obesity bin. Current_Drinker bin. Drank_in_the_past_but_quit bin.
          Never_Drank bin. Current_Smoker bin. Never_Smoked bin. Smokeless bin. Pipe bin. Cigar bin.;

label obesity = 'Binary Obesity'
           obesity3 = 'Multilevel Obesity'
    body_fat="body fat percentage (%)"
;

rename  avg_Absolute_fat_mass_value = Absolute_fat_mas
               avg_Fat_free_mass_value = Fat_free_mass
               avg_Skeletal_muscle_mass_value = Skeletal_muscle_m
               avg_waist_circumference_value = Waist_circum
               avgvisceral_adipose_tissue_value = visceral_adipose
               avg_Resting_energy_expenditure = Resting_energy_exp;

drop work sleep VAR7 ITF DTF Domestic_Travel__Avg___of_days_p TTF Total_Traveled
Work__Avg___of_hrs_per_week
Sleep__Avg___of_hrs_per_night BMI_value Weight_value Height_value Absolute_fat_mass_value
Fat_free_mass_value
Skeletal_muscle_mass_value waist_circumference_value visceral_adipose_tissue_value
Resting_energy_expenditure_value
SMM__LA__value SMM__LL__value SMM__RA__value SMM__RL__value SMM__torso__value
match Avg___of_drinks_per_week;
run;

*** 5. label the important variables;
data secalifestyle1;
```

```
set secalifestyle1;
label Absolute_fat_mas = 'Absolute fat mass value'
         Fat_free_mass = 'Fat free mass value'
         Skeletal_muscle_m = 'Skeletal muscle mass value'
         Waist_circum = 'waist circumference value'
         visceral_adipose = 'visceral adipose tissue value'
   Resting_energy_exp = 'Resting energy expenditure'
         avg_SMM__LA__value='SMM LA value'
   avg_SMM__LL__value='SMM LL value'
   avg_SMM__RA__value='SMM RA value'
   avg_SMM__RL__value='SMM RL value'
   avg_SMM__torso__value='SMM torso value'
         ethnic="ethnicity"
   avg_bmi="BMI"
         avg_Weight_value='weight value'
         avg_Height_value='height value'
   ;
run;

*** 6. merge with exercise table(with total exercise time);
data exercise;
set x.exercise (rename=(mrn=original_mrn));
run;
* first sort data;
proc sort data=exercise;
by original_mrn;
run;

data manu_exea;
merge secalifestyle1(in=a) exercise(in=b);
by original_mrn;
if a;
if worker=1;
run;

*** 7. output the final dataset;
data x.final;
set manu_exea;
run;
```

6.2 SAS code for Data Analysis

```
libname x "h:\ape\data";
%include "h:\ape\code\des_sta.sas";
%include "h:\ape\code\sum_sta.sas";
%include "h:\ape\code\UNI_NUMEXP.sas";
*define a directory where the output tables will be saved;
%let dir = h:\ape\output\;
*import format;
proc format;
value tfn   1 = '< 1'
        2 = '1-6'
        3 = '7-13'
        4 = '> 13'
```

```
             . = 'Missing';
value drink 1 = '0'
             2 = '1-2'
           3 = '3-5'
           4 = '6-10'
           5 = '11-15'
           6 = '16+'
             . = 'Missing';
value drinkm 1 = '0'
             2 = '1-10'
            3 = '11+'
             . = 'Missing';
value drinkf 1 = '0'
             2 = '1-5'
            3 = '6+'
             . = 'Missing';
value working 1 = 'Yes'
               0 = 'No';
value workcat 1 = '[30,35)'
             2 = '[35,40)'
             3 = '[40,45)'
              4 = '[45,50)'
              5 = '[50,55)'
              6 = '[55,60)'
              7 = '[60,65)'
              8 = '[65,70)'
              9 = '>= 70'
               . = 'Missing';
value obcatn  1 = 'Normal'
             2 = 'Overweight'
             3 = 'Obese'
              . = 'Missing';
value bin     0 = "No"
               1 = "Yes";
run;

