DISTRIBUTION AGREEMENT

In presenting this Thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my Thesis in whole or in part in all forms of media, now or hereafter known, including display on world wide web.

I understand that I may select some access restrictions as part of the online submission of this Thesis. I retain all ownership rights to the copyright of the Thesis. I also retain the right to use in future works (such as articles or books) all or part of this Thesis.

Signature of Student Date

THE ASSOCIATION BETWEEN VITAMIN D STATUS AND OBSTRUCTIVE LUNG DISEASE AMONG U.S. ADULTS AGED 40+: 2007-2010 NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES)

By

Mohamed Seedahmed

Degree to be awarded: MPH (Applied Epidemiology)

Executive MPH

Veronika Fedirko, PhD, MPH Committee Chair	**Date**
Jordan Kempker, MD, MSc Field Advisor	**Date**
Jodie Guest, PhD, MPH Associate Director, Applied Epidemiology	**Date**
Melissa Alperin, EdD, MPH, MCHES Chair, EMPH Program	**Date**

THE ASSOCIATION BETWEEN VITAMIN D STATUS AND OBSTRUCTIVE LUNG DISEASE AMONG U.S. ADULTS AGED 40+: 2007-2010 NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES)

By

Mohamed Seedahmed

MPH, Emory University, 2018

Fellow of the American College of Physicians, 2015

MBBS, University of Khartoum, 2006

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 the Executive MPH program

2018

Abstract

For the last few decades, multiple epidemiological studies have shown that low 25-hydroxyvitamin D [25(OH)D] level is a potential risk factor for multiple chronic diseases and conditions including osteoporosis, type 1 diabetes, hypertension, infections and respiratory diseases. The primary goal of our study was to examine the association between vitamin D status and the lung function measured as the ratio of the Forced Expiratory Volume in the 1st second over the Forced Vital Capacity (FEV1/FVC) *via* spirometry screening test. We conducted a multi-year cross-sectional study of individuals aged 40 years and older, who participated in the 2007/2008 and 2009/2010 cycles of the National Health and Nutrition Examination Survey (NHANES). All statistical analyses were conducted with SAS 9.4 using PROC SURVEY procedures. The odds ratios (ORs) and 95% confidence intervals (CIs) for the association between 25(OH)D concentration and FEV1/FVC ratio below the lower limit of normal (LLN) were estimated by crude logistic regression. Further adjustment for *a priori* selected covariates did not affect the results. Of 5,477 participants who were eligible for the study, 378 had a baseline serum 25(OH)D < 30 nmol/L, 3,029 had a baseline 25(OH)D between 30-74 nmol/L and 1,558 had 25(OH)D level ≥ 75 nmol/L. Among individuals with FEV1/FVC ≥ LLN, 8% were vitamin D deficient [25(OH)D<30 nmol/L], and 31% had 25(OH)D level ≥ 75 nmol/L. Compared to those with FEV1/FVC < LLN, 7% were vitamin D deficient, and 33% had 25(OH)D level ≥ 75 nmol/L. In unadjusted analyses (crude model), 25(OH)D was not associated with baseline FEV1/FVC ratio < LLN [OR = 1.06, 95% CI: 0.96-1.16 per 25 nmol/L increases in 25(OH)D]. Furthermore, there was no significant association with baseline FEV1/FVC < LLN when we analyzed 25(OH)D as a categorical variable (*p for trend* = 0.79), with unadjusted ORs of 0.88 (95%: 0.60-1.36) for those with 25(OH)D < 30 nmol/L, and 0.96 (95% CI: 0.77-1.19) for those with 25(OH)D ≥ 75 nmol/L, when compared to individuals with 25(OH)D between 30 and 74 nmol/L. In this study, we did not find an association between 25(OH)D measurements and baseline FEV1/FVC < LLN. Our findings, though statistically non-significant, were contrary to our initial hypothesis. However, if confirmed, could be suggestive for new evidence that higher vitamin D might be associated with obstructive lung disease.

THE ASSOCIATION BETWEEN VITAMIN D STATUS AND OBSTRUCTIVE LUNG DISEASE AMONG U.S. ADULTS AGED 40+: 2007-2010 NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES)

By

Mohamed Seedahmed

MPH, Emory University, 2018

Fellow of the American College of Physicians, 2015

MBBS, University of Khartoum, 2006

Thesis Committee Chair: **Veronika Fedirko, PhD, MPH**

Thesis Field Advisor: **Jordan Kempker, MD, MSc**

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 the Executive MPH program

2018

Acknowledgments

- Author affiliations: Rollins School of Public Health, Emory University. Emory Division of Hospital Medicine. Emory University School of Medicine.

- The author thanks the Thesis committee members for their tremendous assistance and guidance, as well as my wife who continued to offer unconditional love and support.

- Conflict of Interest Statements: None to declare.

TABLE OF CONTENTS

PAGE

I. CHAPTER I: BACKGROUND & LITERATURE REVIEW …..……. 10

II. CHAPTER II: MANUSCRIPT ……..... 15

II A. ABSTRACT ……... 15

II B. INTRODUCTION …….. 16

II C. METHODS ……..... 19

II D. RESULTS …….. 24

II E. DISCUSSION ……. 28

III. CHAPTER III: SUMMARY ……. 31

IV. REFRENCES …….. 32

V. APPENDICES ……..... 37

List of Tables

Table 1. Demographics and Clinical Characteristics of Study Participants by baseline FEV1/FVC ratio by LLN, NHANES, 2007-2010 37

Table 2. Demographics and Clinical Characteristics of Study Participants by 25 (OH) Vitamin D Deficiency Status, NHANES, 2007-2010…….. 40

Table 3. The Crude and Multivariable-adjusted Associations between i25(OH)D and baseline FEV1/FVC < LLN, NHANES, 2007-2010 …… 42

Table 4. GOLD 2017 Classification [a] of Individuals with Post-bronchodilator FEV1/FVC < LLN by Vitamin D levels, NHANES 2007-2010 ……. 43

List of Figures

Figure 1. Study algorithm …….. 44

Figure 2. Modelling Process …….. 45

Figure 3: The distribution of participants with 25(OH)D level by baseline FEV1/FVC% using the LOESS procedure …… 46

Figure 4: Matrix Plot of 25(OH)D and baseline FEV1/FVC% …… 47

I. Chapter I: Background & Literature Review

While vitamin D has been classically described in regard to healthy bone metabolism, in recent years there has been increased attention towards its extraskeletal physiologic actions. In the past few decades, the importance of vitamin D rose due to growing evidence that low levels of 25-hydroxyvitamin D (25(OH)D) have been linked to clinical conditions such as rickets hypertension, ischemic heart disease, diabetes mellitus type 1, few cancers, osteoporosis, infections (Kempker *et al.*, 2016) and respiratory diseases (Hejazi *et al.*, 2016) including asthma, chronic obstructive lung disease (COPD) and respiratory infections. In order to point out a role for vitamin D status on the lung function, we studied the data from The National Health and Nutrition Examination Survey (NHANES) 2007-2010.

Rickets, which is a clinical condition that results mainly from prolonged and severe vitamin D deficiency in children, can lead to growth retardation, decreased in muscle tone and increase the risk of bone fractures. It has become a global health issue since the mid-1600s. In the 18th century, few scientists as Sniadecki and Palm recognized the importance of sun exposure in preventing and curing rickets (Holick, 2009). Later in 1918, the discovery of vitamin D by McCollum et al. and its role in preventing rickets, is considered one of the major medical accomplishment in the global health history (Holick, 2006) (Mccollum, 1975).

At the molecular level, vitamin D is a fat-soluble secosterols compound that physiologically functions as a prohormone in humans. Functionally, vitamin D helps in regulating calcium and phosphorus homeostasis in conjunction with parathyroid (PTH) hormone and calcitonin, balancing adequate mineralization of the skeleton with calcium needs of other organ systems (Deluca, 2004).

There are two major physiological forms of Vitamin D; vitamin D2 and vitamin D3. Vitamin D2, or ergocalciferol, is synthesized by fungi (as mushrooms) or commercially produced by ultraviolet irradiation of ergosterol as mold ergot (Deluca, 2004). Vitamin D3, or cholecalciferol, is naturally synthesized by the skin during exposure to specific frequencies of sunlight. Additionally, vitamin D3 can be commercially produced or cab found naturally in few foods, such as fatty fish, liver, and egg yolks (John G. Haddad, 1992). Both vitamin D2 and D3 require sequential enzymatic conversion in the liver into 25-hydroxyvitamin D (25OHD), and this, in turn, is converted in the kidney into 1,25-dihydroxyvitamin D, the active biological form (Tripkovic *et al.*, 2012).

Although 25(OH)D is not metabolically active, it is the most common and reliable way to assess someone's vitamin D status given its availability in the serum (Saenger *et al.*, 2006). In November 2010, the Institute of Medicine (IOM) released a report on the health effects associated with various serum concentrations of total Vitamin D and made recommendations on levels required for optimal health (Ross *et al.*, 2011). This report defines serum total Vitamin D levels below 30 nmol/L as associated with vitamin D deficiency, levels between 30-50 nmol/L as inadequate, and levels between 50-125 nmol/L as adequate for bone and overall health in healthy individuals, while levels more than 125 nmol/L are linked with potential toxicity (Ross *et al.*, 2011). Using these values, vitamin D deficiency is common among US adults. Data from the National Health & Nutrition Examination Survey (NHANES) 2005 to 2006 revealed that 41.6 percent of adults' participants ≥20 years-old had vitamin D deficiency, with higher prevalence among non-Hispanic blacks (82.1%), those with no college education and those with body-mass-index more than 30 (Forrest and Stuhldreher, 2011). In addition, data from NHANES 2001-2004

demonstrated that older age, female sex, winter season and smoking are associated with vitamin D deficiency (Looker *et al.*, 2008).

Several studies examined the impact of vitamin D deficiency on chronic lung disease include COPD. Clinical investigation revealed that 1,25-dihydroxyvitamin D inhibits the production of the matrix metalloproteinases, which has a major role in the proliferation of the fibroblast, and thus impact the collagen synthesis and tissue remodeling which can lead to pulmonary fibrosis and possible bronchiolitis or small airway disease (Dobak *et al.*, 1994). Other theories suggest that 25(OH)D helps to prevent the development of COPD through ;1) inhibiting the inflammatory intracellular signals as NF-kB and P38 MAP kinase pathways. 2) minimize the impact of oxidative stress in the airway, 3) decreases the release of proteases enzymes which can lead to emphysema (Heulens *et al.*, 2015). Also, some observational studies have shown an association between low levels of vitamin D and risk of respiratory infection (Hejazi *et al.*, 2016). Few basic science studies revealed a significant association between vitamin D receptor (VDR) polymorphism due to low vitamin D level and lower respiratory tract infections that can result into a variety of other respiratory infections which in turn can increase the risk of COPD/asthma exacerbation or acquiring tuberculosis (Roth *et al.*, 2004). These observations opened the researchers' eyes to conduct more studies and to explore the possibility that continue investigating the role of vitamin D on the lung function which can be measured by pulmonary function test.

COPD is a common respiratory disease that result from airflow obstruction, and primarily lead to exhalation-related condition (Buist *et al.*, 2007). In the United States, about 6 % (14.8-million) of the population were diagnosed with COPD in 2010 (Nhlbi, 2012). In 2015, COPD was considered the third leading cause of death in the United States, responsible for almost 5.7%

of deaths, with an estimation of more than 120,000 patients die annually from COPD-related condition (Nchs, 2015). The estimated impact of COPD on the U.S. economy is $49.9 billion in 2010, and roughly $29.5 billion per year as direct cost (Nhlbi, 2012).

It might take years before patients with COPD become symptomatic enough to be diagnosed, and for the most part they present far late seeking medical attention. Primary care physicians can play a major role in early detection to provide early interventions, that can help to control the socioeconomic burden of the COPD. Spirometry is one component of pulmonary function testing that is used to measure a person's airflow, through standardized maneuvers of maximal inspiration and forced exhalation. The test is the gold standard for measuring airflow limitation and diagnosing obstructive lung diseases with the American College of Physicians (ACP), American Thoracic Society (ATS) and American College of Chest Physicians (ACCP) each recommending the use of spirometry to diagnose airflow obstruction in patients with respiratory symptoms (Qaseem *et al.*, 2011). To assist in these diagnostic efforts, spirometry data from NHANES III (1988-1994) were used to develop population reference values that adjust for age, gender and race/ethnicity (Hankinson *et al.*, 1999). Similarly, health examination data from NHANES 2007-2010 estimated the prevalence of COPD among United States adults aged 40-79 years was 10.2% (Tilert *et al.*, 2013). A retrospective cohort study of 96 patients not taking vitamin D supplements demonstrated that severe vitamin D deficiency was associated with more frequent COPD exacerbations and hospitalizations, but there was no significant relationship with spirometry measurements (Malinovschi *et al.*, 2014). Despite these provocative findings, the relationships between vitamin D status and COPD are still inconclusive. Furthermore, it is unknown if these associations depend on the type of vitamin D (D3 or D2) measured.

The goal of this study is to utilize a large, representative sample of US adults, to examine the relationship between vitamin D status and obstructive lung disease pattern among U.S. adult aged 40+ years old.

II. Chapter II: Manuscript Chapter

II A. ABSTRACT

For the last few decades, multiple epidemiologic studies have shown that low 25-hydroxyvitamin D [25(OH)D] level is a potential risk factor for multiple chronic diseases and conditions including osteoporosis, type 1 diabetes, hypertension, infections and respiratory diseases. The primary goal of our study was to examine the association between vitamin D status and the lung function measured as the ratio of the Forced Expiratory Volume in the 1st second over the Forced Vital Capacity (FEV1/FVC) *via* spirometry screening test. We conducted a multi-year cross-sectional study of individuals aged 40 years and older, who participated in the 2007/2008 and 2009/2010 cycles of the National Health and Nutrition Examination Survey (NHANES). All statistical analyses were conducted with SAS 9.4 using PROC SURVEY procedures. The odds ratios (ORs) and 95% confidence intervals (CIs) for the association between 25(OH)D concentration and FEV1/FVC ratio below the lower limit of normal (LLN) were estimated by crude logistic regression. Further adjustment for *a priori* selected covariates did not affect the results. Of 5,477 participants who were eligible for the study, 378 had a baseline serum 25(OH)D < 30 nmol/L, 3,029 had a baseline 25(OH)D between 30-74 nmol/L and 1,558 had 25(OH)D level ≥ 75 nmol/L. Among individuals with FEV1/FVC ≥ LLN, 8% were vitamin D deficient [25(OH)D<30 nmol/L], and 31% had 25(OH)D level >75 nmol/L. Compared to those with FEV1/FVC < LLN, 7% were vitamin D deficient, and 33% had 25(OH)D level ≥ 75 nmol/L. In unadjusted analyses (crude model), 25(OH)D was not associated with baseline FEV1/FVC ratio < LLN [OR = 1.06, 95% CI: 0.96-1.16 per 25 nmol/L increases in 25(OH)D]. Furthermore, there was no significant association with baseline FEV1/FVC < LLN when we analyzed 25(OH)D as a categorical variable (*p for trend* = 0.79), with unadjusted ORs of 0.88 (95%: 0.60-1.36) for those with 25(OH)D < 30 nmol/L, and 0.96 (95% CI: 0.77-1.19) for those with 25(OH)D ≥ 75 nmol/L, when compared to individuals with 25(OH)D between 30 and 74 nmol/L. In this study, we did not find an association between 25(OH)D measurements and baseline FEV1/FVC < LLN. Our findings, though statistically non-significant, were contrary to our initial hypothesis. However, if confirmed, could be suggestive for new evidence that higher vitamin D might be associated with obstructive lung disease.

II B. INTRODUCTION

While vitamin D has been classically described in regard to healthy bone metabolism, in recent years there has been increased attention towards its extraskeletal physiologic actions. In the past few decades, the importance of vitamin D rose due to growing evidence that low levels of 25-hydroxyvitamin D (25(OH)D) have been linked to clinical conditions such as rickets hypertension, ischemic heart disease, diabetes mellitus type 1, few cancers, infections (Kempker *et al.*, 2016), osteoporosis and respiratory diseases (Hejazi *et al.*, 2016) including asthma, chronic obstructive lung disease (COPD) and respiratory infections. In order to point out a role for vitamin D status on the lung function, we studied the data from The National Health and Nutrition Examination Survey (NHANES) 2007-2010.

There are two major physiological forms of Vitamin D; vitamin D2 and vitamin D3. Vitamin D2, or ergocalciferol, is synthesized by fungi (as mushrooms) or commercially produced by ultraviolet irradiation of ergosterol as mold ergot (Deluca, 2004). Vitamin D3, or cholecalciferol, is naturally synthesized by the skin during exposure to specific frequencies of sunlight. Additionally, vitamin D3 can be commercially produced or cab found naturally in few foods, such as fatty fish, liver, and egg yolks (John G. Haddad, 1992). Both vitamin D2 and D3 require sequential enzymatic conversion in the liver into 25-hydroxyvitamin D (25OHD), and this, in turn, is converted in the kidney into 1,25-dihydroxyvitamin D, the active biological form (Tripkovic *et al.*, 2012).

In November 2010, the Institute of Medicine (IOM) released a report on the health effects associated with various serum concentrations of total Vitamin D and made recommendations on levels required for optimal health (Ross *et al.*, 2011). This report defines serum total Vitamin D

levels below 30 nmol/L as associated with vitamin D deficiency, levels between 30-50 nmol/L as inadequate, and levels between 50-125 nmol/L as adequate for bone and overall health in healthy individuals, while levels more than 125 nmol/L are linked with potential toxicity (Ross *et al.*, 2011). Using these values, vitamin D deficiency is common among US adults. Data from the National Health & Nutrition Examination Survey (NHANES) 2005 to 2006 revealed that 41.6 percent of adults' participants ≥20 years-old had vitamin D deficiency, with higher prevalence among non-Hispanic blacks (82.1%), those with no college education and those with body-mass-index more than 30 (Forrest and Stuhldreher, 2011). In addition, data from NHANES 2001-2004 demonstrated that older age, female sex, winter season and smoking are associated with vitamin D deficiency (Looker *et al.*, 2008).

COPD is a common respiratory disease that result from airflow obstruction, and primarily lead to exhalation-related condition (Buist *et al.*, 2007). In the United States, about 5% (14.8-million) of the population were diagnosed with COPD in 2010 (Nhlbi, 2012). In 2015, COPD was considered the third leading cause of death in the United States, responsible for almost 5.7% of deaths, with an estimation of more than 120,000 patients die annually from COPD-related condition (Nchs, 2015). The estimated impact of COPD on the U.S. economy is $49.9 billion in 2010, and roughly $29.5 billion per year as direct cost (Nhlbi, 2012).

It might take years before patients with COPD become symptomatic enough to be diagnosed, and for the most part they present far late seeking medical attention. Primary care physicians can play a major role in early detection to provide early interventions, that can help to control the socioeconomic burden of the COPD. Spirometry is one component of pulmonary function testing that is used to measure a person's airflow, through standardized maneuvers of maximal inspiration and forced exhalation. The test is the gold standard for measuring airflow

limitation and diagnosing obstructive lung diseases with the American College of Physicians (ACP), American Thoracic Society (ATS) and American College of Chest Physicians (ACCP) each recommending the use of spirometry to diagnose airflow obstruction in patients with respiratory symptoms (Qaseem *et al.*, 2011). To assist in these diagnostic efforts, spirometry data from NHANES III (1988-1994) were used to develop population reference values that adjust for age, gender and race/ethnicity (Hankinson *et al.*, 1999). Similarly, health examination data from NHANES 2007-2010 estimated the prevalence of COPD among United States adults aged 40-79 years was 10.2% (Tilert *et al.*, 2013).

Several studies examined the impact of vitamin D deficiency on chronic lung disease include COPD. For example, a clinical investigation revealed that 1,25-dihydroxyvitamin D inhibits the production of the matrix metalloproteinases, which has a major role in the proliferation of the fibroblast, and thus impact the collagen synthesis and tissue remodeling which can lead to pulmonary fibrosis and possible bronchiolitis or small airway disease (Dobak *et al.*, 1994). A retrospective cohort study of 96 patients not taking vitamin D supplements demonstrated that severe vitamin D deficiency was associated with more frequent COPD exacerbations and hospitalizations, but there was no significant relationship with spirometry measurements (Malinovschi *et al.*, 2014). Despite these provocative findings, the relationships between vitamin D status and COPD are still inconclusive. Furthermore, it is unknown if these associations depend on the type of vitamin D (D3 or D2) measured. These observations opened the researchers' eyes to conduct more studies and to explore the possibility that continue investigating the role of vitamin D on the lung function which can be measured by pulmonary function test.