** import data;
data final;
set x.final;
ethnic2=ethnic;
if ethnic ne "Caucasian" then ethnic2="Not Caucasian";
STF_mid2=round(STF_mid*STF_mid,.01);
ITF_mid2=round(ITF_mid*ITF_mid,.01);
DTF_mid2=round(DTF_mid*DTF_mid,.01);
run;

data final_male final_female;
set final;
if Gender='Male' then output final_male;
else output final_female;
run;

data final_male;
set final_male;
if drink=3 or drink=4 then drink=2;
if drink=6 or drink=5 then drink=3;
```

```
format drink drinkm.;
run;

data final_female;
set final_female;
if drink=3 then drink=2;
if drink=6 or drink=5 or drink=4 then drink=3;
format drink drinkf.;
run;


** 1. descriptive statistics;
%let c_var = ethnic obesity obesity3 STFca2 DTFcat2 ITFcat2 workcat
Current_Drinker Never_Drank Drank_in_the_past_but_quit drink
Current_Smoker Never_Smoked Smokeless Pipe Cigar;
%let n_var = Age avg_Height_value avg_Weight_value avg_BMI Absolute_fat_mas Fat_free_mass
body_fat   Skeletal_muscle_m waist_circum visceral_adipose Resting_energy_exp
avg_SMM__LA__value avg_SMM__LL__value avg_SMM__RA__value avg_SMM__RL__value
avg_SMM__torso__value
STF_mid DTF_mid ITF_mid PDIS work_mid exerise sleep_mid ;

%DESCRIPTIVE(DATASET=work.final,
    CLIST= gender &c_var,
    NLIST = &n_var,
    OUTPATH=&dir,
    FNAME=Table 1 Overall,
            DEBUG = F);
%DESCRIPTIVE(DATASET=final_male,
    CLIST= &c_var,
    NLIST = &num_var,
    OUTPATH=&dir,
    FNAME=Table 1 Male,
            DEBUG = F);
%DESCRIPTIVE(DATASET=final_female,
    CLIST= &c_var,
    NLIST = &n_var,
    OUTPATH=&dir,
    FNAME=Table 1 Female,
            DEBUG = F);


*** 2. univariate associations with categorical/continuous TF and BMI (continuous, binary obesity and
multilevel obesiy);
%let cat_var = ethnic obesity obesity3 workcat
Current_Drinker Never_Drank Drank_in_the_past_but_quit drink
Current_Smoker Never_Smoked Smokeless Pipe Cigar;
%let num_var = Age avg_Height_value avg_Weight_value avg_BMI Absolute_fat_mas Fat_free_mass
body_fat
Skeletal_muscle_m waist_circum visceral_adipose Resting_energy_exp
avg_SMM__LA__value avg_SMM__LL__value avg_SMM__RA__value avg_SMM__RL__value
avg_SMM__torso__value
work_mid exerise sleep_mid;
%let cat_var2 = ethnic obesity obesity3 workcat
Current_Drinker Never_Drank Drank_in_the_past_but_quit drink
Current_Smoker Never_Smoked Cigar;
```

```
** 2.1 categorical travel frequency ;
** overall;
%UNI_CAT(dataset = work.final,
            outcome = STFcat2,
            CLIST = gender &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by STF);

%UNI_CAT(dataset = work.final,
            outcome = DTFcat2,
            CLIST = gender &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by DTF);

%UNI_CAT(dataset = work.final,
            outcome = ITFcat2,
            CLIST = gender &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by ITF);

** male;
%UNI_CAT(dataset = final_male,
            outcome = STFcat2,
            CLIST = &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by STF male);

%UNI_CAT(dataset = final_male,
            outcome = DTFcat2,
            CLIST = &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
```

```
            fname = cat_univariate by DTF male);

%UNI_CAT(dataset = final_male,
            outcome = ITFcat2,
            CLIST = &cat_var,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by ITF male);

** female;
%UNI_CAT(dataset = final_female,
            outcome = STFcat2,
            CLIST = &cat_var2,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by STF female);

%UNI_CAT(dataset = final_female,
            outcome = DTFcat2,
            CLIST = &cat_var2,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by DTF female);

%UNI_CAT(dataset = final_female,
            outcome = ITFcat2,
            CLIST = &cat_var2,
   NLIST =  &num_var,
   nonpar = F,
            rowpercent = F,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = cat_univariate by ITF female);


** 2.2 continuous travel frequency;
** overll;
%UNI_NUMEXP(dataset= final,
                        outcome= STF_mid,
                clist= gender &cat_var,
                        nlist= &num_var,
                nonpar = T,
                outpath= &dir,
```

```
                fname=con_univariate by STF,
        debug=F);

%UNI_NUMEXP(dataset= final,
                outcome= DTF_mid,
          clist= gender &cat_var,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by DTF,
        debug=F);

%UNI_NUMEXP(dataset= final,
                outcome= ITF_mid,
          clist= gender &cat_var,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by ITF,
        debug=F);

** male;
%UNI_NUMEXP(dataset= final_male,
                outcome= STF_mid,
          clist= gender &cat_var,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by STF male,
        debug=F);

%UNI_NUMEXP(dataset= final_male,
                outcome= STF_mid,
          clist= gender &cat_var,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by DTF male,
        debug=F);

%UNI_NUMEXP(dataset= final_male,
                outcome= ITF_mid,
          clist= gender &cat_var,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by ITF male,
        debug=F);

** female;
%UNI_NUMEXP(dataset= final_female,
                outcome= STF_mid,
          clist= gender &cat_var2,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
```

```
                fname=con_univariate by STF female,
        debug=F);

%UNI_NUMEXP(dataset= final_female,
                outcome= STF_mid,
          clist= gender &cat_var2,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by DTF female,
        debug=F);

%UNI_NUMEXP(dataset= final_female,
                outcome= ITF_mid,
          clist= gender &cat_var2,
                nlist= &num_var,
          nonpar = T,
          outpath= &dir,
                fname=con_univariate by ITF female,
        debug=F);


***2.3 outcome is BMI/binary obesity/multilevel obesity;
%let c_var = ethnic workcat STFcat2 DTFcat2 ITFcat2
Current_Drinker Drank_in_the_past_but_quit Never_Drank drink
Current_Smoker Never_Smoked Smokeless Pipe Cigar;

%let n_var = Age STF_mid DTF_mid ITF_mid sleep_mid work_mid exercise body_fat avg_Weight_value
avg_Height_value
Absolute_fat_mas Fat_free_mass Skeletal_muscle_m waist_circum visceral_adipose Resting_energy_exp
avg_SMM__LA__value avg_SMM__LL__value avg_SMM__RA__value avg_SMM__RL__value
avg_SMM__torso__value;

%let c_var2 = ethnic workcat STFcat2 DTFcat2 ITFcat2
Current_Drinker Drank_in_the_past_but_quit Never_Drank drink
Current_Smoker Never_Smoked Cigar;

** 2.3.1 BMI;
%UNI_NUMEXP(dataset= final,
                outcome= avg_bmi,
          clist= gender &c_var,
                nlist= &n_var,
          nonpar = T,
          outpath= &dir,
                fname=univariate bmi,
        debug=F);
%UNI_NUMEXP(dataset= final_male,
                outcome= avg_bmi,
          clist= &c_var,
                nlist= &n_var,
          nonpar = T,
          outpath= &dir,
                fname=univariate bmi Male,
        debug=F);
%UNI_NUMEXP(dataset= final_female,
                outcome= avg_bmi,
```

```
                clist= &c_var2,
                        nlist= &n_var,
                nonpar = T,
                outpath= &dir,
                        fname=univariate bmi Female,
             debug=F);

** 2.3.2 obesity;
%UNI_CAT(dataset = final,
            outcome = obesity,
            CLIST = gender &c_var,
   NLIST =  &n_var,
   nonpar = F,
            rowpercent = T,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = univariate obesity);

%UNI_CAT(dataset = final_male,
            outcome = obesity,
            CLIST = &c_var,
   NLIST =  &n_var,
   nonpar = F,
            rowpercent = T,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = univariate obesity Male);

%UNI_CAT(dataset = final_female,
            outcome = obesity,
            CLIST = &c_var2,
   NLIST =  &n_var,
   nonpar = F,
            rowpercent = T,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = univariate obesity Female);


** 2.3.3 obesity3;
%UNI_CAT(dataset = final,
            outcome = obesity3,
            CLIST = gender &c_var,
   NLIST =  &n_var,
   nonpar = F,
            rowpercent = T,
            spread = T,
            orientation = portrait,
            outpath = &dir,
            fname = univariate multiobesity);

%UNI_CAT(dataset = final_male,
            outcome = obesity3,
```

```
          CLIST = &c_var,
  NLIST =  &n_var,
  nonpar = F,
          rowpercent = T,
          spread = T,
          orientation = portrait,
          outpath = &dir,
          fname = univariate multiobesity Male);