The goal of this study is to utilize a large, representative sample of US adults, to examine the association between vitamin D status and obstructive lung disease pattern among U.S. adult aged 40+ years old.

II C. METHODS

Data source and participants' selection

NHANES is a national survey of the National Center for Health Statistics that has produced national vital and health statistics since the 1960s. In 1999, the continuous NHANES was started, and data are collected annually, on a 2-year cycle, generating population-level data. We conducted a multi-year cross-sectional study of individuals who participated in the 2007-2010 cycles.

The survey is conducted through a complex, multistage sampling design, which includes design variables and weights to produce statistical estimates of the civilian, non-institutionalized US population. NHANES data are available for public-use on the National Center for Health Statistics (NCHS) website (https://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm). The data were sorted, merged and concatenated using the unique sequence number given to each NHANES participants, in addition to specific identifier number that code for the 2-year cycle, 2007-2008, and 2009-2010.

NHANES 2007-2010 included spirometry exams as part of the respiratory health data which was sponsored by the National Heart, Lung and Blood Institute of the National Institute of Health, the Centers for Disease Control and Prevention (CDC), National Institute for Health

Statistics and the National Institute for Occupational Safety and Health (NIOSH). The objectives of the spirometry data are to assess the prevalence of obstructive lung disease as asthma and COPD. Additionally, is to update the spirometry reference data that adjust for age, gender and race/ethnicity which was last updated in 1999 by Hankinson et. al. (Hankinson *et al.*, 1999).

We included those participants ≥ 40 years of age, who have completed the spirometry tests. NHANES 2007-2010 have excluded those who were using supplemental oxygen, had a recent eye, chest or abdominal surgery; had active cardiovascular conditions or recent history of stroke or tuberculosis exposure; personal history of coughing blood, retinal detachment or collapsed lung from performing spirometry tests (Cdc, 2007-2008b). Eligible participants performed a baseline spirometry test, then selected subsample of participants with forced expiratory volume in the 1st second/Forced Vital Capacity ratio (FEV1/FVC) below the lower limit of normal (LLN) and/or fixed ratio of 0.70, did repeat a second spirometry test using bronchodilator treatment. The post-bronchodilator test helps to differentiate between asthma (reversible defects) and COPD (permanent defects) (Figure 1).

Variables' and covariates' definitions

An important aspect of NHANES 2007-2010 was the use of the same procedures and equipment that met the American Thoracic Society's (ATS) spirometry recommendations and ensuring high standard quality control.(Cdc, 2007-2008b). The primary outcome of interest was FEV1/FVC ratio was 2-category variable: FEV1/FVC below the LLN (indicates airflow limitation,) and FEV1/FVC above the LLN (indicated normal or restrictive lung disease). Using SAS 9.4 (SAS Institute, Cary, NC), the spirometric measurements for the baseline and post-bronchodilator exams, including FEV1, FVC and FEV1/FVC values, were recoded to calculate

the lower limit of normal (LLN) using the LLN spirometry reference equations from the 1999 Hankison (Hankinson *et al.*, 1999) paper. This helped to develop the dichotomous variable that compared the measured FEV1/FVC ratio for each participant to their reference LLN values.

The planned secondary outcome of interest was to highlight the participants' distribution for those who completed the post-bronchodilator spirometry test and had FEV1/FVC < LLN, by vitamin D status and the severity of airway obstruction using the Global Initiatives for Chronic Obstructive Lung Disease classification (Gold, 2017) (Figure 2). However, because of small sample size, particularly those with GOLD stage 3 and 4 as they were excluded from participating in the spirometry tests of NHANES 2007-2010, we did not report on this outcome.

The primary exposure of interest was serum 25(OH)D concentration. The NHANES 2007-2010 used ultra-high-performance liquid chromatography-tandem mass spectrometry (LC-MS/MS) for all vitamin D measurements. Total vitamin D, D2, D3 and epi D3 concentrations were reported as nmol/L (can be converted to ng/mL by dividing by 2.48). The values below the lower limit of detection (LLOD) were imputed, by dividing each LLOD value by the square root of 2 (LLOD/√2) (Cdc, 2007-2008a). Both serum 25(OH) D and vitamin D3 were categorized based on the 2011 IOM classifications (Ross *et al*., 2011) into 3 major categories: deficiency (<30 nmol/L) inadequate or adequate level (30 - 75 nmol/L) and above normal level (>75 nmol/L). Additionally, to examine the covariates by serum 25(OH) D, we dichotomized the levels using 30 nmol/L as a cutpoint, which was defined as the lower limit of normal recommended by the IOM. (Ross *et al.*, 2011). However, because the optimal 25(OH)D concentration for the normal respiratory function is uncertain, we also explored 25(OH)D as a continuous variable. Scatterplots with LOESS & SPLINE curves of FEV1/FVC<LLN by 25(OH)D and vitamin D3 revealed similar distribution, compared to vitamin D2.

Other covariates data were collected from the NHANES included demographic information, race/ethnicity, body mass index (BMI), family income, education level, marital status, season during which phlebotomy was performed, behaviors and dietary information as a self-reported antacid and tobacco usage, in addition to C-reactive protein level (CRP). The risk of developing obstructive lung disease due to smoking tobacco is well established in the literature as the main risk factor (Kuempel *et al.*, 2009), the same study also concluded that age and race are significant predictors of emphysema severity. Therefore, we created tobacco-pack-year variable to assess the impact of tobacco smoking. A 4-category tobacco pack-year status was created and categorized as zero (never or less than 100-cigarete during lifetime), 1-5, 6-20 and >20 pack-year. BMI is considered a risk factor for obstructive lung disease, one study indicated that overweight, BMI 25-30, is associated with lower risk of all-cause mortality among patients with COPD (Guo *et al.*, 2016). BMI was classified into 4 categories to study the impact of the underweight and obese groups compared to those with normal or over- weight. A 2-category season variable was conducted and reported in NHANES 2007-2010 in a 6-month interval. Dahl et. al. study concluded that CRP is a strong independent predictor of COPD outcomes (Dahl *et al.*, 2007). Moreover, one study suggested that individuals with high levels of CRP and vitamin D can have higher levels of pro-inflammatory cytokines (Azizieh *et al.*, 2016). Therefore, we decided to include CRP which was created as 2-category variable using 1 mg/L as the cutoff point.

Statistical Analysis

All statistical analyses were conducted with SAS 9.4 (SAS Institute, Cary, NC), using PROC SURVEY procedures to attribute to the complex sampling design. The NHANES 2007-2010 sample weights were calculated for the two 2-year cycle combined (Afzal *et al.*, 2014) and

were used in every analysis. Per the NHANES analytic guidelines for the cycles between 1999-2010 (Analytic, 2013), we applied the Taylor series method and managed missing data as not missing completely at random for variance estimation, to ensure that the regression model will be performed on the domain of nonmissing observations. Moreover, domain analysis was used to examine the study group, participants aged 40+, and therefore to provide more accurate variance estimation.

Following the descriptive analysis, we compared two different multivariable logistic regression models; Crude and a priori. The a priori model was developed based on explanatory factors/covariates shown to be related to 25(OH)D serum level and the risk of having FEV1/FVC below LLN as an indicator for potential obstructive lung disease as discussed above. These covariates included age, gender, race/ethnicity, BMI, CRP level, tobacco smoking and the season of the exam. For bivariate analyses, we defined significance as non-overlapping 95% confidence intervals or $P < 0.05$.

We developed the selected model following pre-specified multistep approach. The process started with multicollinearity assessment for the eight covariates included in the a priori model. We found no condition indices (CNIs) above 30 (highest was 23.4), and therefore no collinearity was diagnosed. In the following step, we concluded that there were no significant interaction terms between the primary exposure 25(OH)D and the other seven covariates. The mainstay of the modeling processing was to compute the odds ratio (OR) of the 25(OH)D before and after dropping each covariate, mainly to determine if there was meaningful change (>10%) from the a priori model. As a result, none of the eight covariates did fulfill the confounder definition. Consequently, the crude model was chosen to be the final model.

II D. RESULTS

Of the 20,686 participants from NHANES 2007-2010, 20,015 had conducted the survey in the mobile examination centers (MEC). 5,584 were eligible for the analyses after the exclusions criteria were applied (Figure 1).

Study Participants' Characteristics by the Baseline FEV1/FVC

Of the eligible participants in the study, only 5,477 completed high quality spirometry exams. Fourteen and half of a percent (*n*=777) of the study population had a baseline FEV1/FVC <LLN (Table 1). Those with baseline FEV1/FVC < LLN were significantly more likely to be non-Hispanic White (61%, *p*=<.0001) vs. 37% (*p*=<.0001), to have BMI between 25-30 kg/m^2 vs. 27% (*p*=<0001), chance to be in the sixth decade vs 60% (*p*=0.80), to have 25(OH)D between 30-74 nmol/L, or a higher mean 25(OH)D serum concentration; 71 nmol/L (*p*=<.0001) vs higher chance to be a smoker of 20 pack-year (45%, *p*=<0.0001). On the other hand, 57.3%, insignificantly (*p*=0.2), likely to be male compared to 42.7% to be female, and 60% (*p*=<0.15) to conduct the exam between May 1st through October 31th.

Moreover, those with baseline FEV1/FVC below the LLN compared to those with above the LLN, were significantly more likely to be non-Hispanic White (61%, *p*=<.0001) vs. 47% (*p*=<.0001) to be non-Hispanic Black, higher mean age 56.1 vs 54.7 *p*=<0001), 37% chance to have BMI between 25-30 kg/m^2 vs. >/= 30 kg/m^2 (42%, *p*=<0001), 25% (*p*=<0.002) chance to have high school degree vs. higher chance (26%,=<0.002) to carry AA degree or some college, chance to have higher mean 25(OH)D serum concentration; 71 nmol/L vs 69.8 (*p*=<.0001), with almost equal percentage distribution of 25(OH)D categories in the two groups, higher chance to be a smoker of 20 pack-year (45%, *p*=<0.0001) vs. 56% (*p*=<0.0001) chance to be a non-smoker.

Study Participants' Characteristics as related to 25(OH)D

Among the study participants, only eight percent (*n*=569) had 25(OH)D level below 30 nmol/L. We calculated the proportions of the different covariates by 25(OH)D level, (Table 2). Those with 25(OH)D < 30 nmol/L were significantly more likely to be non-Hispanic Black (52%, *p*=<.0001) vs. 58% (p=0.001), likely to be female vs. 48% (*p*=<.0001), to have BMI between 18.5-30 kg/m^2 vs. 27% (*p*=<0001), chance to be in the fourth decade with mean age 57.5 (+/- 0.75, *p*=<.0001) vs. 78% (*p*=0<.0001), to have a poverty-index-ratio more or equal than 1.0 vs. 82% (p=<.0001), to have CRP level below 1.0 mg/dL vs. 62% (*p*=<0.0001), chance to be a nonsmoker vs. 62% (*p*=<0.0001), to conduct the exam between November 1st through April 30th. We also calculated the weighted means for different spirometry measurements; among those with 25(OH)D < 30 nmol/L compared to those with 25(OH)D > 30 nmol/L, had a higher mean measured baseline FEV1/FVC% (76.3% +/- 0.74, *p*= <.0001) vs. 78.4% (+/- 0.19, *p*= <.0001), predicted baseline FEV1/FVC% vs. 68.5 % (+/- 0.17, *p*= <.0001), Lower Limit of Normal baseline FEV1/FVC%.

Adjusted Vs. Unadjusted Analysis for the Association of Baseline FEV1/FVC (Primary Outcome) with 25(OH)D Level (Primary Exposure)

In unadjusted analyses (crude model), 25(OH)D was not associated with baseline FEV1/FVC ratio < LLN [OR = 1.06, 95% CI: 0.96-1.16 per 25 nmol/L increases in 25(OH)D]. Furthermore, there was no significant association with baseline FEV1/FVC < LLN when we analyzed 25(OH)D as a categorical variable (*p for trend* = 0.79), with unadjusted ORs of 0.88

(95%: 0.60-1.36) for those with 25(OH)D < 30 nmol/L, and 0.96 (95% CI: 0.77-1.19) for those with 25(OH)D ≥ 75 nmol/L, when compared to individuals with 25(OH)D between 30 and 74 nmol/L.

In the multivariable analyses, further adjustment for age, sex, gender, BMI, CRP level, smoking tobacco and season of the exam (*a priori* selected covariates) – did not impact the result using the 10% rule, and subsequently there was no association between serum 25(OH)D level (as a continues) with baseline FEV1/FVC ratio below LLN (aOR= 0.97, 95% CI: 0.87-1.08, p *for trend*=0.55). Moreover, there was no significant association with baseline FEV1/FVC < LLN when we analyzed 25(OH)D as a categorical variable (*p for trend* = 0.22), with adjusted ORs of 0.91 (95%: 0.55-1.50) for those with 25(OH)D < 30 nmol/L, and 0.81 (95% CI: 0.63-1.03) for those with 25(OH)D ≥ 75 nmol/L, when compared to individuals with 25(OH)D between 30 and 74 nmol/L.

Following similar steps, we examined the association of vitamin D3 with baseline FEV1/FVC below LLN. The two different models: crude and a priori did not reveal any significant association as well between vitamin D3 and the risk to have baseline FEV1/FVC below LLN.

Vitamin D and vitamin D3 as related to the GOLD Classification (Secondary Outcome)

We performed a subgroup analysis of the different types of Vitamin D levels by the Global Initiatives for Chronic Obstructive Lung Disease (GOLD) classification (Table 4). Of the 326 participants who had post-bronchodilator FEV1/FVC < LLN and had measured serum vitamin D level as well, 96% (*n*= 313) were classified as GOLD 1 or 2, mainly because the study excluded participants on supplemental oxygen or history of respiratory illness.

When we compared to GOLD 2 participants, those in GOLD 1 had a higher proportion (43% vs. 39%) among 25(OH)D ≥75 nmol/L group, lower proportion (53% vs. 56%) among those 25(OH)D 30-74 nmol/L group, similar proportion (4% vs. 5%) among those 25(OH)D <30nmol/L group.

II E. Discussion

In the current study of the individuals who participated in NHANES 2007-2008 and 2009-2010, we applied a strict modeling strategy that resulted in using an unadjusted analysis as the final model, because we failed to statistically prove the presence of any confounder variable that can influence both the vitamin D status and baseline FEV1/FVC<LLN that can lead to a doubtful association. As a result, we concluded that our results were inconsistent with our a priori hypothesis that 25(OH)D level is a potential predictor for obstructive lung disease (baseline FEV1/FVC < LLN). We did not find a statistically significant association between 25(OH)D level and baseline FEV1/FVC below the lower limit of normal. But, surprisingly when we examined the relationship of 25(OH)D as categorical variable, we found an insignificant increase of the odds ratio from 0.88 for those with 25(OH)D level < 30 nmol/L, to 0.96 for those with 25(OH)D level >/= 75 nmol/L respectively. To our knowledge, this is the first study to highlight a possibility of an inverse relationship between 25(OH)D and baseline FEV1/FVC < LLN.

Reported Demographics' Proportions by FEV1/FVC<LLN of the Study Participants

Comparing the proportions from our study to the reported prevalence in 2011 by the Behavioral Risk Factor surveillance system, only 14.5% of those included in the study had FEV1/FVC<LLN with a potential to have obstructive lung disease, vs 6.0% of U.S. population diagnosed with COPD in 2010 (Nhlbi, 2012). When we reviewed the data by race, majority of those with baseline FEV1/FVC< LLN were non-Hispanic White (8.8% vs 6.3%), compared to non-Hispanic Black (3% vs 6.1%). By gender, in our study 8.3% males had baseline

FEV1/FVC<LLN vs. 5.2%. While the proportion of female with baseline FEV1/FVC<LLN were 6.3% vs. 6.7%. The likelihood explanation of these discrepancies that COPD is underdiagnosed in the reality, mainly due to the strict guidelines which are recommending spirometry exams only for those with respiratory symptoms (Qaseem *et al.*, 2011).

GOLD Classification and Vitamin D Status

Those participants in GOLD 3&4 groups were underestimated due to the small sample size, likely because the study excluded participants who were on supplemental oxygen. The reported proportions by GOLD classification is surprising as well, as we were expecting higher proportions than 4 and 5% for those with 25(OH)D level < 30 nmol/L. These results open more inquiries about how accurate are the results from previous studies that reported COPD patients are more likely to have vitamin D deficiency. Therefore, are we providing more harm when we prescribe vitamin D supplements for patients with severe COPD? We will need to have more studies that examine this relationship over time through an observational study.

Study Strengths and Limitations

The major strength of this study that the data was collected to increase its generalizability to represent a large proportion of noninstitutionalized U.S. population, through a multistage complex sampling design. Therefore, preventing the potential sample-related bias (selection, channeling and interviewer). Other strengths of the NHNAES data are the high quality and standard techniques in measuring vitamin D levels, as well as conducting spirometry examinations. Consequently, that will limit selection and performance bias.

Our study had several limitations. First, the observational cross-sectional design limits our ability to predict that the relationship between baseline FEV1/FVC<LLN is directly due to high vitamin D level. Second, it is difficult to exclude the possibility that one of the seven covariates can account for the association between 25(OH)D and baseline FEV1/FVC. In addition, the recall bias is a major threat to the NHANES data integrity. For example, even with designing a thorough tobacco smoking survey to cross-examine most of the answers. But, still the NHANES data had more than five-thousand missing data that impacted the final examination of the ability to exclude smoking tobacco as potential confounder.

In conclusion, we did not find an association between 25(OH)D measurements and baseline FEV1/FVC below the LLN. Our findings, though statistically non-significant, were contrary to our initial hypothesis. However, if confirmed, could be suggestive for new evidence that higher vitamin D might be associated with obstructive lung disease. This raise an important question that demand the need to conduct future studies, primarily to examine the relationship and to determine the causal nature if there is an association. We recommend starting with another observational study; either a cohort-study to examine the impact of vitamin D on the baseline FEV1/FVC over extended period of time would be an ideal study design, or to repeat another multi-year NHANES cross-sectional study that will include more participants from other cycles.

III. SUMMARY (Chapter III)

The role of vitamin D as potential correctable risk factor in the development of obstructive lung disease is still a hot debate in the clinical literature. As we discussed earlier, number of observational studies linked vitamin D deficiency and the development of other forms of chronic lung disease as; respiratory infections, obstructive lung disease and even pulmonary hypertension. Other studies investigated the exact mechanisms of how vitamin D can help to prevent the development of respiratory diseases, including COPD. The common denominator between most of the clinical investigations is that low vitamin D level is considered a risk factor on itself for chronic lung disease development.

Although our study failed to prove a significant association between vitamin D level and the baseline FEV1/FVC below the lower limit of normal. But, other findings came inconsistent and possibly conflicting with what have been reported in the literature. This study shine a light on the possibility that higher level of vitamin D could be unbeneficial or associated with potential obstructive lung disease. It also indicates that a lot of work needed to be done before we conclude that vitamin D supplement is beneficial or harmful. Therefore, we recommend to conduct another observational study; an ideal study will be, to conduct a cohort-study to examine the impact of vitamin D on the baseline FEV1/FVC over extended period of time. Another option as well, is to repeat another multi-year NHANES cross-sectional study that include more participants from other cycles.

.

IV. References

1. Afzal, S. and Lange et. al. Plasma 25-hydroxyvitamin D, lung function and risk of chronic obstructive pulmonary disease. *Thorax 2012; 1468-3296.*
2. NHANES Analytic Guidelines for 1999-2010. *2013 Report* {https://www.cdc.gov/nchs/data/series/sr_01/sr01_056.pdf}.
3. Azizieh, F. and Alyahya et. al. Association between levels of vitamin D and inflammatory markers in healthy women. *J Inflamm Res 2016; 1178-7031.*
4. Banerjee, J. and Roy et. al. Association of Body Mass Index (BMI) with Lung Function Parameters in Non-asthmatics Identified by Spirometric Protocols. *J Clin Diagn Res 2014; 2249-782X.*
5. Black, P. N. et. al. Relationship between serum 25-hydroxyvitamin d and pulmonary function in the third national health and nutrition examination survey. *Chest 2005; 0012-3692.*
6. Buist, A. S. and McBurnie et. al. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet 2007; 1474-547X.*
7. Kempker, J. A. et. al. Associations Between Vitamin D Level and Hospitalizations With and Without an Infection in a National Cohort of Medicare Beneficiaries. *Am J Epidemiol. 2016 May 15;183(10):920-9.*
8. NHANES 2007-2010 Data Documentation, Codebook, and Frequencies {https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/VID_E.htm}
9. NHANES 2007-2008 Spirometry Procedures Manual {https://www.cdc.gov/nchs/data/nhanes/nhanes_07_08/spirometry.pdf}.