%UNI_CAT(dataset = final_female,
          outcome = obesity3,
          CLIST = &c_var2,
  NLIST =  &n_var,
  nonpar = F,
          rowpercent = T,
          spread = T,
          orientation = portrait,
          outpath = &dir,
          fname = univariate multiobesity Female);


*** 3. univariate analysis between multilevel besity and categorical TF;
** 3.1 with multinomial regression model;
ods rtf file='h:\ape\output\uni-multinominal.doc';
title 'STF-male';
proc logistic data= final_male;
class obesity3(ref='Normal') STFcat2(ref= '< 1')/ param=ref;
model obesity3 = STFcat2/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
title 'ITF-male';
proc logistic data= final_male;
class obesity3(ref='Normal') ITFcat2(ref= '< 1')/ param=ref;
model obesity3 = ITFcat2/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
title 'STF-female';
proc logistic data= final_female;
class obesity3(ref='Normal') STFcat2(ref= '< 1')/ param=ref;
model obesity3 = STFcat2/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
title 'ITF-female';
proc logistic data= final_female;
class obesity3(ref='Normal') ITFcat2(ref= '< 1')/ param=ref;
model obesity3 = ITFcat2/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
ods rtf close;

** 3.2 cumulative logits model;
** 3.2.1 first prepare the dataset;
data maleclm ;
set final_male;
format obesity3;
informat obesity3;
run;
data femaleclm ;
set final_female;
format obesity3;
```

```
informat obesity3;
run;

** 3.2.2 fit models;
ods rtf file='h:\ape\output\uni-ordinal.doc';
title 'STF-Male';
proc logistic data= maleclm;
class STFcat2(ref= '< 1') / param=ref ;
model obesity3 = STFcat2 /unequalslopes;
run;
title 'ITF-Male';
proc logistic data= maleclm;
class ITFcat2(ref= '< 1') / param=ref;
model obesity3 = ITFcat2;
run;
title 'STF-Female';
proc logistic data= femaleclm;
class STFcat2(ref= '< 1') / param=ref;
model obesity3 = STFcat2;
run;
title 'ITF-Female';
proc logistic data= femaleclm;
class ITFcat2(ref= '< 1') / param=ref;
model obesity3 = ITFcat2;
run;
ods rtf close;


*** 4. multivariable analysis between multilevel besity and categorical TF;
** cat_var = ethnic2 Current_Drinker Current_Smoker;
** num_var = Age sleep_mid work_mid total_per_week body_fat;

** 4.1 multinomial regression model;
ods rtf file='h:\ape\output\mul-multinomial.doc';
title 'STF-male';
proc logistic data= final_male;
class obesity3(ref='Normal') STFcat2(ref= '< 1') ethnic2(ref='Caucasian') / param=ref;
model obesity3 = STFcat2 ethnic2 Current_Drinker Age sleep_mid body_fat / LINK=glogit
AGGREGATE SCALE=NONE ;
run;
*********** STFcat2 ethnic2 current_Drinker Age sleep_mid body_fat;
title 'ITF-male';
proc logistic data= final_male;
class obesity3(ref='Normal') ITFcat2(ref= '< 1') ethnic2(ref='Caucasian') / param=ref;
model obesity3 = ITFcat2 ethnic2 Current_Drinker sleep_mid body_fat / LINK=glogit AGGREGATE
SCALE=NONE ;
run;
*********** ITFcat2 ethnic2 current_Drinker sleep_mid body_fat;
title 'STF-female';
proc logistic data= final_female;
class obesity3(ref='Normal') STFcat2(ref= '< 1') ethnic2(ref='Caucasian') / param=ref;
model obesity3 = STFcat2 Age body_fat/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
*********** STFcat2 Age body_fat;
title 'ITF-female';
proc logistic data= final_female;
```

```
class obesity3(ref='Normal') ITFcat2(ref= '< 1') ethnic2(ref='Caucasian') / param=ref;
model obesity3 = ITFcat2 Age body_fat/ LINK=glogit AGGREGATE SCALE=NONE ;
run;
*********** ITFcat2 Age body_fat;
ods rtf close;