10. CDC/NHANES Overview {https://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/nhanes_overview_brochure.pdf}.

11. Overview: Weighting in NHANES { https://www.cdc.gov/nchs/tutorials/nhanes/surveydesign/weighting/intro.htm}.

12. Dahl, M. et. al. C-reactive protein as a predictor of prognosis in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med 2007; 250-5.*

13. DeLuca, H. F. et. al. Overview of general physiologic features and functions of vitamin D. *Am J Clin Nutr 2004;8(1):18-24.*

14. Dobak, J. et. al. 1,25-Dihydroxyvitamin D3 increases collagen production in dermal fibroblasts. *J Dermatol Sci 1994.*

15. Forrest, K. Y. et. al. Prevalence and correlates of vitamin D deficiency in US adults. *Nutr Res. 2011.*

16. 2017 Report: Pocket Guide to COPD Diagnosis, Management and Prevention { http://goldcopd.org/wp-content/uploads/2016/12/wms-GOLD-2017-Pocket-Guide.pdf}.

17. Guo, Y. et. al. Body mass index and mortality in chronic obstructive pulmonary disease: A dose-response meta-analysis. *Medicine (Baltimore). 2016; doi: 10.1097/MD.*

18. Hankinson, J. L. et. al. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med 1999 Jan;159(1):179-87.*

19. Hansdottir, S. et. al. Respiratory epithelial cells convert inactive vitamin D to its active form: potential effects on host defense. *J Immunol. 2008 Nov 15;181(10):7090-9.*

20. Hejazi, M. E. et. al. A review of Vitamin D effects on common respiratory diseases: Asthma, chronic obstructive pulmonary disease, and tuberculosis. *J Res Pharm Pract. 2016 Jan-Mar;5(1):7-15. doi: 10.4103/2279-042X.176542.*

21. Heulens, N. et. al. Innate immune modulation in chronic obstructive pulmonary disease: moving closer toward vitamin D therapy. *J Pharmacol Exp Ther. 2015 May;353(2):360-8. doi: 10.1124/jpet.115.223032.*

22. Holick, M. F. et. al. Resurrection of vitamin D deficiency and rickets. *J Clin Invest. 2006 Aug;116(8):2062-72.*

23. Holick, M. F. et. al. Shining light on the vitamin D: Cancer connection IARC report. *Dermatoendocrinol. 2009 Jan;1(1):4-6.*

24. John G. Haddad, M.D. et. al. Vitamin D — Solar Rays, the Milky Way, or Both?. N *Engl J Med 1992; 326:1213-1215*

25. Kuempel, E. D. et. al. Contributions of dust exposure and cigarette smoking to emphysema severity in coal miners in the United States. *Am J Respir Crit Care Med. 2009 Aug 1;180(3):257-64.*

26. Looker, A. C. et. al. Serum 25-hydroxyvitamin D status of the US population: 1988-1994 compared with 2000-2004. *Am J Clin Nutr. 2008 Dec;88(6):1519-27.*

27. Malinovschi, A. et. al. Severe vitamin D deficiency is associated with frequent exacerbations and hospitalization in COPD patients. *Respir Res. 2014 Dec 13;15:131.*

28. McCollum et. al. Nutrition classics from The Journal of Biological Chemistry 53:293-312, 1922. Studies of experimental rickets. XXI. An experimental demonstration of the existence of a vitamin which promotes calcium depositor. *Nutr Rev. 1975 Feb;33(2):48-50.*

29. Murray, C. J. et. al. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet. 1997 May 24;349(9064):1498-504.*

30. *Health, United States 2015 with Special Feature on Racial and Ethnic Health Disparities*. Murray CJ et. al. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet. 1997 May 24;349(9064):1498-504.*

31. NHANES Participants { https://www.cdc.gov/nchs/nhanes/participant.htm}.

32. NHANES Analytic Guidelines for 1999-2010 { https://www.cdc.gov/nchs/data/series/sr_01/sr01_056.pdf}.

33. Morbidity & Mortality Chart Book on CVS, lung and blood diseases. { https://www.nhlbi.nih.gov/files/docs/research/2012_ChartBook_508.pdf}.

34. Obermajer, J. et. al. [Jedrzej Sniadecki and the treatment of rickets] *Cas Lek Cesk. 1985 Jul 19;124(29):919-21*

35. Qaseem, A. et. al. Diagnosis and management of stable chronic obstructive pulmonary disease: a clinical practice guideline update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society. *Ann Intern Med. 2011 Aug 2;155(3):179-91.*

36. Ross, A. C. et. al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab. 2011 Jan;96(1):53-8.*

37. Roth, D. E. et. al. Association between vitamin D receptor gene polymorphisms and response to treatment of pulmonary tuberculosis. *J Infect Dis. 2004 Sep 1;190(5):920-7.*

38. Saenger, A. K. et. al. Quantification of serum 25-hydroxyvitamin D(2) and D(3) using HPLC-tandem mass spectrometry and examination of reference intervals for diagnosis of vitamin D deficiency. *Am J Clin Pathol. 2006 Jun;125(6):914-20.*

39. Tilert, T. et. al. Estimating the U.S. prevalence of chronic obstructive pulmonary disease using pre- and post-bronchodilator spirometry: the National Health and Nutrition Examination Survey (NHANES) 2007-2010. *Respir Res. 2013 Oct 9;14:103.*

40. Tripkovic, L. et. al. Comparison of vitamin D2 and vitamin D3 supplementation in raising serum 25-hydroxyvitamin D status: a systematic review and meta-analysis. *Am J Clin Nutr. 2012 Jun;95(6):1357-64.*

V. Appendices

V A. Table 1

Table 1. Demographics and Clinical Characteristics of Study Participants by Baseline FEV1/FVC ratio below and above LLN [a], NHANES, 2007-2010.

Characteristic	FEV1/FVC< LLN [b]		FEV1/FVC ≥ LLN [b]		p[c]
	N	%	N	%	
	777	14.5	4700	85.5	
AGE, years					
40 – 49	195	25	1459	31	*<.0001*
50 – 59	202	26	1270	27	
60 - 69	210	27	1219	26	
>/= 70	170	22	752	16	
Mean ± SE [d]	56.1 ± 0.41		54.7 ± 0.22		*<.0001*
Race					
Mexican American	72	9	852	18	*<.0001*
Other Hispanic	55	7	536	11	
Non-Hispanic White	474	61	2230	47	
Non-Hispanic Black	154	20	892	20	
Other race or Multi-racial	22	3	190	4	
Gender					
Male	445	57.3	2280	48.5	*0.21*
Female	332	42.7	2420	51.5	
BMI [e], kg/m^2					
< 18.5	26	3	33	1	*<.0001*
18.5 – 25	258	33	946	20	
25 - 30	283	37	1716	37	
≥30	205	27	1987	42	
Missing [l]	5		18		
Education Level					
< 9th grade	85	11	624	13	*0.002*
9-11th grade	146	19	715	15	
High school/GED [i] or equivalent	220	28	1073	23	
AA [j] degree or some college	196	25	1239	26	
College graduate or higher	129	17	1044	22	
Missing [l]	1		5		
Marital Status					
Married	454	60	2908	62	*0.32*
Divorced or Separated	223	27	1236	27	
Never Married	100	13	552	11	

Missing [l]	0		4		
Income-to-Poverty ratio [f]					
Below poverty level (<1.0)	130	18	691	16	*0.40*
Above poverty level (≥1.0)	594	82	3562	84	
Missing [l]	53		447		
Total 25(OH)-vitamin D, nmol/L					
< 30	49	7	329	8	*0.80*
30 – 74	422	60	2607	61	
≥ 75	229	33	1329	31	
Mean ± SE	71 ± 1.14		69.8 ± 0.90		*<.0001*
Missing [l]	77		435		
25(OH)-vitamin D_3, nmol/L					
< 30	72	10	469	11	*0.73*
30 – 74	441	63	2696	63	
≥ 75	187	27	1101	26	
Mean ± SE	66.9 ± 1.15		65.7 ± 0.88		*<.0001*
Missing [l]	77		435		
CRP [k], mg/dL					
< 1.0	664	90	4096	90	*0.23*
≥1	72	10	434	10	
Missing [l]	41		170		
Antacid Intake [g]					
Yes	112	14	627	13	*0.87*
No	665	86	4071	87	
Missing [l]	.		2		
Tobacco, Pack-Years					
0	203	28	2530	56	*<.0001*
1 – 5	78	11	528	12	
6 – 20	121	16	638	14	
> 20	332	45	800	18	
Missing [l]	43		204		
Six-month examination time period					
November 1 – April 30	307	40	2165	46	*0.15*
May 1 – October 31	470	60	2535	54	

[a] Lower Limit of Normal (LLN).
[b] Column Percentage from the total N. The percentage is calculated from the respective column based on Baseline FEV1/FVC.
[c] Rao-Scott Chi-Square Test.
[d] SE, standard error.
[e] Body Mass Index

[f] Income-to-Poverty Ratio is the ratio of family or unrelated individual income to their appropriate poverty threshold; calculated by dividing family income by the poverty guidelines, specific to family size, as well as the appropriate year and state.
[d] Lower limit of detection divided by the square root of 2 [LLOD/sqrt (2)].
[g] All types of antacids in the past 30 days of the interview, including calcium-containing products except calcium acetate.
[h] Chi-square tests cannot be computed for the table of age by baseline FEV1/FVC because at least one table cell has 0 frequency.
[i] General Educational Development.
[j] Associate Degree
[k] C-Reactive Protein
[l] Missing values & Percent: represent column percent from the total N based on baseline FEV1/FVC.

V B. Table 2

Table 2. Demographics and Clinical Characteristics of Study Participants by 25 (OH) Vitamin D Deficiency Status [a], NHANES, 2007-2010.

Characteristic	25(OH)D ≥ 30 nmol/L[a]		25(OH)D < 30 nmol/L[a]		p [b]
	N	%[c]	N	%[c]	
	6410	92	569	8	
AGE, years					
40 – 49	1622	25	152	27	<.0001
50 – 59	1453	23	150	26	
60 - 69	1574	25	136	24	
≥ 70	1761	27	131	23	
Mean ± SE [d]	57.3 ± 0.24		57.5 ± 0.75		<.0001
Race					
Mexican American	1067	17	74	13	<.0001
Other Hispanic	681	11	43	8	
Non-Hispanic White	3465	54	132	23	
Non-Hispanic Black	936	15	294	52	
Other race or Multi-racial	259	4	26	5	
Gender					
Male	3195	49.8	239	42	0.001
Female	3215	50.2	330	58	
BMI [e], kg/m^2					
< 18.5	83	1	14	2	<.0001
18.5 – 25	1481	23	107	19	
25 – 30	2340	37	156	28	
≥ 30	2419	39	274	50	
Missing [j]	87		18		
Education Level					
< 9th grade	1019	16	85	15	<.0001
9-11th grade	1012	16	122	21	
High school/GED [h] or equivalent	1501	23	134	24	
AA [i] degree or some college	1565	24	151	27	
College graduate or higher	1300	21	77	13	
Missing [j]	13		.		
Marital Status					
Married	3825	60	264	47	<.0001
Previously Married or Separated	1899	30	207	36	
Never Married or Living with partner	683	10	97	17	

Missing [j]	3		1		
Income-to-Poverty ratio [f]					
Below poverty level (<1.0)	1043	18	114	22	*<.0001*
Above poverty level (≥ 1.0)	4790	82	401	78	
Missing [j]	577		54		
CRP (mg/dL)					
< 1.0	5743	90	465	82	*<.0001*
≥ 1	663	10	104	18	
Missing [j]	4		.		
Antacid Intake [g]					
Yes	856	13	57	10	*0.046*
No	5551	87	512	90	
Missing [j]	3		.		
Tobacco Pack-Years					
Zero	3255	77	251	62	*<.0001*
1 to 5	658	3	57	5	
6 to 20	830	7	89	11	
> 20	1368	14	143	22	
Missing [j]	299		29		
Six-month examination time period					
November 1 through April 30	2766	43	350	62	*<.0001*
May 1 through October 31	3644	57	219	38	
Spirometry Measurements					
Baseline Percentage Predicted FEV1	94.5 ± 0.43		89.5 ± 1.18		*<.0001*
Baseline Percentage Predicted FVC	97.6 ± 0.34		92.2 ± 1.10		*<.0001*
Baseline FEV1/FVC%	75.4 ± 0.30		76.3 ± 0.74		*<.0001*
Predicted Baseline FEV1/FVC% [f]	77.5 ± 0.10		78.4 ± 0.19		*<.0001*
LLN Baseline FEV1/FVC % [f]	68.0 ± 0.07		68.5 ± 0.17		*<.0001*

[a] 30 nmol/L is the cutoff used to define vitamin D deficiency status.
b Rao-Scott Chi-Square Test
[c] Column Percentage from the total N. The percentage is calculated from the respective column based on vitamin D level.
[d] Plus-minus values are: means +/- SEM (standard error of the mean).
[e] Body Mass Index
[f] Standardized reference values FEV1/FVC % = b0 (intercept PRD) + b1 * age; using predicted/LLN equation
[g] Income-to-Poverty ratio ($Income/$Threshold) = Represent the ratio of family to their appropriate poverty threshold. It was calculated by dividing family income by the poverty guidelines, specific to family size, as well as the appropriate year and state
[h] General Educational Development.
[I] Associate Degree
[j] Missing values & percent: represent column percent from the total N, based on baseline FEV1/FVC

V C. Table 3

Table 3. The Crude and Multivariable-adjusted Associations between i25(OH)D and baseline FEV1/FVC < LLN [a], NHANES, 2007-2010.

Vitamin D Measurements	Crude Model [b]				*A Priori* Model [c]			
	OR	**95% CI**	***p***	***p* trend**	**OR**	**95% CI**	***p***	***p* trend**
Total 25(OH)D, nmol/L								
< 30	0.88	0.60-1.36	*0.55*	*0.79*	0.91	0.55-1.50	*0.69*	*0.22*
30 - 74 (*Ref)*	---	---	---		---	----	---	
≥ 75	0.96	0.77-1.19	*0.69*		0.81	0.63-1.03	*0.09*	
per 25 nmol/L	1.06	0.96-1.16	*0.24*	*0.24*	0.97	0.87-1.08	*0.55*	*0.55*
25(OH)D_3, nmol/L								
< 30	0.92	0.65-1.31	*0.65*	*0.73*	0.90	0.60-1.37	*0.62*	*0.24*
30 - 74 (*Ref)*	---	---	---		---	---	---	
≥ 75	0.93	0.74-1.16	*0.49*		0.80	0.62-1.04	*0.09*	
per 25 nmol/L	1.05	0.94-1.16	*0.34*	*0.34*	0.97	0.87-1.08	*0.57*	*0.57*

[a] Lower Limit of Normal
[b] Crude model included only vitamin D measurement in the model.
[b] A priori model included the following covariates: Age Race Gender BMI CRP antacid smoking season.

V D. Table 4

Table 4. GOLD 2017 Classification [a] of Individuals with Post-bronchodilator FEV1/FVC < LLN by Vitamin D levels, NHANES 2007-2010 [b]

Serologic Measurements Of Vitamin Ds	GOLD Classification (FEV1 % Predicted)						Row Total	
	GOLD 1 (≥ 80%)		GOLD 2 (50-79 %)		GOLD 3&4 (<50%)			
	N	%	N	%	N	%	N	%
Total 25 (OH) Vitamin D (nmol/L)								
< 30	8	4	5	5	2	15	15	5
30-74	109	53	61	56	8	62	178	55
≥ 75	88	43	42	39	3	23	133	40
Column Total	205	63	108	33	13	4	326	100
Vitamin D3 levels (nmol/L)								
< 30	12	6	8	7	2	15	22	7
30-74	128	62	62	58	10	77	200	61
≥ 75	65	32	38	35	1	8	104	32
Total	205	63	108	33	13	4	326	100

[a] GOLD 2017 classification: 2017 update from the Global Initiative for Chronic Obstructive Lung Disease classification, using Percentage Predicted post-bronchodilator FEV1. It helps to assess the severity of airway obstruction for prognosis.

[b] This table represent subset of participants (N=530) who were selected for post-bronchodilator 2nd Spirometry test, after they had low FEV1/FVC below the lower-limit of normal (LLN) in their 1st spirometry test.

V E. Figure 1

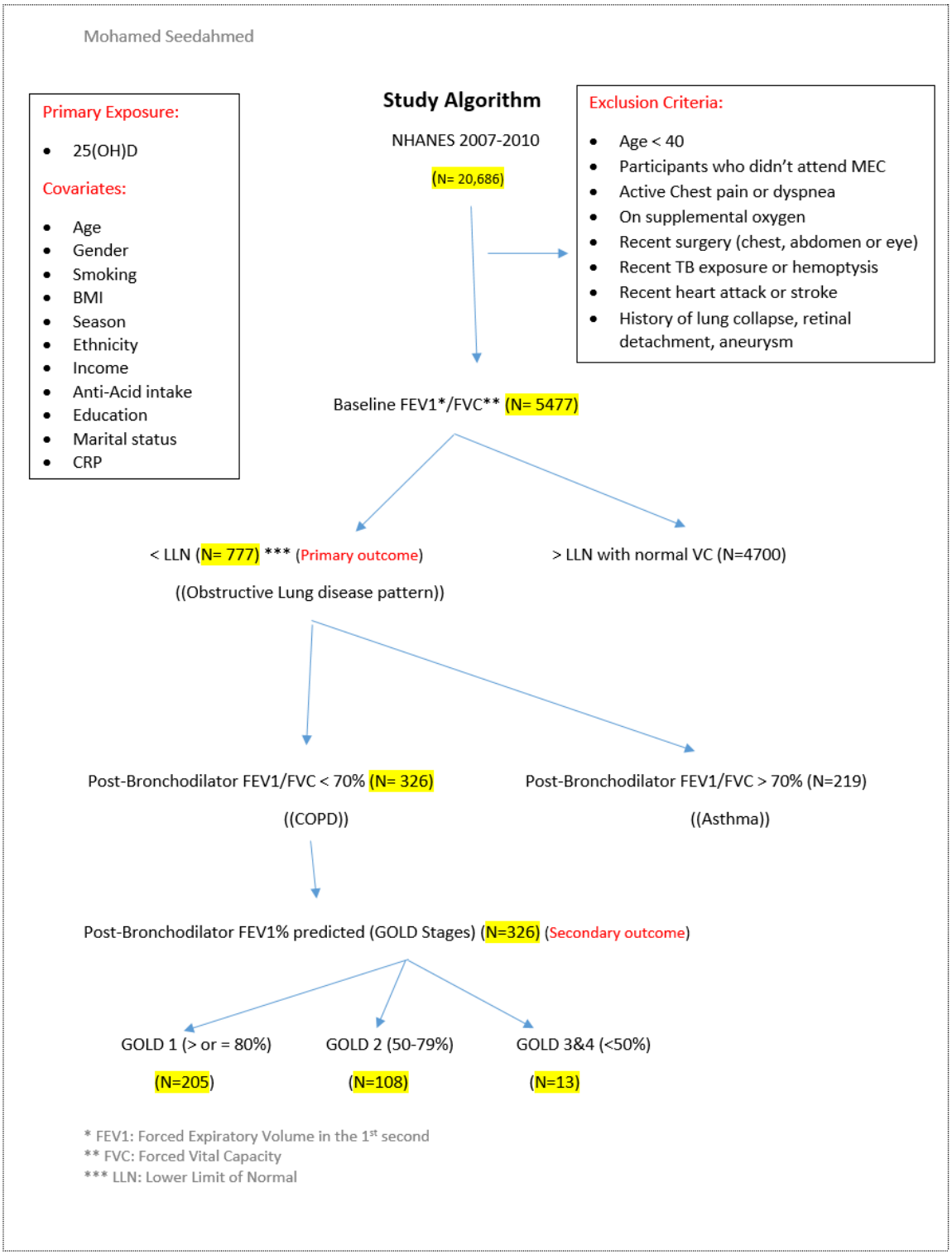

Figure 1: Study Algorithm

V F. Figure 2

Appendix - Modelling Process for Table 3

(Total Vitamin D)

Model		Vitamin D		Within 10% of the a priori model Non confounding Zone (0.87-.1.07)	p-value
		aOR	aOR C.I.		
A Priori	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.97	0.87-1.08	---	0.55
1	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP antacid smoking season	0.97	0.87-1.08	Yes	0.56
2	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.99	0.90-1.10	Yes	0.88
3	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.97	0.87-1.07	Yes	0.51
4	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	1.05	0.95-1.16	Yes	0.34
5	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.97	0.87 – 1.08	Yes	0.55
6	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.92	0.83-1.02	Yes	0.11
7	Baseline FEV/FVC = Vit D Age Race Gender BMI CRP smoking season	0.97	0.87 – 1.08	Yes	0.60
Crude	Baseline FEV/FVC = Vit D	1.06	0.96-1.16	Yes	0.24

Figure 2: Modeling Process

V G. Figure 3

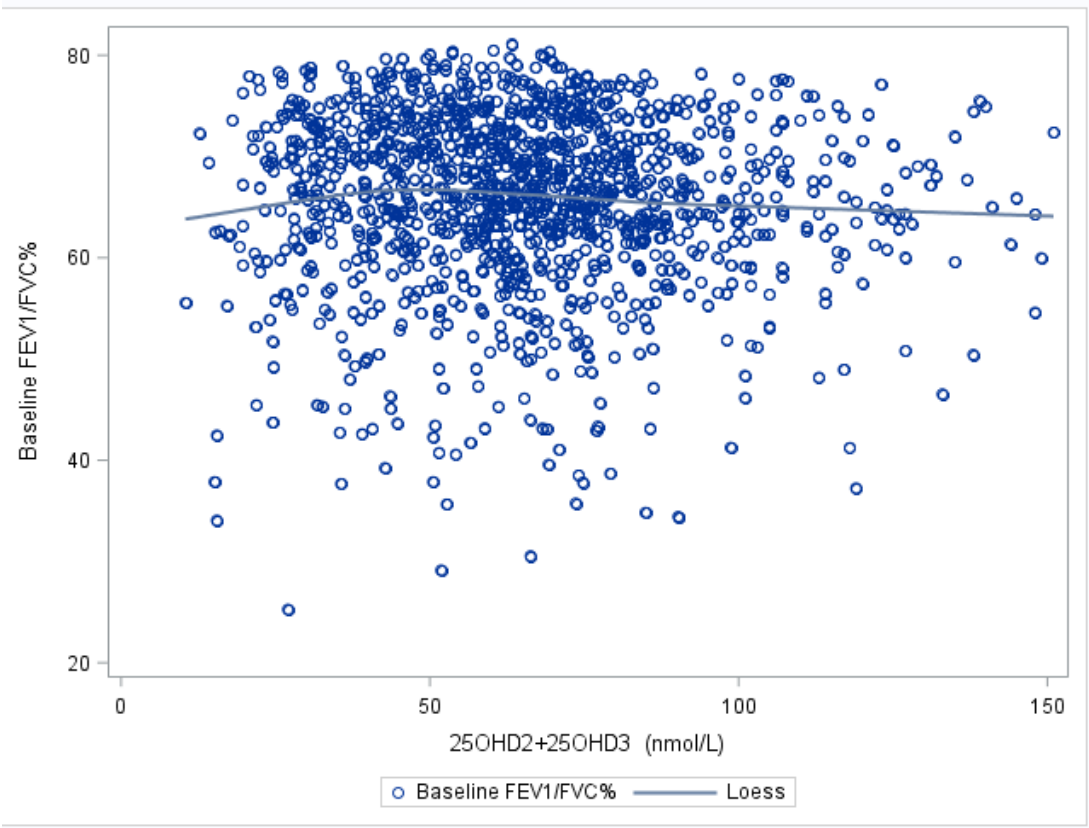

Figure 3: The distribution of participants with 25(OH)D level by baseline FEV1/FVC% using the LOESS procedure

V H. Figure 4

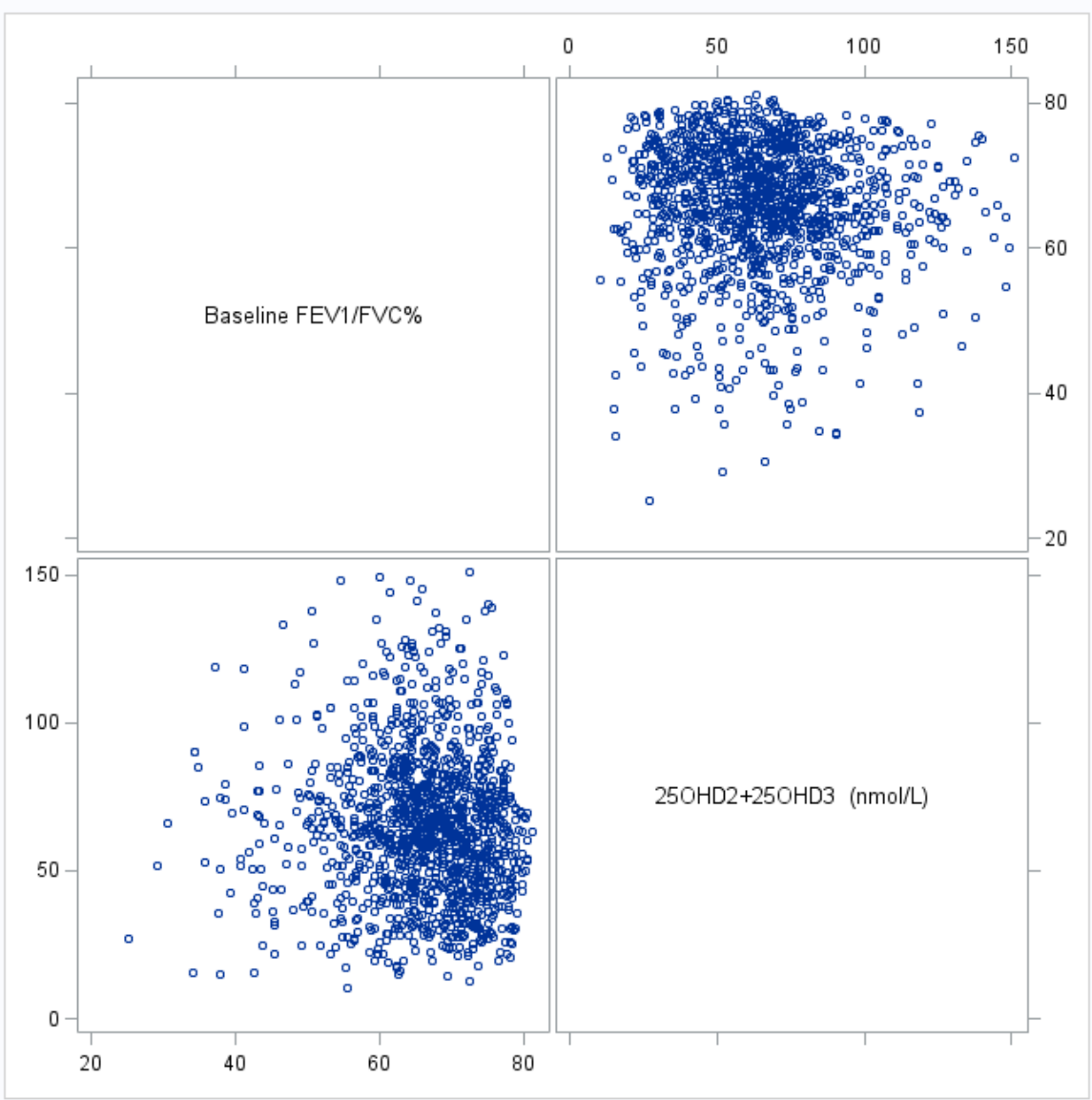

Figure 4: Matrix Plot of 25(OH)D and baseline FEV1/FVC%

V I. Data Dictionary

Demographic Variables (DEMO_E or F)

Variable	Description	Response/Count	
RIDSTATR	Interview and Examination Status of the Sample Person (0-150 Years)	1 Interviewed Only 2 Both Interviewed and MEC examined . Missing	387+ 284 9762+ 10253 = 20015 0 (both datasets)
SDDSRVYR	NHANES Public Release Year	5 NHANES 07/08 6 NHANES 09/10 . Missing	10149 10537 0 (both datasets)
SEQN	Respondent sequence number		
RIDAGEYR	Age in years both gender 0-150 years	Numeric (0-79, >= 80)	
RIAGENDR	Indicates gender	1 male →Ref 2 females . Missing	
RIDRETH1	Race, uniform coding	1 Mexican American 2 other Hispanic 3 Non-Hispanic White →Ref 4 Non-Hispanic Black 5 Other Race-Including Multi-Racial . Missing	
DMDMARTL	Marital status	1 Married →Ref 2 Widowed 3 Divorced 4 Separated 5 Never married 6 Living with partner 77 Refused 99 Don't know	
INDFMPIR	Income-to-Poverty ratio; ratio of household family income to poverty threshold	0 to 4.99 range 5 Value greater than or equal to 5.00	

DMDEDUC2	Education Level - Adults 20+ (20 years -150 years)	1 < 9th grade 2 9- 11thGrade (include 12th grade with no diploma) 3 High school grad/GED or Equivalent→Ref 4 AA degree or some college 5 College graduate or above 7 Refused 9 Don't know . Missing
RIDEXMON	Six-month time period when the examination was performed	1 November 1 through April 30 2 May 1 through October 31 →Ref
WTMEC2YR	Full Sample 2 Year MEC Exam Weight (for both interviewed and MEC examined sample persons)	0 to 192770.73006 (07/08) 0 to 158146.91752 (09/10) . Missing
SDMVSTRA	Masked Variance Pseudo-Stratum	59 to 74 . Missing
SDMVPSU	Masked Variance Pseudo-PSUs (Primary Sampling Unit)	1 to 2 . Missing

Masked Variance Units: Masked Variance Strata and Masked Variance Units or "MVUs" are included in the Demographics data file. The MVUs are a collection of secondary sampling units that are aggregated into groups for the purpose of variance estimation. The variance estimates that are produced, using the masked strata and MVUs, closely approximate the variances that would have been estimated using the "true" sample design variance units that are based on the actual survey sample strata and primary sampling units (PSUs). MVUs are used to protect the confidentiality of information provided by survey participants and to reduce disclosure risks. The MVUs are described in the NHANES Analytic Guidelines. Analysts should review the Guidelines carefully prior to analyzing the survey data.

Sample Weights: The 2-year sample weights (WTINT2YR, WTMEC2YR) should be used for all NHANES 2007–2008 analyses. There are no 4-year weights in this file. The 4-year weights were only provided with the NHANES 2001–2002 release file. Detailed instructions for linking earlier datasets (1999–2000, 2001–2002, 2003-2004, and 2005-2006) are provided in the NHANES Analytic Guidelines.

Dietary variables (DSQTOT_E or F)

Variable	Description	Response
DSQTVD	Vitamin D (D2+D3) in mcg for both males and females	Numeric . Missing
DSQTVC	Vitamin C in mg	Numeric . Missing
DSD010AN	Antacid intake in the past 30 days	1 Yes 2 No →Ref 7 Refused 9 Don't know . Missing
DSQTTFAT	Total fat intake in (gm)	Numeric . Missing
DSQTSFAT	Total saturated fatty acids in (gm)	Numeric . Missing
DSQTCARB	Carbohydrate intake in (gm)	Numeric . Missing
DSQTSELE	Selenium in (mcg)	Numeric . Missing

Body measurement variables (BMX_E or F) & Spirometry variables (SPX_E or F)

Variable	Description	Response
BMXHT	Standing Height (cm) with a range 81.5 to 203.8	Numerical
BMXBMI	BMI	Numerical (12.5 to 73.43) . missing
ENQ010	Breathing problem require oxygen	1 Yes 2 No 7 Refused 9 Don't know . Missing
ENQ020	Problem taking deep breath?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ010	Have a current painful ear infection?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ020	Have you/Has SP ever had eye surgery?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ030	Eye surgery in the last 3 months?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ040	Ever had open chest/abdominal surgery?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ050	Chest/abdominal surgery last 3 months?	1 Yes 2 No

		7 Refused 9 Don't know . Missing
SPQ060	Tuberculosis in last year?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ070a	Ever told had an aneurysm?	1 Yes 77 Refused 99 Don't know . Missing
SPQ070b	Ever told had a collapsed lung?	2 Yes . Missing
SPQ070c	Ever told had a detached retina?	3 Yes . Missing
SPQ070d	Ever told had a stroke?	4 Yes . Missing
SPQ070e	Ever told had a heart attack?	5 Yes . Missing
SPQ080	Stroke in the last 3 months?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ090	Heart attack in last 3 months?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPQ100	Coughed up blood past month?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPAENA	Check item If answered No in SPQ010, SPQ030, SPQ050, SPQ060, SPQ080, SPQ090, SPQ100, ENQ010, ENQ020, and Missing in SPQ070a, SPQ070b,	

	SPQ070c set Spirometry to eligible, Otherwise set Spirometry Not Done/Safety Exclusion	
ENQ100	Had respiratory illness in the last 7 days?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SPXNSTAT	Spirometry First Test Status (baseline)	1 Complete exam 2 Partial exam 3 Not done 4 Safety exclusion . Missing
SPXNFVC	Baseline FVC (mL)	Numeric . Missing
SPXNEV *	Baseline Extrapolated Volume (mL) **Not used in the clinical practice**	Numeric . Missing
SPXNFEV1	Baseline FEV 1 (mL)	Numeric . Missing
SPXNFEV6	Baseline FEV 6 (ml)	Numeric . Missing
SPXNPEF	Baseline PEF (mL/s)	Numeric . Missing
SPXNF257	Baseline FEF 25-75% (mL/s)	Numeric . Missing
SPXNCMT	Spirometry First Test Comment (baseline spirometry)	2 SP refusal 3 No time 4 Physical limitation 5 Communication problem 6 Equipment failure 7 SP ill/Emergency 14 Interrupted 27 Proxy no information 56 Came late/ left early 84 SP with child 99 Other . Missing

Variable	Description	Response
SPXNFET	Baseline Forced Expiratory Time (s)	Numeric . Missing
SPXNQFVC	Baseline FVC Quality Attribute	A Exceeds ATS data collection standards B Meets ATS data collection standards C Potentially usable value, but does not meet all ATS standards D Questionable result, use with caution F Results are not valid . Missing
SPXNQFV1	Baseline FEV1 Quality Attribute	A Exceeds ATS data collection standards B Meets ATS data collection standards C Potentially usable value, but does not meet all ATS standards D Questionable result, use with caution F Results are not valid . Missing
SPXNQEFF	Baseline Effort Quality Attribute	A All 6 spirometry quality curve attributes are acceptable B The curves had a large time to peak flow or a non-repeatable peak flow C The curves had either < 6 seconds of exhalation or no plateau D The curves had either a cough or large extrapolated volume . Missing
SPDBRONC	Selected for Bronchodilator (Best test FEV1/FVC ratio below limit of normal and/or less than 70%)	1 Yes 2 No . Missing
SPXBSTAT	Spirometry 2nd Test Status (baseline)	1 Complete exam

		2 Partial exam 3 Not done 4 Safety exclusion . Missing
SPXBCMT	Spirometry 2nd Test Comment (baseline spirometry)	2 SP refusal 3 No time 4 Physical limitation 5 Communication problem 6 Equipment failure 7 SP ill/Emergency 14 Interrupted 27 Proxy no information 56 Came late/ left early 84 SP with child 99 Other . Missing
SPXBFVC	2nd FVC (mL)	Numeric . Missing
SPXBEV	2nd Extrapolated Volume (mL)	Numeric . Missing

SPXBFEV1	2nd FEV 1 (mL)	Numeric . Missing
SPXBFEV6	2nd FEV 6 (ml)	Numeric . Missing
SPXBPEF	2nd PEF (mL/s)	Numeric . Missing
SPXBF257	2nd FEF 25-75% (mL/s)	Numeric . Missing
SPXBFET	2nd test Forced Expiratory Time (s)	Numeric . Missing
SPXBQFVC	2nd test FVC Quality Attribute	A Exceeds ATS data collection standards B Meets ATS data collection standards C Potentially usable value, but does not meet all ATS standards D Questionable result, use with caution F Results are not valid

		. Missing
SPXBQFV1	2nd test FEV1 Quality Attribute	A Exceeds ATS data collection standards B Meets ATS data collection standards C Potentially usable value, but does not meet all ATS standards D Questionable result, use with caution F Results are not valid . Missing
SPXBQEFF	2nd test Effort Quality Attribute	A All 6 spirometry quality curve attributes are acceptable B The curves had a large time to peak flow or a non-repeatable peak flow C The curves had either < 6 seconds of exhalation or no plateau D The curves had either a cough or large extrapolated volume . Missing

Lab Values (CRP & VID)

Variable	Description	Response
LBXCRP	C-reactive protein(mg/dL)	Numeric . Missing
URXUCD	Cadmium, urine (ug/L)	Numeric . Missing
URDUCDLC	Urinary Cadmium comment code	0 At or above the detection limit 1 Below lower detection limit . Missing
URXUPB	Lead, urine (ug/L)	Numeric . Missing
URDUPBLC	Urinary Lead comment code	0 At or above the detection limit 1 Below lower detection limit . Missing
URXUCO	Cobalt, urine (ug/L)	Numeric . Missing
URDUCOLC	Urinary Cobalt (ug/L) comment code	0 At or above the detection limit 1 Below lower detection limit . Missing
LBX4PA	4-pyridoxic acid (nmol/L)	Numeric . Missing
LBXVIDMS	25OHD2+25OHD3 SI (nmol/L)	Numeric (5.41 to 213 SI) . Missing
LBDVIDLC	25OHD comment code	0 At or above the detection limit 1 Below lower detection limit . Missing
LBXVD2MS	25OHD2 SI (nmol/L)	Numeric (1.45 to 183 SI) . Missing
LBDVD2LC	25OHD2 comment code	0 At or above the detection limit 1 Below lower detection limit . Missing
LBXVD3MS	25OHD3 SI (nmol/L)	Numeric (3.02 to 212 SI) . Missing
LBDVD3LC	25OHD3 comment code	0 At or above the detection limit 1 Below lower detection limit . Missing

LBXVE3MS	epi-25OHD3 (nmol/L)	Numeric (1.16 to 23.1 SI) . Missing
LBDVE3LC	epi-25OHD3 comment code	0 At or above the detection limit 1 Below lower detection limit . Missing

Questionnaire variables (DUQ, HUQ, HIQ and SMQ)

Variable	Description	Response
PAQ690k	Use or change air filter/air cleaner	20 . Missing
DUQ220U	Last time used marijuana or hashish/unit	1 Days 2 Weeks 3 Months 4 Years 7 Refused 9 Don't know . Missing
DUQ230	# days used marijuana or hashish/month	1 to 30 (numeric) 77 Refused 99 Don't know . Missing
DUQ270Q	Last time you used cocaine, in any form	0 to 365 (numeric) 7777 Refused 9999 Don't know . Missing
DUQ280	# of days used cocaine/month	1 to 30 (numeric) 77 Refused 99 Don't know . Missing
DUQ310Q	Last time used heroin	0 to 90 (numeric) 7777 Refused 9999 Don't know . Missing
DUQ320	# of days used heroin/month	1 to 30 (numeric) 77 Refused 99 Don't know . Missing
DUQ430	Ever been in rehabilitation program	1 Yes 2 No 7 Refused 9 Don't know . Missing
HIQ011	Covered by health insurance	1 Yes

		2 No 7 Refused 9 Don't know . Missing
HUQ010	General health condition	1 Excellent 2 Very good 3 Good 4 Fair 5 Poor 7 Refused 9 Don't know . Missing
HUQ071	Overnight hospital patient in last year	1 Yes 2 No 7 Refused 9 Don't know . Missing
HUD080	#times overnight hospital patient/last year	1 to 5 6 6 times or more 77777 Refused 99999 Don't know . Missing
IMQ020	Received hepatitis B 3 dose series	1 Yes, at least 2 doses 2 Yes, Less than 2 doses 3 No 7 Refused 9 Don't know . Missing
IND247	Total savings/cash assets for the family	1 Less than $ 500 2 $505 - $ 1000 3 $ 1001 - $ 2000 4 $ 2001 - $3000 5 $ 3001 - $ 4000 6 $ 4001 - $ 5000 77 Refused 99 Don't know . Missing
INDFMMPC	Family monthly poverty level index category	1 <= 1.30 2 1.31 to 1.85 3 > 1.85