** 4.2 model selection for cumulative logit model;
ods rtf file='h:\ape\output\mul-ordinal.doc';
title 'STF-Male';
proc logistic data= maleclm;
class STFcat2(ref= '< 1') ethnic2(ref='Caucasian')/ param=ref;
model obesity3 = STFcat2 ethnic2 current_Drinker Age sleep_mid body_fat;
run;
title 'ITF-Male';
proc logistic data= maleclm;
class ITFcat2(ref= '< 1') ethnic2(ref='Caucasian')/ param=ref;
model obesity3 = ITFcat2 ethnic2 current_Drinker Age sleep_mid body_fat;
run;
title 'STF-Female';
proc logistic data= femaleclm;
class STFcat2(ref= '< 1') ethnic2(ref='Caucasian')/ param=ref;
model obesity3 = STFcat2 Age sleep_mid body_fat;
run;
title 'ITF-Female';
proc logistic data= femaleclm;
class ITFcat2(ref= '< 1') ethnic2(ref='Caucasian')/ param=ref;
model obesity3 = ITFcat2 Age work_mid body_fat;
run;
title 'STF-Male-no drink';
proc logistic data= maleclm;
class STFcat2(ref= '< 1') ethnic2(ref='Caucasian') / param=ref;
model obesity3 = STFcat2 ethnic2 Age sleep_mid body_fat;
run;
title 'ITF-Male-nodrink';
proc logistic data= maleclm;
class ITFcat2(ref= '< 1') ethnic2(ref='Caucasian')/ param=ref;
model obesity3 = ITFcat2 ethnic2 Age sleep_mid body_fat;
run;
ods rtf close;


***5. further check p-value with exact test, and compare results with above univariate and multivariable
analysis;
ods rtf file='h:\ape\output\p-value check.doc';
title 'multilevel obesity-STF for male';
proc freq data=final_male;
tables obesity3*STFcat2 /expected chisq;
run;
title 'multilevel obesity-STF for female';
proc freq data=final_female;
tables obesity3*STFcat2 /expected exact;
run;
title 'multilevel obesity-ITF for male';
proc freq data=final_male;
tables obesity3*ITFcat2 /expected exact;
run;
```

```
title 'multilevel obesity-ITF for female';
proc freq data=final_female;
tables obesity3*ITFcat2 /expected exact;
run;
ods rtf close;

***6. study the association between continuous TF and multilevel obesity by;
** 6.1 first draw scatter plots;
ods rtf file='h:\ape\output\scatter plots-continuous TF.doc';
proc sgpanel data=final;
  PANELBY gender;
  title "BMI versus Summed Travel Frequency";
  scatter x=STF_mid y=avg_BMI;
run;
proc sgpanel data=final;
  PANELBY gender;
  title "BMI versus International Travel Frequency";
  scatter x=ITF_mid y=avg_BMI;
run;
ods rtf close;

**6.2 univariate (add squared term);
ods rtf file='h:\ape\output\uni-continuous TF(squared).doc';
title 'bmi-STF for male';
proc genmod data=final_male plots=all;
model avg_BMI= STF_mid STF_mid2 ;
run;
title 'bmi-STF for female';
proc genmod data=final_female plots=all;
model avg_BMI= STF_mid STF_mid2 ;
run;
title 'bmi-ITF for male';
proc genmod data=final_male plots=all;
model avg_BMI= ITF_mid ITF_mid2 ;
run;
title 'bmi-ITF for female';
proc genmod data=final_female plots=all;
model avg_BMI= ITF_mid ITF_mid2 ;
run;
ods rtf close;

**6.1 multivariable (add squared term);
ods rtf file='h:\ape\output\mul-continuous TF(squared).doc';
title 'bmi-STF for male';
proc genmod data=final_male plots=all;
model avg_BMI= STF_mid STF_mid2 sleep_mid body_fat Current_Drinker;
run;
title 'bmi-STF for female';
proc genmod data=final_female plots=all;
model avg_BMI= STF_mid STF_mid2 Age sleep_mid body_fat;
run;
title 'bmi-ITF for male';
proc genmod data=final_male plots=all;
model avg_BMI= ITF_mid ITF_mid2 sleep_mid body_fat Current_Drinker;
run;
title 'bmi-ITF for female';
```

```
proc genmod data=final_female plots=all;
model avg_BMI= ITF_mid ITF_mid2 Age sleep_mid body_fat;
run;
ods rtf close;
```