		7 Refused 9 Don't know . Missing
MCQ010	Ever been told you have asthma	1 Yes 2 No 7 Refused 9 Don't know . Missing
MCQ160G	Ever been told you have Emphysema	1 Yes 2 No 7 Refused 9 Don't know . Missing
MCQ160K	Ever been told you have Chronic Bronchitis	1 Yes 2 No 7 Refused 9 Don't know . Missing
MCQ050	Emergency care visit for asthma/past year	1 Yes 2 No 7 Refused 9 Don't know . Missing
OCQ260	Description of job/work situation	1 An employee of a private company, business, or individual for wages, salary, or commission. 2 A federal government employee 3 A state government employee 4 A local government employee 5 Self-employed in own business, professional practice or farm. 6 Working without pay in family business or farm 77 Refused 99 Don't know

		. Missing
OCQ510	Ever had work exposure to mineral dusts?	1 Yes 2 No 7 Refused 9 Don't know . Missing
OCQ530	Ever had work exposure to organic dusts?	1 Yes 2 No 7 Refused 9 Don't know . Missing
OCQ550	Ever exposed to exhaust fumes at work?	1 Yes 2 No 7 Refused 9 Don't know . Missing
PUQ100	In the past 7 days, Products used in home to control insects?	1 Yes 2 No 7 Refused 9 Don't know . Missing
PUQ110	In the past 7 days, Products used to kill weeds?	1 Yes 2 No 7 Refused 9 Don't know . Missing
PAD680	Minutes sedentary activity	Numeric 7777 Refused 9999 Don't know . Missing
RXDCOUNT	Number of prescription medicines taken	Numeric . Missing
RDQ031	Coughing most days - over 3 months period (Age 40 years to 150 years)	1 Yes 2 No 7 Refused 9 Don't know . Missing
RDQ050	Bring up phlegm most days – 3 months period (Age 40 years to 150 years)	1 Yes 2 No 7 Refused

		9 Don't know . Missing
RDQ134	Doctor prescribe wheezing medication in the past 12 months (Age 1 year to 150 years)	1 Yes 2 No 7 Refused 9 Don't know . Missing
SLQ070A	Sleep disorder: Do you have Sleep Apnea?	1 Sleep apnea 7 Refused 9 Don't know . Missing
SMD410	Does anyone smoke inside home?	1 Yes 2 No 7 Refused 9 Don't know . Missing
SMQ680	Used tobacco/nicotine last 5 days? (Age 12 years to 150 years)	1 Yes 2 No 7 Refused 9 Don't know . Missing
SMQ020	Smoked at least 100 cigarettes in life	1 Yes 2 No 7 Refused 9 Don't know . Missing
SMQ030	Age started smoking cigarettes regularly	7 to 74 0 Never smoked 80 80 years or older 777 Refused 999 Don't know . Missing
SMQ040	Do you now smoke cigarettes	1 Every day 2 Some days 3 Not at all 7 Refused 9 Don't know . Missing

SMQ050Q	How long since quit smoking cigarettes	1 to 66 77777 Refused 99999 Don't know . Missing
SMQ050U	Unit of measure (day/week/month/year)	1 Days 2 Weeks 3 Months 4 Years 7 Refused 9 Don't know . Missing
SMD055	Age last smoked cigarettes regularly	10 to 78 80 80 years or older 777 Refused 999 Don't know . Missing
SMD057	# cigarettes smoked per day when quit	2 to 90 Range of values 1 1 cigarette or less 95 95 cigarettes or more 777 Refused 999 Don't know . Missing
SMD641	# days smoked cigs during past 30 days	0 to 30 77 Refused 99 Don't know . Missing
SMD650	Average # cigarettes/day during past 30 days	2 to 90 Range of values 1 1 cigarette or less 95 95 cigarettes or more 777 Refused 999 Don't know . Missing

New Variables and Categorization

Variable	Description	Response
WTMEC4YR	Full Sample 4 Year MEC Exam Weight (because we are combining 4 years of datasets)	WTMEC4YR = WTMEC2YR/2
SEG	Study Eligible Group Age	1 Age >= 40 0 Age < 40
INDFMPIR_cat	Income-to-Poverty ratio	2 <1 Below poverty level 1 = >1 Above poverty level →Ref
Age_cat	Categorization of Age in decades	6 0-29 5 30-39 4 40- 49 3 50- 59 2 60- 69 →Ref 1 >= 70 . Missing
BMI_Cat	Categorization of BMI	4 < 18.5 (underweight) 3 18.5 - < 30 (normal/overweight) →Ref 2 30 - 39 (obese) 1 => 40 (Morbid obese) . Missing
CRP_Cat	Categorization of CRP (mg/dL)	2 < 1.0 → Ref 1 > 1.0 . Missing
LBXVIDMS_cat	Categories of total 25 (OH) Vitamin D levels in nmol/L	3 < 30 (Deficiency) 2 30 - 74 (Normal) → Ref 1 >75 (Above normal)
LBXVD3MS_cat	Categories of Vitamin D3 levels in nmol/L	3 < 30 (Deficiency) 2 30 - 74 (Normal) → Ref 1 >75 (Above normal)

LBXVD2MS_cat	Categories of Vitamin D2 levels in nmol/L	2 < 2.05 (non-detectable) →Ref 1 > 2.05 (Severe Deficiency/normal)
pk_yrs_cat	Categories of Tobacco Pack-Years (1 = pack/day for one year)	0 0 (Ref) 1 >0-5 2 >5-20 3 >20
Baseline_fevfvcratio	Baseline FEV1/FVC ratio	Numeric . Missing
Baseline_fevfvcratio_percent	Baseline FEV1/FVC in percentage	Numeric . Missing
POST_fev2fvc2ratio	post-bronchodilator FEV1/FVC ratio	Numeric . Missing
POST_fev2fvc2ratio _percent	post-bronchodilator FEV1/FVC in percentage	Numeric . Missing
Baseline_fev_ppd	Percentage Predicted baseline FEV1	Numeric . Missing
Baseline_fvc_ppd	Percentage Predicted baseline FVC	Numeric . Missing
POST_fev2_ppd	Percentage Predicted post-bronchodilator FEV1	Numeric . Missing
POST_fvc2_ppd	Percentage Predicted post-bronchodilator FVC	Numeric . Missing
ratio70_baseline	Baseline FEV1/FVC ratio using 70% cutoff	1 FEV1/FVC ratio < 70 0 FEV1/FVC ratio >= 70
ratio70_POST	post-bronchodilator FEV1/FVC ratio using 70% cutoff	1 FEV1/FVC ratio < 70 0 FEV1/FVC ratio >= 70
ratioLLN_baseline	Baseline FEV1/FVC ratio using LLN cutoff	1 if Ratio < LLN 0 if Ratio >= LLN
ratioLLN_POST	post-bronchodilator FEV1/FVC ratio using LLN cutoff	1 if Ratio < LLN 0 if Ratio >= LLN

FevfevcPRD_ref	Standardized reference values for FEV1/FVC using predicted equation FEV1/FVC % = b0 (intercept PRD) + b1 * age	Numerical in percentage
fevfvcLLN_ref	Standardized reference values for FEV1/FVC using 5th percentile LLN equation FEV1/FVC % = b0 (intercept LLN) + b1 * age	Numerical in percentage
GOLD	GOLD Classification based on Percentage Predicted post-bronchodilator FEV1	GOLD 1 (> or = 80%) GOLD 2 (50-79 %) GOLD 3&4 (<50 %)
LBXVIDMS_cat_by30nmolperL	Total 25 (OH) Vitamin D levels into two categories nmol/L	1 < 30 nmol/L (Deficiency) 2 >30 nmol/L (Normal/Above normal)”

V J. SAS Codes

```
*** Merge and concatenating the data;
libname c "H:\Thesis\NHANES III\DATA";
run;
data c.newseedahmed;
merge c.bmx_e
      c.bmx_f
      c.crp_e
      c.crp_f
       c.demo_e
      c.demo_f
       c.dsqtot_e
      c.dsqtot_f
       c.duq_e
      c.duq_f
       c.hiq_e
      c.hiq_f
       c.huq_e
      c.huq_f
       c.smq_e
      c.smq_f
       c.spx_e
      c.spx_f
       c.vid_e
      c.vid_f;
       by seqn;
run;
*In general you want to sort and keep the last one per group;
proc sort data= c.newseedahmed out= c.newseedahmed ;
by seqn SDDSRVYR;
run;
data c.newseedahmed1;
set c.newseedahmed;
by seqn SDDSRVYR ;
run;
```

```
libname c "H:\Thesis\NHANES III\DATA ";
run;
********** Limit dataset to those who went to MEC and create
SEG variable (Study Eligible Group) for those with Age >= 40
years;
DATA newseedahmed_MEC;
SET c.newseedahmed1;
if RIDSTATR=2;
if RIDAGEYR >= 40 then SEG=1; else SEG=0;
```

```
label SEG= "Study Eligible Group";
label RIDRETH1= "Race";
label RIAGENDR= "Gender";
label DMDEDUC2= "Education Level-Adults 20+";
label DMDMARTL= "Marital status";
label DSD010AN = "Antacid intake in the past 30 days";
********** Create 4-Year weight variable;
WTMEC4YR= WTMEC2YR/2;
run;
DATA newseedahmed2;
SET newseedahmed_MEC;
******** Label RIDEXMON = Six-month time period when the
examination was performed ;
Label RIDEXMON= "Examination period";
********* Categorization of Age;
If ( RIDAGEYR < 30) then age_cat = 6;
If (30 <=RIDAGEYR < 40) then age_cat = 5;
If (40 <=RIDAGEYR < 50) then age_cat = 4;
If (50 <=RIDAGEYR < 60) then age_cat = 3;
If (60 <=RIDAGEYR < 70) then age_cat = 2;
If (RIDAGEYR > = 70) then age_cat = 1;
If (RIDAGEYR = .) then age_cat = .;
label age_cat =" Age categories in decades";
********* categorization of Income-to-Poverty ratio into 2
categories= < 1 or = > 1 ;
If ( INDFMPIR < 1) then INDFMPIR_cat = 2;
If (INDFMPIR => 1 ) then INDFMPIR_cat = 1;
If (INDFMPIR = .) then INDFMPIR_cat = .;
label INDFMPIR_cat = "Income-to-Poverty ratio";

********* Categorization of baseline CRP,
** Cutoff was obtained from "C-reactive Protein As a Predictor
of Prognosis in Chronic Obstructive Pulmonary Disease"
Morten Dahl et al Study;
If (LBXCRP < 1.0) then CRP_cat = 2;
If (LBXCRP = > 1.0) then CRP_cat = 1;
If (LBXCRP = .) then CRP_cat = .;
label CRP_cat =" CRP (mg/dl) ";
********* Categorization of BMI;
If ( BMXBMI < 18.5) then BMI_cat = 4;
If (18.5 <= BMXBMI < 25) then BMI_cat = 3;
If (25 <= BMXBMI < 30) then BMI_cat = 2;
If (BMXBMI >= 30) then BMI_cat = 1;
If (BMXBMI = .) then BMI_cat = .;
label BMI_cat =" BMI Categories ";
********* categorization of total 25OH vitamin D
```

```
********* The CDC has released total 25(OH)D based on 2011
Institute of Medicine (IOM)to:
< 30 (Deficiency), 30-75 (WNL), >75 (Above Normal);
If ( LBXVIDMS < 30) then LBXVIDMS_cat = 3;
If (30 <=LBXVIDMS < 75) then LBXVIDMS_cat = 2;
If (LBXVIDMS => 75) then LBXVIDMS_cat = 1;
If (LBXVIDMS = .) then LBXVIDMS_cat = .;
label LBXVIDMS_cat =" 2011 IOM Total 25OH Vitamin D levels
Categories (nmol/L) ";
********* Categorization of Vitamin D2 (LBXVD2MS) , D3
(LBXVD3MS) and epiD3 (LBXVE3MS) levels based on For NHANES 2007-
2010
********* The method described here is designed to detect serum
25OHD2 and 25OHD3 isomers at values from approximately 2-150
nmol/L.
When 25OHD3 values are <12.5 nmol/L, the results are verified by
re-analysis.  There is no threshold level for repeats for 25OHD2
or epi-25OHD3.
Values greater than the highest calibrator are diluted with PBS-
4% albumin and confirmed through repeat testing.
The difference between repeat values should be within 15% for
25OHD3 and 20% for epi-25OHD3 and 25-OHD2. Otherwise, another
repeat needs to be done.
********* The reportable ranges of serum concentrations are as
follows:
/// 25OHD2 (2.05-75.0) nmol/L
/// 25OHD3 (2.23-150) nmol/L
/// 3-epi-25OHD3 (1.64-60.0) nmol/L
************
The detection limits were constant for all of the analysis in
the data set. The SAS variable name ending in LC (ex., LBXVD2LC)
indicates whether the result was below the limit of
detection: the value “0” means that the result was at or above
the limit of detection, “1” indicates that the result was below
the limit of detection.
For analytes with analytic results below the lower limit of
detection (ex., LBXVIDLC=1), an imputed fill value was placed in
the analyte results field.
This value is the lower limit of detection divided by the square
root of 2 (LLOD/sqrt (2)). The other variable prefixed LBX (ex.,
LBXVIDMS) provides the result for the analyte.
*********
********* The lower limits of detection (LLOD in nmol/L) for
Vitamin D metabolites are:
25-hydroxyvitamin D2----> LBXVD2MS---->  2.05 nmol/L
25-hydroxyvitamin D3---->  LBXVD3MS---->  2.23 nmol/L
epi-25-hydroxyvitamin D3---->  LBXVE3MS--->  1.64 nmol/L;
```

```
********* Actual ranges of values from the file codebook and
confirming accuracy using Proc Univariate are:
25-hydroxyvitamin D2+25-hydroxyvitamin D3 (Total)-----> LBXVIDMS
------> (5.41 to 213)SI,  n=0 below LLOD
25-hydroxyvitamin D2----> LBXVD2MS---->  (1.45 to 183)SI , n=
13,078 below LLOD
25-hydroxyvitamin D3---->  LBXVD3MS---->  (3.02 to 260) SI, n=0
below LLOD
epi-25-hydroxyvitamin D3---->  LBXVE3MS--->  (1.16 to 37.8) SI ,
n= 2,655 below LLOD
*************************************;
if ( LBXVD3MS < 30) then LBXVD3MS_cat= 3;
If ( 30 <= LBXVD3MS < 75) then LBXVD3MS_cat = 2;
If (LBXVD3MS => 75) then LBXVD3MS_cat = 1;
If (LBXVD3MS = .) then LBXVD3MS_cat = .;
label LBXVD3MS_cat ="Vitamin D3 levels in nmol/L";
if ( LBXVD2MS < 2.05) then LBXVD2MS_cat= 2;
If ( LBXVD2MS = > 2.05 ) then LBXVD2MS_cat = 1;
If (LBXVD2MS = .) then LBXVD2MS_cat = .;
label LBXVD2MS_cat ="Vitamin D2 levels in nmol/L";
run;
DATA newseedahmed3;
SET newseedahmed2;
*********I will use  SPDBRONC which is yes or no for baseline
FEV1/FVC < LLN or < 70% as my primary outcome, but also, I will
use SPXBFEV1 ( postbronchodilator FEV1) to assess for GOLD
stages;
********* I will set any variable, that I'm using in the
analysis, that has in the response option 7= refused or 9= don't
know to missing also I will make all spirometry variables
missing if the quality attribute was F or D;
IF DMDMARTL = 77 or DMDMARTL =99 THEN DMDMARTL = .;
IF DMDEDUC2 = 7 or DMDEDUC2 =9 THEN DMDEDUC2 = .;
IF DSD010AN = 7 or DSD010AN =9 THEN DSD010AN = .;
IF SPXNQFVC = 'D' or SPXNQFVC ='F' THEN SPXNFVC=.;
IF SPXNQFV1 = 'D' or SPXNQFV1 ='F' THEN do;
spxnfev1 = .;
spxnfev3 = .;
spxnfev5 = .;
spxnfev6 = .;
spxnev = .;
end;
IF SPXBQFVC = 'D' or SPXBQFVC ='F' THEN SPXBFVC=.;
IF SPXBQFV1 = 'D' or SPXBQFV1 ='F' THEN do;
spxbfev1 = .;
spxbfev3 = .;
spxbfev5 = .;
```

```
spxbfev6 = .;
spxbev = .;
end;
IF DMDMARTL = 77 or DMDMARTL =99 THEN DMDMARTL = .;
IF DSD010AN = 7 or DSD010AN =9 THEN DSD010AN = .;
IF DMDEDUC2 = 7 or DMDEDUC2 =9 THEN DMDEDUC2 = .;
IF DUQ220U = 7 or DUQ220U =9 THEN DUQ220U = .;
IF DUQ270Q= 7777 or DUQ270Q =9999 THEN DUQ270Q = .;
IF DUQ310Q = 7777 or DUQ310Q =9999 THEN DUQ310Q = .;
IF HIQ011 = 7 or HIQ011 =9 THEN HIQ011 = .;
IF HUQ010 = 7 or HUQ010 =9 THEN HUQ010 = .;
IF HUD080= 77777 or HUD080 =99999 THEN HUD080 = .;
IF MCQ010 = 7 or MCQ010 =9 THEN MCQ010 = .;
IF MCQ160G = 7 or MCQ160G =9 THEN MCQ160G = .;
IF MCQ160K = 7 or MCQ160K =9 THEN MCQ160K = .;
IF SMQ020 = 7 or SMQ020 =9 THEN SMQ020 = .;
IF SMQ030 = 777 or SMQ030 =999 THEN SMQ030 = .;
IF SMQ040 = 7 or SMQ040 =9 THEN SMQ040 = .;
IF SMQ050Q = 77777 or SMQ050Q =99999 THEN SMQ050Q = .;
IF SMQ050U = 7 or SMQ050U =9 THEN SMQ050U = .;
IF SMD055 = 777 or SMD055 =999 THEN SMD055 = .;
IF SMD057 = 777 or SMD057 =999 THEN SMD057 = .;
IF SMD641 = 77 or SMD641 =99 THEN SMD641 = .;
IF SMD650 = 777 or SMD650 =999 THEN SMD650 = .;
********** Tobacco pack-years;
IF smq050u = 1 then time_quit = (smq050q/365);
IF smq050u = 2 then time_quit = (smq050q/52);
IF smq050u = 3 then time_quit = (smq050q/12);
IF smq050u = 4 then time_quit = smq050q;
IF . < time_quit < 1 THEN time_quit = 0;
IF smq020 = 2 THEN pk_yrs = 0;
IF smq040 = 1 THEN pk_yrs = ((ridageyr - smd030)*(smd650/20));
IF smq040 = 2 THEN pk_yrs = (((ridageyr-time_quit)-
smd030)*(smd057/20));
IF smq040 = 3 THEN pk_yrs = (((ridageyr-time_quit)-
smd030)*(smd057/20));
IF pk_yrs < 0 then pk_yrs = .;
LABEL pk_yrs = "Tobacco Pack-Years ( 1 = pack/day for one
year)";
IF pk_yrs = 0  THEN pk_yrs_cat = 0;
IF 0 < pk_yrs <= 5 THEN pk_yrs_cat = 1;
IF 5 < pk_yrs <= 20 THEN pk_yrs_cat = 2;
IF pk_yrs >20  THEN pk_yrs_cat = 3;
IF pk_yrs= . then pk_yrs_cat = .;
LABEL  pk_yrs_cat = 'Tobacco Pack-Years';
********** Marital Status categorization;
IF DMDMARTL= . then DMDMARTL_cat = .;
```

```
IF DMDMARTL = 1 THEN DMDMARTL_cat = 1;
IF DMDMARTL = 2 THEN DMDMARTL_cat = 2;
IF DMDMARTL = 3 THEN DMDMARTL_cat = 2;
IF DMDMARTL = 4 THEN DMDMARTL_cat = 2;
IF DMDMARTL = 5 THEN DMDMARTL_cat = 3;
IF DMDMARTL = 6 THEN DMDMARTL_cat = 3;
LABEL  DMDMARTL_cat = 'Marital Status new categorization';

********* Create Standardized References for baseline Spirometry
Value and FEV1/FVC using prediction equations that utilize the
HtPRD coefficient as spirometric parameters;
*********Lung function parameter= b0 + b1 * age + b2 * age2 + b3
* height**2 (PRD);
IF riagendr = 1 and ridreth1 = 3 and ridageyr < 20 THEN
fevprd_ref = -0.7453 + (-.04106 * ridageyr) + (0.004477
*(ridageyr**2)) + (0.00014098 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 3 and ridageyr >= 20 THEN
fevprd_ref = 0.5536 + (-.01303 * ridageyr) + (-0.000172
*(ridageyr**2)) + (0.00014098 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr < 20 THEN
fevprd_ref = -0.7453 + (-.04106 * ridageyr) + (0.004477
*(ridageyr**2)) + (0.00014098 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr >= 20 THEN
fevprd_ref = 0.5536 + (-.01303 * ridageyr) + (-0.000172
*(ridageyr**2)) + (0.00014098 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr < 20 THEN
fevprd_ref =  -0.7048 + (-0.05711 * ridageyr) + (0.004316
*(ridageyr**2)) + (0.00013194 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr >= 20 THEN
fevprd_ref = 0.3411 + (-0.02309 * ridageyr) + (0.00013194
*(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr < 20 THEN
fevprd_ref = -0.8218 + (-0.04248 * ridageyr) + (0.004291
*(ridageyr**2)) + (0.00015104 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr >= 20 THEN
fevprd_ref = 0.6306 + (-0.02928 * ridageyr) + (0.00015104
*(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr < 20 THEN
fevprd_ref = -0.8218 + (-0.04248 * ridageyr) + (0.004291
*(ridageyr**2)) + (0.00015104 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr >= 20 THEN
fevprd_ref = 0.6306 + (-0.02928 * ridageyr) + (0.00015104
*(bmxht**2));

IF riagendr = 2 and ridreth1 = 3 and ridageyr < 18 THEN
fevprd_ref = -0.8710 + (.06537 * ridageyr)+ (0.00011496
*(bmxht**2));
```

```
IF riagendr = 2 and ridreth1 = 3 and ridageyr >= 18 THEN
fevprd_ref = 0.4333 + (-.00361 * ridageyr) + (-0.000194
*(ridageyr**2)) + (0.00011496 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr < 18 THEN
fevprd_ref = -0.8710 + (.06537 * ridageyr)+ (0.00011496
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr >= 18 THEN
fevprd_ref = 0.4333 + (-.00361 * ridageyr) + (-0.000194
*(ridageyr**2)) + (0.00011496 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr < 18 THEN
fevprd_ref =  -0.9630 + (0.05799 * ridageyr) + (0.00010846
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr >= 18 THEN
fevprd_ref = 0.3433 + (-0.01283 * ridageyr) + (-0.000097
*(ridageyr**2)) + (0.00010846 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr < 18 THEN
fevprd_ref = -0.9641 + (0.06490 * ridageyr) + (0.00012154
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr >= 18 THEN
fevprd_ref = 0.4529 + (-0.01178 * ridageyr)+ (-0.000113
*(ridageyr**2)) + (0.00012154 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr < 18 THEN
fevprd_ref = -0.9641 + (0.06490 * ridageyr) + (0.00012154
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr >= 18 THEN
fevprd_ref = 0.4529 + (-0.01178 * ridageyr)+ (-0.000113
*(ridageyr**2)) + (0.00012154 *(bmxht**2));

IF riagendr = 1 and ridreth1 = 3 and ridageyr < 20 THEN
fvcprd_ref = -0.2584 + (-0.20415 * ridageyr) + (0.010133
*(ridageyr**2)) + (0.00018642 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 3 and ridageyr >= 20 THEN
fvcprd_ref = -0.1933 + (0.00064 * ridageyr) + (-0.000269
*(ridageyr**2)) + (0.00018642 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr < 20 THEN
fvcprd_ref = -0.2584 + (-0.20415 * ridageyr) + (0.010133
*(ridageyr**2)) + (0.00018642 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr >= 20 THEN
fvcprd_ref = -0.1933 + (0.00064 * ridageyr) + (-0.000269
*(ridageyr**2)) + (0.00018642 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr < 20 THEN
fvcprd_ref =  -0.4971 + (-0.15497 * ridageyr) + (0.007701
*(ridageyr**2)) + (0.00016643 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr >= 20 THEN
fvcprd_ref = -0.1517 + (-0.01821 * ridageyr) + (0.00016643
*(bmxht**2));
```

```
IF riagendr = 1 and ridreth1 = 1 and ridageyr < 20 THEN
fvcprd_ref = -0.7571 + (-0.09520 * ridageyr) + (0.006619
*(ridageyr**2)) + (0.00017823 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr >= 20 THEN
fvcprd_ref = 0.2376 + (-0.00891 * ridageyr) + (-0.000182
*(ridageyr**2)) + (0.00017823 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr < 20 THEN
fvcprd_ref = -0.7571 + (-0.09520 * ridageyr) + (0.006619
*(ridageyr**2)) + (0.00017823 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr >= 20 THEN
fvcprd_ref = 0.2376 + (-0.00891 * ridageyr) + (-0.000182
*(ridageyr**2)) + (0.00017823 *(bmxht**2));

IF riagendr = 2 and ridreth1 = 3 and ridageyr < 18 THEN
fvcprd_ref = -1.2082 + (0.05916 * ridageyr)+ (0.00014815
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 3 and ridageyr >= 18 THEN
fvcprd_ref = -0.3560 + (0.01870 * ridageyr) + (-0.000382
*(ridageyr**2)) + (0.00014815 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr < 18 THEN
fvcprd_ref = -1.2082 + (0.05916 * ridageyr)+ (0.00014815
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr >= 18 THEN
fvcprd_ref = -0.3560 + (0.01870 * ridageyr) + (-0.000382
*(ridageyr**2)) + (0.00014815 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr < 18 THEN
fvcprd_ref =  -0.6166 + (-0.04687 * ridageyr) + (0.003602
*(ridageyr**2)) + (0.00013606 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr >= 18 THEN
fvcprd_ref = -0.3039 + (0.00536 * ridageyr) + (-0.000265
*(ridageyr**2)) + (0.00013606 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr < 18 THEN
fvcprd_ref = -1.2507 + (0.07501 * ridageyr) + (0.00014246
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr >= 18 THEN
fvcprd_ref = 0.1210 + (0.00307 * ridageyr)+ (-0.000237
*(ridageyr**2)) + (0.00014246 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr < 18 THEN
fvcprd_ref = -1.2507 + (0.07501 * ridageyr) + (0.00014246
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr >= 18 THEN
fvcprd_ref =  0.1210 + (0.00307 * ridageyr)+ (-0.000237
*(ridageyr**2)) + (0.00014246 *(bmxht**2));

********* for FEV1/FVC % = b0 (interceptPRD) + b1 * age ;
IF riagendr = 1 and ridreth1 = 3  THEN fevfvcprd_ref = 87.340 +
(-0.2066* ridageyr);
```

```
IF riagendr = 1 and ridreth1 = 5  THEN fevfvcprd_ref= 87.340  +
(-0.2066* ridageyr);
IF riagendr = 1 and ridreth1 = 4  THEN fevfvcprd_ref = 89.239 +
(-0.1828 * ridageyr);
IF riagendr = 1 and ridreth1 = 1  THEN fevfvcprd_ref = 90.024 +
(-0.2186 * ridageyr);
IF riagendr = 1 and ridreth1 = 2  THEN fevfvcprd_ref = 90.024 +
(-0.2186 * ridageyr);

IF riagendr = 2 and ridreth1 = 3  THEN fevfvcprd_ref = 90.809+
(-0.2125* ridageyr);
IF riagendr = 2 and ridreth1 = 5  THEN fevfvcprd_ref= 90.809 +
(-0.2125* ridageyr);
IF riagendr = 2 and ridreth1 = 4  THEN fevfvcprd_ref = 91.655 +
(-0.2039 * ridageyr);
IF riagendr = 2 and ridreth1 = 1  THEN fevfvcprd_ref = 92.360 +
(-0.2248 * ridageyr);
IF riagendr = 2 and ridreth1 = 2  THEN fevfvcprd_ref = 92.360 +
(-0.2248 * ridageyr);

fevprd_refml = fevprd_ref*1000;
fvcprd_refml = fvcprd_ref *1000;
fevfvcPRD_ref_ratio= fevfvcprd_ref/100;
********* Create Percentage predicted values for Baseline FEV1
and FVC;
baseline_fev_ppd = round(((spxnfev1/fevprd_refml)*100),.1);
LABEL baseline_fev_ppd = " Percentage Predicted baseline FEV1 ";
baseline_fvc_ppd = round(((spxnfvc/fvcprd_refml)*100),.1);
LABEL baseline_fvc_ppd = " Percentage Predicted baseline FVC";

********* Create Percentage predicted values for post-
bronchodilator FEV1 and FVC;
POST_fev2_ppd = round(((spxbfev1/fevprd_refml)*100),.1);
LABEL POST_fev2_ppd= "Percentage Predicted post-bronchodilator
FEV1";
POST_fvc2_ppd = round(((spxbfvc/fvcprd_refml)*100),.1);
LABEL POST_fvc2_ppd = "Percentage Predicted post-bronchodilator
FVC";

********** Create Baseline FEV1/FVC ratio and
FEV1/FVC%******************;
baseline_fevfvcratio = (spxnfev1/spxnfvc);
LABEL baseline_fevfvcratio= "Baseline FEV1/FVC ratio";
baseline_fevfvcratio_percent = (baseline_fevfvcratio* 100);
LABEL baseline_fevfvcratio_percent  = "Baseline FEV1/FVC%";
```

```
********** Create Post-pronchodilator FEV1/FVC ratio
**********************;
POST_fev2fvc2ratio = (SPXBFEV1/SPXBFVC );
LABEL POST_fev2fvc2ratio  = "post-bronchodilator FEV1/FVC
ratio";
POST_fev2fvc2ratio_percent = POST_fev2fvc2ratio* 100;
LABEL POST_fev2fvc2ratio_percent  = "post-bronchodilator
FEV1/FVC in percentage ";

*** Create a dichotomous variable for Baseline FEV1/FVC ratio
using 0.70 cutoff ;
IF baseline_fevfvcratio < 0.70 THEN ratio70_baseline = 1;
IF baseline_fevfvcratio >= 0.70 THEN ratio70_baseline = 0;
if baseline_fevfvcratio = . then ratio70_baseline = .;
LABEL ratio70_baseline = "Dichotomous Baseline FEV1/FVC ratio
using 70% cutoff";

*** Create a dichotomous variable for Post-bronchodilator
FEV1/FVC ratio using 0.70 cutoff ;

IF  POST_fev2fvc2ratio < 0.70 THEN ratio70_POST = 1;
IF POST_fev2fvc2ratio >= 0.70 THEN ratio70_POST = 0;
if POST_fev2fvc2ratio = . then ratio70_POST = .;
LABEL ratio70_POST = " Dichotomous post-bronchodilator FEV1/FVC%
using 70% cutoff";

********** Categorization of post-bronchodilator FEV1 based on -
---> GOLD classification;
If ratio70_POST=1  and POST_fev2_ppd < 50 then GOLD = 3;
If ratio70_POST=1  and (50 <=POST_fev2_ppd  < 80) then GOLD = 2;
If ratio70_POST=1  and (POST_fev2_ppd  > = 80) then GOLD = 1;
If ratio70_POST=.  and (POST_fev2_ppd = .) then GOLD = .;
label GOLD ="GOLD Classification based on Percentage Predicted
post-bronchodilator FEV1";

********* Create Standardized References for baseline Spirometry
Value and FEV1/FVC ratio using 5th Percentile
as the lower limit of the normal range (LLN) which means only 5
percent of healthy individuals will have a result below the LLN
- Using LLN rather than a fixed ratio cutoff to define air flow
obstruction reduces the misclassification that happened with the
fixed ratio approach
- Taking in consideration age, height, race and sex
********** For Prediction equation:
********** Lung function parameter= b0 + b1 * age + b2 * age2 +
b3 * height**2 (LLN);
```

```
IF riagendr = 1 and ridreth1 = 3 and ridageyr < 20 THEN
fevLLN_ref = -0.7453 + (-.04106 * ridageyr) + (0.004477
*(ridageyr**2)) + ( 0.00011607 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 3 and ridageyr >= 20 THEN
fevLLN_ref = 0.5536 + (-.01303 * ridageyr) + (-0.000172
*(ridageyr**2)) + (0.00011607 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr < 20 THEN
fevLLN_ref = -0.7453 + (-.04106 * ridageyr) + (0.004477
*(ridageyr**2)) + (0.00011607 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr >= 20 THEN
fevLLN_ref = 0.5536 + (-.01303 * ridageyr) + (-0.000172
*(ridageyr**2)) + (0.00011607 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr < 20 THEN
fevLLN_ref =  -0.7048 + (-0.05711 * ridageyr) + (0.004316
*(ridageyr**2)) + (0.00010561 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr >= 20 THEN
fevLLN_ref = 0.3411 + (-0.02309 * ridageyr) + (0.00010561
*(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr < 20 THEN
fevLLN_ref = -0.8218 + (-0.04248 * ridageyr) + (0.004291
*(ridageyr**2)) + (0.00012670 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr >= 20 THEN
fevLLN_ref = 0.6306 + (-0.02928 * ridageyr) + (0.00012670
*(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr < 20 THEN
fevLLN_ref = -0.8218 + (-0.04248 * ridageyr) + (0.004291
*(ridageyr**2)) + (0.00012670 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr >= 20 THEN
fevLLN_ref = 0.6306 + (-0.02928 * ridageyr) + (0.00012670
*(bmxht**2));

IF riagendr = 2 and ridreth1 = 3 and ridageyr < 18 THEN
fevLLN_ref = -0.8710 + (.06537 * ridageyr)+ (0.00009283
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 3 and ridageyr >= 18 THEN
fevLLN_ref = 0.4333 + (-.00361 * ridageyr) + (-0.000194
*(ridageyr**2)) + (0.00009283 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr < 18 THEN
fevLLN_ref = -0.8710 + (.06537 * ridageyr)+ (0.00009283
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr >= 18 THEN
fevLLN_ref = 0.4333 + (-.00361 * ridageyr) + (-0.000194
*(ridageyr**2)) + (0.00009283 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr < 18 THEN
fevLLN_ref =  -0.9630 + (0.05799 * ridageyr) + (0.00008546
*(bmxht**2));
```

```
IF riagendr = 2 and ridreth1 = 4 and ridageyr >= 18 THEN
fevLLN_ref = 0.3433 + (-0.01283 * ridageyr) + (-0.000097
*(ridageyr**2)) + (0.00008546  *(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr < 18 THEN
fevLLN_ref = -0.9641 + (0.06490 * ridageyr) + (0.00009890
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr >= 18 THEN
fevLLN_ref = 0.4529 + (-0.01178 * ridageyr)+ (-0.000113
*(ridageyr**2)) + (0.00009890 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr < 18 THEN
fevLLN_ref = -0.9641 + (0.06490 * ridageyr) + (0.00009890
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr >= 18 THEN
fevLLN_ref = 0.4529 + (-0.01178 * ridageyr)+ (-0.000113
*(ridageyr**2)) + (00.00009890 *(bmxht**2));

IF riagendr = 1 and ridreth1 = 3 and ridageyr < 20 THEN
fvcLLN_ref = -0.2584 + (-0.20415 * ridageyr) + (0.010133
*(ridageyr**2)) + (0.00015695 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 3 and ridageyr >= 20 THEN
fvcLLN_ref = -0.1933 + (0.00064 * ridageyr) + (-0.000269
*(ridageyr**2)) + (0.00015695 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr < 20 THEN
fvcLLN_ref = -0.2584 + (-0.20415 * ridageyr) + (0.010133
*(ridageyr**2)) + (0.00015695 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 5 and ridageyr >= 20 THEN
fvcLLN_ref = -0.1933 + (0.00064 * ridageyr) + (-0.000269
*(ridageyr**2)) + (0.00015695 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr < 20 THEN
fvcLLN_ref =  -0.4971 + (-0.15497 * ridageyr) + (0.007701
*(ridageyr**2)) + (0.000013670 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 4 and ridageyr >= 20 THEN
fvcLLN_ref = -0.1517 + (-0.01821 * ridageyr) + (0.000013670
*(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr < 20 THEN
fvcLLN_ref = -0.7571 + (-0.09520 * ridageyr) + (0.006619
*(ridageyr**2)) + (0.0001497 *(bmxht**2));
IF riagendr = 1 and ridreth1 = 1 and ridageyr >= 20 THEN
fvcLLN_ref = 0.2376 + (-0.00891 * ridageyr) + (-0.000182
*(ridageyr**2)) + (0.0001497  *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr < 20 THEN
fvcLLN_ref = -0.7571 + (-0.09520 * ridageyr) + (0.006619
*(ridageyr**2)) + (0.0001497  *(bmxht**2));
IF riagendr = 1 and ridreth1 = 2 and ridageyr >= 20 THEN
fvcLLN_ref = 0.2376 + (-0.00891 * ridageyr) + (-0.000182
*(ridageyr**2)) + (0.0001497 *(bmxht**2));
```

```
IF riagendr = 2 and ridreth1 = 3 and ridageyr < 18 THEN
fvcLLN_ref = -1.2082 + (0.05916 * ridageyr)+ (0.00012198
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 3 and ridageyr >= 18 THEN
fvcLLN_ref = -0.3560 + (0.01870 * ridageyr) + (-0.000382
*(ridageyr**2)) + (0.00012198*(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr < 18 THEN
fvcLLN_ref = -1.2082 + (0.05916 * ridageyr)+ (0.00012198
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 5 and ridageyr >= 18 THEN
fvcLLN_ref = -0.3560 + (0.01870 * ridageyr) + (-0.000382
*(ridageyr**2)) + (0.00012198  *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr < 18 THEN
fvcLLN_ref =  -0.6166 + (-0.04687 * ridageyr) + (0.003602
*(ridageyr**2)) + (0.00010916 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 4 and ridageyr >= 18 THEN
fvcLLN_ref = -0.3039 + (0.00536 * ridageyr) + (-0.000265
*(ridageyr**2)) + (0.00010916 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr < 18 THEN
fvcLLN_ref = -1.2507 + (0.07501 * ridageyr) + (0.00011570
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 1 and ridageyr >= 18 THEN
fvcLLN_ref = 0.1210 + (0.00307 * ridageyr)+ (-0.000237
*(ridageyr**2)) + (0.00011570 *(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr < 18 THEN
fvcLLN_ref = -1.2507 + (0.07501 * ridageyr) + (0.00011570
*(bmxht**2));
IF riagendr = 2 and ridreth1 = 2 and ridageyr >= 18 THEN
fvcLLN_ref =  0.1210 + (0.00307 * ridageyr)+ (-0.000237
*(ridageyr**2)) + (0.00011570 *(bmxht**2));

********* for FEV1/FVC % = b0 (interceptLLN) + b1 * age ;
IF riagendr = 1 and ridreth1 = 3   THEN fevfvcLLN_ref = 78.388 +
(-0.2066* ridageyr);
IF riagendr = 1 and ridreth1 = 5   THEN fevfvcLLN_ref=  78.388
+ (-0.2066* ridageyr);
IF riagendr = 1 and ridreth1 = 4   THEN fevfvcLLN_ref = 78.822 +
(-0.1828 * ridageyr);
IF riagendr = 1 and ridreth1 = 1   THEN fevfvcLLN_ref = 80.925+
(-0.2186 * ridageyr);
IF riagendr = 1 and ridreth1 = 2   THEN fevfvcLLN_ref = 80.925 +
(-0.2186 * ridageyr);

IF riagendr = 2 and ridreth1 = 3   THEN fevfvcLLN_ref = 81.015+
(-0.2125* ridageyr);
IF riagendr = 2 and ridreth1 = 5   THEN fevfvcLLN_ref=  81.015+
(-0.2125* ridageyr);
```

```
IF riagendr = 2 and ridreth1 = 4   THEN fevfvcLLN_ref = 80.978 +
(-0.2039 * ridageyr);
IF riagendr = 2 and ridreth1 = 1   THEN fevfvcLLN_ref = 83.044 +
(-0.2248 * ridageyr);
IF riagendr = 2 and ridreth1 = 2   THEN fevfvcLLN_ref = 83.044 +
(-0.2248 * ridageyr);

fevLLN_refml = fevLLN_ref *1000;
fvcLLN_refml = fvcLLN_ref *1000;
fevfvcLLN_ref_ratio=fevfvcLLN_ref/100;

********* Create Percentage predicted values for Baseline FEV1
and FVC using LLN cutoff;
baseline_fev_LLN = round(((spxnfev1/fevLLN_refml)*100),.1);
LABEL baseline_fev_LLN = " baseline LLN FEV1 ";
baseline_fvc_LLN = round(((spxnfvc/fvcLLN_refml)*100),.1);
LABEL baseline_fvc_LLN = " baseline LLN FVC";

************ Create Percentage predicted values for post-
bronchodilator LLN FEV1 and FVC using LLN cutoff;
POST_fev2_LLN = round(((spxbfev1/fevLLN_refml)*100),.1);
LABEL POST_fev2_LLN= "post-bronchodilator LLN FEV1";
POST_fvc2_LLN = round(((spxbfvc/fvcLLN_refml)*100),.1);
LABEL POST_fvc2_LLN = "post-bronchodilator LLN FVC";


************ Create a dichotomous variable for Baseline FEV1/FVC
% using LLN cutoff;
IF  baseline_fevfvcratio_percent < fevfvcLLN_ref THEN
ratioLLN_baseline = 1;
IF baseline_fevfvcratio_percent >= fevfvcLLN_ref THEN
ratioLLN_baseline = 0;
if baseline_fevfvcratio= . then ratioLLN_baseline = .;
LABEL ratioLLN_baseline = " Dichotomous Baseline FEV1/FVC% using
LLN cutoff";

************ Create a dichotomous variable for post-
bronchodilator LLN FEV1/FVC % ;
IF  POST_fev2fvc2ratio_percent< fevfvcLLN_ref THEN ratioLLN_POST
= 1;
IF POST_fev2fvc2ratio_percent >= fevfvcLLN_ref THEN
ratioLLN_POST = 0;
if POST_fev2fvc2ratio = . then ratioLLN_POST = .;
LABEL ratioLLN_POST = " Dichotomous post-bronchodilator
FEV1/FVC% using LLN cutoff";
run;
```

```
*****************************************************************
*****************************************************************
***************
Here we have permanently associated the variables with their
respective formats
therefore, in the future analytic procedures, we will no long
have to use the format statement
anymore, increasing efficiency and consolidation
*****************************************************************
********;
data c.newseedahmed_clean;
set newseedahmed3;
options fmtsearch=(c.newformat);
FORMAT
age_cat         age_cat.
INDFMPIR_cat    INDFMPIR_cat.
CRP_cat         CRP_cat.
bmi_cat         bmi_cat.
LBXVIDMS_cat    LBXVIDMS_cat.
LBXVD2MS_cat    LBXVD2MS_cat.
LBXVD3MS_cat    LBXVD3MS_cat.
pk_yrs_cat      pk_yrs_cat.
RIDEXMON  RIDEXMON.
GOLD        GOLD.
SEG         seg.
RIDRETH1    RIDRETH.
RIAGENDR    RIAGENDR.
DMDEDUC2    DMDEDUC.
DMDMARTL_cat    DMDMARTL_cat.
DSD010AN    DSD010AN.
ratioLLN_baseline ratioLLN_baseline.
ratioLLN_POST ratioLLN_POST.
ratio70_POST  ratio70_POST.
ratio70_baseline ratio70_baseline.
;
run;
```

```
libname c "H:\Thesis\NHANES III\DATA";
run;
***** Formatting the data;
proc format library=c.newformat;
options fmtsearch=(c.newformat);

value age_cat
                        1 = ">/= 70"
                             2 = "60-69 (Ref)"
```

```
                              3 = "50-59"
                              4 = "40-49"
                              5 = "30-39"
                              6 = "00-29";

value INDFMPIR_cat       1 = " Above poverty level= (Ref) "
                         2 = "Below poverty level";

value CRP_cat            1 = " >/= 1.0  "
                         2 = "< 1.0 =Ref";


value BMI_cat            1 = " >/= 30   (Obese/Morbid Obese) "
                         2 = " 25 - 30  (Overweight)"
                              3 = "18.5 - 25 (Normal) = Ref"
                              4 = " < 18.5   (Underweight) ";


value LBXVIDMS_cat
                         1 = ">/= 75 = (Above normal)"
                              2 = "30 - 74= (inadequate/adequate
=Ref)"
                              3 = " < 30  = (Deficiency)";

value RIDEXMON           1 = "November 1 through April 30 "
                         2 = "May 1 through October 31 = Ref";

Value LBXVD3MS_cat
                         1 = " >/= 75 = (Above normal)"
                              2 = " 30 - 74 =
((inadequate/adequate=Ref)"
                              3 = " < 30 = (Deficiency)";

Value LBXVD2MS_cat
                              1 = " >/= 2.05 (Deficiency)"
                              2 = "  < 2.05 = (Undetactable
=Ref)";


value pk_yrs_cat         0 = "0 (<100 cigarettes in life)= Ref"
                         1 = "1-5"
                              2 = "6-20"
                         3 = ">20";
```

```
Value SEG                1 = "Age >= 40 Years --> Study Eligible
Group"
                         0 = "Age < 40 Years";

Value RIDRETH            1 = "Mexican American"
                         2 = "other Hispanic"
                         3 = "Non-Hispanic White--- Ref"
                         4 = "Non-Hispanic Black"
                         5 = "Other Race-Including Multi-Racial";

Value RIAGENDR           1 = "male---- Ref"
                         2 = "female";

Value DMDEDUC            1  = "< 9th grade"
                         2  = "9- 11thGrade (include 12th grade
with no diploma)"
                         3  = "High school grad/GED or Equivalent-
--(Ref)"
                         4  = "AA degree or some college"
                         5  = "College graduate or above";

Value DMDMARTL_cat       1 = "Married---- Ref"
                         2 = "Previously married (divorced,
widowed, separated)"
                         3 = "Never married or Living with partner
";


Value DSD010AN           1  = "Yes"
                         2  = "No---Ref";

Value ratioLLN_baseline 1 = " Baseline FEV1/FVC < LLN"
                         0 = "Baseline FEV1/FVC  >= LLN";

Value ratio70_baseline 1 = " Baseline FEV1/FVC < 70%"
                         0 = "Baseline FEV1/FVC  >= 70%";

Value ratioLLN_POST       1 = "Post-bronchodilator FEV1/FVC <
LLN"
                         0 = "Post-bronchodilator FEV1/FVC >=
LLN";


Value ratio70_POST       1 = "Post-bronchodilator FEV1/FVC < 70%"
                         0 = "Post-bronchodilator FEV1/FVC >=
70%";
```

```
value GOLD                  1 = "GOLD 1(>/= 80%)->Ref"
                            2 = "GOLD 2(50-79%)"
                                 3 = "GOLD 3&4(<50%)";


value LBXVIDMS_cat_by30nmolperL 1=  "Total 25OH Vitamin D levels
>= 30 nmol/L"
                                  2=  "Total 25OH Vitamin D levels
< 30 nmol/L";

Run;
```

```
****** Table 1, by baseline characteristics FEV1/FVC ratio using
LLN ;
ods html close;ods html;
PROC SURVEYFREQ data=c.newseedahmed_clean NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
TABLES SEG*( RIDRETH1 RIAGENDR BMI_Cat DMDEDUC2 DMDMARTL_cat
INDFMPIR_cat LBXVIDMS_cat LBXVD3MS_cat LBXVD2MS_cat CRP_Cat
DSD010AN pk_yrs_cat RIDEXMON )*
ratioLLN_baseline/ expected row col chisq lrchisq wchisq
wllchisq;
RUN;
******Table 1, Age_cat by baseline characteristics FEV1/FVC
ratio using LLN ;
ods html close;ods html;
PROC SURVEYFREQ data=c.newseedahmed_clean NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
TABLES age_cat*ratioLLN_baseline/expected row col chisq lrchisq
wchisq wllchisq;
RUN;
********* Table 1, Age as continues;
proc surveymeans  data=c.newseedahmed_clean varmethod=taylor
all;
```

```
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg*ratioLLN_baseline;
var RIDAGEYR  ;
 ods output Statistics=MyStat;
run;

*************************************************************
*******;
******GRAPHS of Table 1, by baseline characteristics FEV1/FVC
ratio using LLN ;
ods html close;ods html;
ods graphics on;
PROC SURVEYFREQ data=c.newseedahmed_clean NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
TABLES SEG*age_cat*ratioLLN_baseline/plots = mosaicplot plots =
wtfreqplot risk or plots =(oddsratioplot relriskplot);
TABLES SEG*( RIDRETH1 RIAGENDR BMI_Cat DMDEDUC2 DMDMARTL
INDFMPIR_cat LBXVIDMS_cat LBXVD3MS_cat LBXVD2MS_cat CRP_Cat
DSD010AN pk_yrs_cat RIDEXMON )*
ratioLLN_baseline/plots = mosaicplot plots = wtfreqplot risk or
plots =(oddsratioplot relriskplot);
RUN;
ods graphics off;
**********
Continuous vitamin D by baseline FEV1/FVC;
proc surveymeans  data=c.newseedahmed_clean varmethod=taylor
all;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg*ratioLLN_baseline;
var LBXVIDMS  ;
 ods output Statistics=MyStat;
run;
**********
Continuous vitamin D3 by baseline FEV1/FVC;
proc surveymeans  data=c.newseedahmed_clean varmethod=taylor
all;
options fmtsearch=(c.newformat);
```

```
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg*ratioLLN_baseline;
var LBXVD3MS  ;
 ods output Statistics=MyStat;
run;
```

```
data newseedahmed_table2;
set c.newseedahmed_clean;
options fmtsearch=(c.newformat);
********* categorization of total 25OH vitamin D to two
categories using 30 nmol/L as a cutoff point;
********* The CDC has released total 25(OH)D based on 2011
Institute of Medicine (IOM)to:
< 30 (deficiency), 30-50 (WNL) , 50-125 (Above normal) , >125
(Toxic level);
If (  LBXVIDMS < 30) then LBXVIDMS_cat_by30nmolperL = 2;
If ( LBXVIDMS >= 30) then LBXVIDMS_cat_by30nmolperL = 1;
If (LBXVIDMS = .) then LBXVIDMS_cat_by30nmolperL = .;
label LBXVIDMS_cat_by30nmolperL ="Total 25OH Vitamin D levels
Categories using 30 nmol/L as a cutoff point ";
format LBXVIDMS_cat_by30nmolperL LBXVIDMS_cat_by30nmolperL.;
run;
ods html close;ods html;
* Table 2, Health demographics by Total 25OH Vitamin D levels
Categories using 30 nmol/L as a cutoff point ;
PROC SURVEYFREQ data=newseedahmed_table2 NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
TABLES SEG* (RIDRETH1 RIAGENDR BMI_Cat DMDEDUC2 DMDMARTL_cat
INDFMPIR_cat CRP_Cat DSD010AN pk_yrs_cat
RIDEXMON)*LBXVIDMS_cat_by30nmolperL/expected row col chisq
lrchisq wchisq wllchisq  ;
RUN;
*Table 2, Age_cat by Total 25OH Vitamin D levels Categories
using 30 nmol/L as a cutoff point ;
ods html close;ods html;
PROC SURVEYFREQ data=newseedahmed_table2 NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
```

```
TABLES age_cat*LBXVIDMS_cat_by30nmolperL/expected row col chisq
lrchisq wchisq wllchisq;
RUN;
**** Table2, Age as continues;
proc surveymeans  data=newseedahmed_table2 nomcar
varmethod=taylor all;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg*LBXVIDMS_cat_by30nmolperL;
var RIDAGEYR  ;
 ods output Statistics=MyStat;
run;
******;
Proc means  data=newseedahmed_table2;
options fmtsearch=(c.newformat);
class seg;
var LBXVIDMS;
run;

proc freq data=newseedahmed_table2;
options fmtsearch=(c.newformat);
tables  LBXVIDMS_cat_by30nmolperL*seg;
run;

********* Spirometry measurments by vitamin D;
proc surveymeans  data=newseedahmed_table2 varmethod=taylor all;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg*LBXVIDMS_cat_by30nmolperL;
var   Baseline_fev_ppd Baseline_fvc_ppd
Baseline_fevfvcratio_percent fevfvcprd_ref FevfevcLLN_ref ;
 ods output Statistics=MyStat;
run;
```

```
libname c "H:\Thesis\NHANES III\DATA";
run;
**** all variables adjusted by the season
The option LINK=GLOGIT option requests that the procedure fit a
generalized logit model;
********* Total Vitamin D, Model 1 include all variables and
interactions;
*collin macro;
OPTIONS MPRINT SYMBOLGEN mlogic;
```

```
* COLLINEARITY DIAGNOSTICS USING THE INFORMATION MATRIX;
* Original MACRO FROM SAS-L BY MATHEW ZACK;
* Modified 26 April 2005 by Jim Singleton to handle covariates
included in class statement;
* Modified November 2010 by Kristin Wall and Kevin Delaney to
     INCLUDE CODE FOR GENMOD, MIXED and GLIMMIX
       and
     Explicitly name the output dataset containing Collinearity
Diagnostics;

*The MACRO contains four named parameters:

     COVDSN=DATASETNAME is the input dataset containing the
Covariance Matrix output from
     the LOGISTIC, MIXED, GLIMMIX, PHREG or GENMOD procedures

     OUTPUT=DATASETNAME is the name you choose to contain the
output Collinearity Diagnostics

     PROCDR=SAS procedure that produced the collinearity Matrix
output.
     Currently, Valid values include:
     Logistic, Phreg, Genmod, Glimmix and Mixed
     To maintain consistency with previous versions of the
MACRO, this parameter is not required for Logistic or Phreg

     PARMINFO=Dataset generated with the Statement:
          ods output genmod.parminfo=parms
     That contains the names of the variables included in the
model;

*IN PROC LOGISTIC OR PROC PHREG SPECIFY THE COVOUT AND THE
OUTEST=DATASETNAME ;
*  OPTIONS IN THE PROC STATEMENT.
* IF USING LOGISTIC OR PHREG Only COVDSN and OUTPUT are
required;
*    %COLLIN(COVDSN=DATASETNAME, OUTPUT=DATASETNAME2);

*IF USING PROC LOGISTIC for CONDITIONAL LOGISTIC REGRESSION ie.
for Matched data you need to tell SAS to
     Drop the intercept column from the COVOUT DATASET: ;
     /*   proc logistic data=Data1 covout
outest=test(DROP=intercept);
          strata ID;
          model outcome(event='1')=Gall hyper;
          run;
```

```
          */
*Doing so will allow you to use the COVOUT DATASET for
conditional Logistic regression like any other Logistic
     output:
      %COLLIN(COVDSN=DATASETNAME, OUTPUT=DATASETNAME2); *In the
example above COVDSN=TEST;

*If using PROC SURVEYLOGISTIC add the following ODS OUTPUT
STATEMENT to your code;
 *ods output surveylogistic.covb=DATASETNAME;
*Also, add the /covb option to the MODEL statement, e.g. model
outcome=exp covars/covb;
*Then call the Macro as:

     %Collin(COVDSN=DATASETNAME,PROCDR=SURVEYLOGISTIC,OUTPUT=DAT
ASETNAME2);

* When using this Macro with GLIMMIX:
* Use the /covb option after the model statement and include the
line of code:  ods output glimmix.CovB=DATASETNAME;
* Call macro as: %COLLIN(COVDSN=DATASETNAME, PROCDR=GLIMMIX,
OUTPUT=);

*When using the Macro with GENMOD:  ;

* IF REPEATED IS NOT USED (UNCLUSTERED DATA -> NO GEE) THEN ;
* ADD COVB TO THE MODEL STATEMENT (MODEL / COVB) and include the
following two statements immediately
     Before  ;
*ods output genmod.parminfo=parms;
*ods output genmod.covb=covdsn;

* IF REPEATED IS USED FOR CLUSTERED DATA THEN ;
* ADD COVB TO THE REPEATED STATEMENT (REPEATED / COVB);
 *ods output genmod.parminfo=parms;
 *ods output genmod.geercov=covdsn;

* When using GENMOD Call the MACRO with PROCDR=GENMOD  and
PARMINFO=parms
          (from the ODS OUTPUT STATEMENT)
*    %COLLIN(COVDSN=COVDSN, PROCDR=GENMOD, PARMINFO=Parms,
OUTPUT=DATASETNAME);

* When using this Macro with MIXED:
* Use the /covb option after the model statement and include the
line of code:  ods output mixed.CovB=DATASETNAME;
```

```
* Call macro as: %COLLIN(COVDSN=DATASETNAME, PROCDR=MIXED,
OUTPUT=);



%MACRO COLLIN(COVDSN=, PROCDR=, PARMINFO=,OUTPUT=);
OPTIONS MPRINT SYMBOLGEN mlogic;
%* MACRO TO CALCULATE COLLINEARITY DIAGNOSTICS FROM ;
%*  VARIANCE-COVARIANCE MATRIX IN NONLINEAR REGRESSION;

%* REF: DAVIS CE, HYDE JE, BANGDIWALA SI, NELSON JJ.;
%*       AN EXAMPLE OF DEPENDENCIES AMONG VARIABLES IN A;
%*       CONDITIONAL LOGISTIC REGRESSION.  IN: MOOLGAVKAR SH,;
%*       PRENTICE RL, EDS.  MODERN STATISTICAL METHODS IN;
%*       CHRONIC DISEASE EPIDEMIOLOGY.  NEW YORK:;
%*       JOHN WILEY & SONS, INC., 1986:140-7.;

%let Drop=%str();

%* MAKE GENMOD COVARIANCE OUTPUT SIMILAR ENOUGH TO LOGISTIC AND
PHREG THAT THIS MACRO WILL
%* WORK.;

%IF %Upcase(&PROCDR)=GENMOD %THEN %DO;

%* FOR SOME INEXPLICABLE REASON, SAS DOES NOT RECORD THE
VARIABLE NAMES IN THE OUTPUT;
%* VARIANCE-COVARIANCE DATA SET. THIS NEXT SECTION OF CODE
REPLACES THE PARM VARIABLE;
%* WITH THE NAMES OF THE VARIABLES AND RENAMES PARM TO _NAME_ TO
CONFORM TO THE OUTPUT;
%* DATA SETS GENERATED BY LOGISTIC AND GENMOD.;

%* IF THERE ARE MORE THAN 9 VARIABLES IN THE MODEL STATEMENT,
SAS WILL STOP PROCESSING;
%* ON THE DATA NEXT_2 STEP DECLARING THE BY VARIABLE (PARM) IS
NOT IN THE CORRECT SORTED;
%* ORDER. THIS DOESNT HAPPEN FOR LESS THAN NINE VARIABLES. WHEN
YOU SORT THE DATA SET;
%* ON PARM, THE SORT DOES NOT TAKE PLACE AS EXPECTED, MESSING UP
THE VARIANCE-COVARIANCE;
%* MATRIX. THE PROBLEM IS THAT THE VALUES OF PARM PROGRESS AS
PARM1, PARM2, PARM3, ...;
%* PARM9, PARM10, ETC. WHEN YOU SORT ON PARM, PARM10, PARM11
THROUGH PARM19 SORT AFTER;
```

```
%* PARM1 AND BEFORE PARM2, DUE TO THE WAY SORTING WORKS ON
CHARACTER VARIABLES. THE ONLY;
%* WAY TO FIX THIS IS TO RENAME THE VARIABLES TO PARM01, PARM02,
ETC. SO THE SORTING WORKS;
%* CORRECTLY.;

DATA NEXT_1; SET &PARMINFO;
ATTRIB PARNUM FORMAT=$12.;
PARNUM=PARAMETER;
IF PARNUM = 'Prm1' THEN PARNUM = 'Prm01';
IF PARNUM = 'Prm2' THEN PARNUM = 'Prm02';
IF PARNUM = 'Prm3' THEN PARNUM = 'Prm03';
IF PARNUM = 'Prm4' THEN PARNUM = 'Prm04';
IF PARNUM = 'Prm5' THEN PARNUM = 'Prm05';
IF PARNUM = 'Prm6' THEN PARNUM = 'Prm06';
IF PARNUM = 'Prm7' THEN PARNUM = 'Prm07';
IF PARNUM = 'Prm8' THEN PARNUM = 'Prm08';
IF PARNUM = 'Prm9' THEN PARNUM = 'Prm09';

RENAME PARNUM=PARM;

RUN;
PROC SORT;
 BY PARM;
RUN;

DATA NEXT_1A; SET &COVDSN;
ATTRIB PARM FORMAT=$12.;
PARM=ROWNAME;
IF PARM = 'Prm1' THEN PARM = 'Prm01';
IF PARM = 'Prm2' THEN PARM = 'Prm02';
IF PARM = 'Prm3' THEN PARM = 'Prm03';
IF PARM = 'Prm4' THEN PARM = 'Prm04';
IF PARM = 'Prm5' THEN PARM = 'Prm05';
IF PARM = 'Prm6' THEN PARM = 'Prm06';
IF PARM = 'Prm7' THEN PARM = 'Prm07';
IF PARM = 'Prm8' THEN PARM = 'Prm08';
IF PARM = 'Prm9' THEN PARM = 'Prm09';

RUN;
PROC SORT;
 BY PARM;
RUN;

DATA NEXT_2(DROP=EFFECT); MERGE NEXT_1A(IN=IN1A) NEXT_1(IN=IN1);
BY PARM; IF IN1A;
PARM=EFFECT;
```

```
RENAME PARM=_NAME_;
RUN;

   %* IN SOME OUTPUT VARIANCE-COVARIANCE MATRICES, THERE WILL BE
A RECORD FOR;
   %* SCALE. DELETE THIS RECORD.;
   DATA NEXT_3; SET NEXT_2;
   IF _NAME_='SCALE' THEN DELETE;
   RUN;
   %* INSERT A DUMMY RECORD FOR ESTIMATE TO SIMULATE COVARIANCE
OUTPUT FROM LOGISTIC
   %*  AND PHREG.;
   DATA NEXT_4;
   _NAME_= 'ESTIMATE';
   OUTPUT;
   RUN;
   DATA NEXT_5; SET NEXT_4 NEXT_3;

   RUN;
proc print; run;

 %END;

%* MAKE MIXED COVARIANCE OUTPUT SIMILAR ENOUGH TO LOGISTIC AND
PHREG THAT THIS MACRO WILL WORK.;
%* Use the /covb option after the model statement and inlcude
the line of code:  ods output CovB=dataset1;
%* Call macro as: %COLLIN(COVDSN=, PROCDR=MIXED,
PARMINFO=dataset1);

%IF %upcase(&PROCDR)=MIXED %THEN %DO;
DATA NEXT_1 (Keep=_NAME_ col:); SET &COVDSN;
     RENAME EFFECT=_NAME_;
     if Row = 1     then  RowName = 'Prm01';
     if Row = 2     then  RowName = 'Prm02';
     if Row = 3     then  RowName = 'Prm03';
     if Row = 4     then  RowName = 'Prm04';
     if Row = 5     then  RowName = 'Prm05';
     if Row = 6     then  RowName = 'Prm06';
     if Row = 7     then  RowName = 'Prm07';
     if Row = 8     then  RowName = 'Prm08';
     if Row = 9     then  RowName = 'Prm09';
RUN;

data next_2(Drop=covars);
set next_1;
```

```
array cols col:;
covars=dim(cols);
call symput("Numcols",left(covars));
run;

data next_2a;
attrib dummy length=$1000;
retain dummy ;
set next_2;
if sum(of Col1-COL&Numcols)=0 then do;
if dummy=" " then dummy="Drop= COL"||trim(left(put(_N_,8.)));
else if dummy ne " " then do;

dummy=trim(left(dummy))||" COL"||trim(left(put(_N_,8.)));

end;
Call symput("Drop",dummy);
delete;
end;
run;

DATA NEXT_3; SET NEXT_2a(&drop);
   IF _NAME_='SCALE' THEN DELETE;
   RUN;

   %* INSERT A DUMMY RECORD FOR ESTIMATE TO SIMULATE COVARIANCE
OUTPUT FROM LOGISTIC
   %*  AND PHREG.;
   DATA NEXT_4;                ATTRIB _NAME_ FORMAT=$12.;

   _NAME_= 'ESTIMATE';
   OUTPUT;
   RUN;

   DATA NEXT_5; SET NEXT_4 NEXT_3; ATTRIB _NAME_ FORMAT=$12.;
   RUN;

proc print; run;

 %END;


%* MAKE GLIMMIX COVARIANCE OUTPUT SIMILAR ENOUGH TO LOGISTIC AND
PHREG THAT THIS MACRO WILL WORK.;

%IF %Upcase(&PROCDR)=GLIMMIX %THEN %DO;
DATA NEXT_1 (Keep=_NAME_ COL:); SET &COVDSN;
```

```
     RENAME EFFECT=_NAME_;
     if Row = 1     then  RowName = 'Prm01';
     if Row = 2     then  RowName = 'Prm02';
     if Row = 3     then  RowName = 'Prm03';
     if Row = 4     then  RowName = 'Prm04';
     if Row = 5     then  RowName = 'Prm05';
     if Row = 6     then  RowName = 'Prm06';
     if Row = 7     then  RowName = 'Prm07';
     if Row = 8     then  RowName = 'Prm08';
     if Row = 9     then  RowName = 'Prm09';
RUN;

data next_2(Drop=covars);
set next_1;

array cols col:;
covars=dim(cols);
call symput("Numcols",left(covars));
run;

data next_2a;
attrib dummy length=$1000;
retain dummy ;
set next_2;
if sum(of Col1-COL&Numcols)=0 then do;
if dummy=" " then dummy="Drop= COL"||trim(left(put(_N_,8.)));
else if dummy ne " " then do;

dummy=trim(left(dummy))||" COL"||trim(left(put(_N_,8.)));
end;
Call symput("Drop",dummy);
delete;
end;
run;

*Depending on the reference coding used in GENMOD and MIXED the
Covariance MATRIX output by the procedure may have
     Columns and corresponding Rows with all Zeros.
The MACRO Variable DROP (created in NEXT_2A) Isolates and
removes these extraneous columns before we get to
Manipulating the matrix in the IML code below;

DATA NEXT_3; SET NEXT_2a(&drop);
   IF _NAME_='SCALE' THEN DELETE;
   RUN;
```

```
   %* INSERT A DUMMY RECORD FOR ESTIMATE TO SIMULATE COVARIANCE
OUTPUT FROM LOGISTIC
   %*  AND PHREG.;
   DATA NEXT_4;                ATTRIB _NAME_ FORMAT=$12.;

   _NAME_= 'ESTIMATE';
   OUTPUT;
   RUN;

   DATA NEXT_5; SET NEXT_4 NEXT_3; ATTRIB _NAME_ FORMAT=$12.;
   RUN;

proc print; run;

%END;

%* MAKE SURVEYLOGISTIC COVARIANCE OUTPUT SIMILAR ENOUGH TO
LOGISTIC AND PHREG THAT THIS MACRO WILL WORK.;

%IF %Upcase(&PROCDR)=SURVEYLOGISTIC %THEN %DO;
DATA NEXT_1 ; SET &covdsn;
     RENAME Parameter=_NAME_;

RUN;

DATA NEXT_3; SET NEXT_1;
   IF _NAME_='SCALE' THEN DELETE;
   RUN;

   %* INSERT A DUMMY RECORD FOR ESTIMATE TO SIMULATE COVARIANCE
OUTPUT FROM LOGISTIC
   %*  AND PHREG.;
   DATA NEXT_4;     ATTRIB _NAME_ FORMAT=$32.;

   _NAME_= 'ESTIMATE';
   OUTPUT;
   RUN;

   DATA NEXT_5; SET NEXT_4 NEXT_3; ATTRIB _NAME_ FORMAT=$32.;
   RUN;

proc print; run;

%END;

%IF &PROCDR=%str()
     or %upcase(&PROCDR)=LOGISTIC
```

```
     or %upcase(&PROCDR)=PHREG
%THEN %DO;
   DATA NEXT_5; SET &COVDSN;
   RUN;

%END;

proc print data=next_5; run;

%IF (NEXT_5 NE ) %THEN %DO;

OPTION MPRINT;

%LET __STOP=0;

PROC IML;
  USE NEXT_5;
  READ ALL VAR {_NAME_} INTO _VARNAME;

  _NRVNAME=NROW(_VARNAME);

  IF (_NRVNAME>1) THEN DO;
     _VARNAM2=_VARNAME(|2:_NRVNAME, |);
     NMISSING=J(NROW(_VARNAM2),1,.);
     LABELS={"EIGENVAL","CONDINDX","         "};
     _VARNAM2=LABELS//_VARNAM2;
     FREE _VARNAME LABELS;
     READ ALL VAR _NUM_ INTO VARCOV(|COLNAME=_NVNAME|);
     _NRCVC=NCOL(VARCOV);
     LASTVNAM=_NVNAME(|1,_NRCVC|);
     IF (LASTVNAM="_LNLIKE_") THEN
VARCOV2=VARCOV(|2:_NRVNAME,1:_NRCVC-1|);
     IF (LASTVNAM^="_LNLIKE_") THEN
VARCOV2=VARCOV(|2:_NRVNAME,|);

%* IF COVARIANCE MATRIX IS FROM PROC GENMOD USING THE REPEATED
MEASURES DESIGN;
%* THEN THE LOWER DIAGONAL WILL HAVE THE CORRELATIONS AND THE
UPPER DIAGONAL WILL HAVE;
%* THE COVARIANCES. THIS NEXT SECTION OF CODE REPLACES THE LOWER
DIAGONAL WITH THE UPPER;
%* DIAGONAL TO MAKE A SYMMETRIC COVARIANCE MATRIX. IF THE MATRIX
IS SYMMETRICAL ALREADY;
%* THEN THE NEXT SECTION OF CODE WILL NOT AFFECT ANYTHING.;

        VC2_C = NCOL(VARCOV2);
        VC2_R = NROW(VARCOV2);
```

```
         DO CL=1 TO VC2_C;
            DO RW=1 TO VC2_R;
               VARCOV2(|RW,CL|) = VARCOV2(|CL,RW|);
            END;
         END;

%* PRINT THE VARIANCE-COVARIANCE MATRIX FOR DIAGNOSTIC PURPOSES;
       PRINT VARCOV2;

      FREE VARCOV _NRCVC LASTVNAM VC2_C VC2_R CL;
      COVBINV=INV(VARCOV2);
      SCALE=INV(SQRT(DIAG(COVBINV)));
      R=SCALE*COVBINV*SCALE;
      FREE COVBINV SCALE;
      CALL EIGEN(MUSQR,V,R);
      FREE R;
      SROOTMUS=SQRT(MUSQR);
      CI=1/(SROOTMUS/MAX(SROOTMUS));
      PHI=(V##2)*DIAG(MUSQR##(-1));
      SUMPHI=PHI(|,+|);
      PI=PHI#(SUMPHI##(-1));
      FREE PHI SUMPHI SROOTMUS V;
      FINAL=(MUSQR||CI||NMISSING||PI`)`;
      FREE PI MUSQR CI NMISSING;
      _NCFINAL=NCOL(FINAL);
      _NRFINAL=NROW(FINAL);
      FINAL2=J(_NRFINAL,_NCFINAL,0);
      _NCFP1=_NCFINAL+1;
      __VDP="VDP";
      DO I=1 TO _NCFINAL;
         FINAL2(|,_NCFP1-I|)=FINAL(|,I|);
         X=CHAR(I,3);
         Y=COMPRESS(CONCAT(__VDP,X));
         IF I=1 THEN _VDPNAME=Y;
            ELSE _VDPNAME=_VDPNAME||Y;
      END;
      FREE FINAL _NRFINAL _NCFINAL I X Y;
      CREATE &output FROM FINAL2(|ROWNAME=_VARNAM2
COLNAME=_VDPNAME|);
      APPEND FROM FINAL2(|ROWNAME=_VARNAM2|);
      FREE _VARNAM2 _VDPNAME FINAL2;
   END;
   IF (_NRVNAME=1) THEN DO;
      X="1";
      CALL SYMPUT("__STOP",LEFT(X));
      PRINT " ";
```

```
     PRINT
"*********************************************************";
     PRINT "YOU NEED TO SPECIFY THE  COVOUT  OPTION";
     PRINT " IN EITHER PROC LOGISTIC OR PROC PHREG.";
     PRINT " THIS PROGRAM WILL NOT CALCULATE COLLINEARITY
DIAGNOSTICS.";
     PRINT
"*********************************************************";
     PRINT " ";
  END;
  QUIT;
RUN;

%IF (&__STOP EQ 0) %THEN %DO;
   PROC PRINT DATA=&output LABEL NOOBS;
     ID _VARNAM2;
      Title7 "Input DATASET &COVDSN, Submitted &sysdate9";
     TITLE8 "COLLINEARITY DIAGNOSTICS FOR NONLINEAR MODELS
USING";
     TITLE9 "THE INFORMATION MATRIX:  EIGENVALUES, CONDITION
INDEXES,";
     TITLE10 "AND VARIANCE DECOMPOSITION PROPORTIONS (VDP'S)";

     LABEL _VARNAM2="VARIABLE";
   RUN;
%END;

%END;
%ELSE %DO;
   %PUT;
   %PUT
"*********************************************************";
   %PUT "WHEN YOU INVOKE THIS MACRO, YOU HAVE TO SPECIFY THE
NAME";
   %PUT " OF A SAS DATA SET THAT CONTAINS THE VARIANCE-
COVARIANCE";
   %PUT " MATRIX FROM EITHER PROC LOGISTIC OR PROC PHREG.";
   %PUT;
   %PUT "YOU CAN CREATE THIS MATRIX BY INCLUDING THE FOLLOWING
OPTIONS";
   %PUT " ON THE PROC STATEMENT:  COVOUT  AND  OUTEST=SASDSN,";
   %PUT " WHERE SASDSN IS THE NAME OF THE SAS DATA SET
CONTAINING";
   %PUT " THE VARIANCE-COVARIANCE MATRIX.";
   %PUT
"*********************************************************";
   %PUT;
```

```
%END;

PROC DATASETS;
DELETE NEXT_1 NEXT_1A NEXT_2 Next_2a NEXT_3 NEXT_4 NEXT_5;
RUN;
QUIT;

title;


%MEND COLLIN;
%INCLUDE collin;
ODS OUTPUT SURVEYLOGISTIC.COVB=collin_info;
data newseedahmed4;
set c.newseedahmed_clean;
options fmtsearch=(c.newformat);
******* create new continues vitamin D and D3 variebles to
include in the modeling processing
to divide by 25 in order to create a new continuous variable,
for which a change in 1 unit will
equal to a change in 25 nmol/L of vitamin D;
VitaminD_25_continuous= LBXVIDMS/25;
VitaminD3_25_continuous= LBXVD3MS/25;
label VitaminD_25_continuous ="new continuous vitamin D/25";
label VitaminD3_25_continuous ="new continuous vitamin D3/25";
run;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=  VitaminD_25_continuous  RIDAGEYR
BMXBMI  RIDRETH1  RIAGENDR   LBXCRP   pk_yrs
RIDEXMON/covb vadjust=none ;
          run;
%collin(COVDSN=collin_info,PROCDR=SURVEYLOGISTIC,OUTPUT=collin_i
nfo2)
```

```
data newseedahmed4;
set c.newseedahmed_clean;
options fmtsearch=(c.newformat);
******* create new continues vitamin D and D3 variebles to
include in the modeling processing
to divide by 25 in order to create a new continuous variable,
for which a change in 1 unit will
equal to a change in 25 nmol/L of vitamin D;
VitaminD_25_continuous= LBXVIDMS/25;
VitaminD3_25_continuous= LBXVD3MS/25;
label VitaminD_25_continuous ="new continuous vitamin D/25";
label VitaminD3_25_continuous ="new continuous vitamin D3/25";
run;
*******Step 2 for evaluation of the Final Model for different
Vitamin Ds, Interaction assessment after I re-categorize BMI and
smoking to avoid high number of parameters>number of clusters
that can lead to inability to calculate F-statistics;
**********************************************
Under full-rank parameterizations, Type 3 effect tests are
replaced by joint tests. The
      joint test for an effect is a test that all the parameters
associated with that effect
      are zero. Such joint tests might not be equivalent to Type
3 effect tests under GLM
      parameterization
**********************************************
**********************************************
Step 1:  asses the significant of interaction terms for VitD As
NEW CONTINOUS variables
###1: with age;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*RIDAGEYR  /link=glogit expb CLPARM;
run;
**** ###2: with Gender;
```

```
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*RIAGENDR /link=glogit expb CLPARM;
run;
**** ###3: with Race;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*RIDRETH1/link=glogit expb CLPARM;
run;
**** ###4: with BMI;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*BMXBMI/link=glogit expb CLPARM;
run;
```

```
**** ###5: with BMI;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*LBXCRP /link=glogit expb CLPARM;
run;
**** ###6: with smoking;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*pk_yrs/link=glogit expb CLPARM;
run;
**** ###6: with Season;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD_25_continuous  RIDAGEYR
RIAGENDR RIDRETH1  BMXBMI   LBXCRP   pk_yrs RIDEXMON
VitaminD_25_continuous*RIDEXMON /link=glogit expb CLPARM;
```

```
run;
************
************
************
****** Assess for confounders
a Priori model for Total Vitamin D;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=  VitaminD_25_continuous  RIDAGEYR
RIDRETH1  RIAGENDR BMXBMI   LBXCRP  pk_yrs RIDEXMON
/link=glogit expb  clparm;
run;
****** Crude model for Total Vitamin D;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=  VitaminD_25_continuous    /link=glogit
expb  clparm;
run;
****** a Priori model for Vitamin D3;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
```

```
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD3_25_continuous     RIDAGEYR
RIDRETH1  RIAGENDR BMXBMI   LBXCRP  pk_yrs RIDEXMON /link=glogit
expb CLPARM;
run;
**** Crude model for vitamin D3;
proc surveylogistic data=newseedahmed4 nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline  = VitaminD3_25_continuous /link=glogit
expb CLPARM;
run;
*********** a priori model Vit d as categorical;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
LBXVIDMS_cat (ref= "30 - 74= (inadequate/adequate =Ref)"
param=ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=   LBXVIDMS_cat RIDAGEYR  RIDRETH1
RIAGENDR BMXBMI   LBXCRP  pk_yrs RIDEXMON /link=glogit expb
clparm;
run;
****** Crude model for Total Vitamin D as categorical ;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
```

```
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
LBXVIDMS_cat (ref= "30 - 74= (inadequate/adequate =Ref)"
param=ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=   LBXVIDMS_cat   /link=glogit expb
clparm;
run;
********** a priori model: Vit D3 as categorical;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
 LBXVD3MS_cat  (ref= " 30 - 74 = ((inadequate/adequate=Ref)"
param=ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=   LBXVD3MS_cat RIDAGEYR  RIDRETH1
RIAGENDR BMXBMI   LBXCRP  pk_yrs RIDEXMON /link=glogit expb
clparm;
run;
********** crude model: Vit D3 as categorical;
proc surveylogistic data=newseedahmed4  nomcar VARMETHOD=TAYLOR
;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
Domain seg;
class
RIDRETH1  (ref = "Non-Hispanic White--- Ref"  param = ref)
RIAGENDR   (ref = "male---- Ref" param = ref)
 LBXVD3MS_cat  (ref= " 30 - 74 = ((inadequate/adequate=Ref)"
param=ref)
RIDEXMON (ref = "May 1 through October 31 = Ref" param = ref);
model ratioLLN_baseline=   LBXVD3MS_cat   /link=glogit expb
clparm;
run;
```

```
**** Table 4,
GOLD stages by Total Vitamin D ;
ods graphics on;
proc surveyfreq data=c.newseedahmed_clean  NOMCAR VARMETHOD =
TAYLOR ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
table SEG*LBXVIDMS_cat*GOLD/ CL (TYPE = WILSON) DEFF expected
row col chisq lrchisq wchisq wllchisq;
run;
ods graphics off;

**** GOLD stages by Vitamin D 3;
ods graphics on;
prosurveyfreq data=c.newseedahmed_clean  NOMCAR VARMETHOD =
TAYLOR  ;
options fmtsearch=(c.newformat);
stratum SDMVSTRA;
cluster SDMVPSU;
weight WTMEC4YR;
table SEG*LBXVD3MS_cat*GOLD/ CL (TYPE = WILSON) DEFF expected
row col chisq lrchisq wchisq wllchisq plots = mosaicplot plots =
wtfreqplot risk or plots =(oddsratioplot relriskplot);
run;
ods graphics off;
```