ASSOCIATION BETWEEN HIGH CAFFEINE CONSUMPTION AND LOWERED RENAL FUNCTION AMONG NORMENSIVE ADULTS IN THE UNITED STATES

A Thesis
Submitted to
the Temple University Graduate School

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Epidemiology

---

by
Hiroshi James Palomares Inuzuka
Diploma Date August 2021

Examining Committee Members:

Robin Taylor Wilson, Advisor Chair, Department of Epidemiology and Biostatistcs
Jingwei Wu, Department of Epidemiology and Biostatistcs
Jennifer Orlet Fisher, Department of Social and Behavioral Sciences, College of Public Health

ABSTRACT

This cross-sectional study aims to evaluate the association of caffeine intake with renal function among adults between ages 18-55. Participants of the National Health and Nutritional Examination Survey (NHANES) survey for the three consecutive years (2013-2014, 2015-2016, and 2017-2018) were used. A weighted multivariable linear regression analysis of the caffeine concentration was conducted. Greater intake was associated with lowered renal function. This association persisted when limiting the daily caffeine intake to 2000 mg/day or less. Among younger adults, ages 18 to 39 the beta coefficient was about 50 percent larger than the beta coefficient for individuals ages 40 to 55. This suggests that caffeine intake may have a greater impact on renal function among younger adults.

While greater caffeine intake was associated with reduced renal function in this cross-sectional study, further investigation such as an experimental study should be performed to confirm the findings of this thesis.

This thesis is dedicated to my mom and dad. For all the pain through highschool and undergraduate studies, thank you for being there.

ACKNOWLEDGMENTS

I would like to acknowledge my thesis chair and committee members for their infinite patience going through revision after revision, with minimal amount of time. It was a painful process of trial and error but their faith, especially from my thesis chair, in my abilities helped me push beyond what I believed I could do. My committee members, who were understanding in the first semester and helped answer my questions for data processing, feedback, and guidance moving forward. Thank you for your support through these interesting times.

Secondly, I would like to acknowledge my parents and my baby sister, who despite the 2020 COVID pandemic gave their emotional support through it all. The year has been unexpected, chaotic, and pushed people in ways we would've never imagined. It is in these days we appreciate what epidemiology has done and could do.

Finally, I would also like to acknowledge my therapy counseler who helped me find my steps again. This thesis would not be possible with declining mental health.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Background

As caffeine's prevalence in culture grows, studying its possible effects grows as well. Coffee, tea, chocolate products, soft and energy drinks are the primary sources of caffeine. The molecule is typically known for it's stimulant effects, increasing alertness, and enhancing mood quickly following consumption due to its rapid absorption into the bloodstream (Cappelletti, Piacentino, Sani, and Aromatario, 2015). A previously completed survey done by the National Health and Nutrition Examination Survey (NHANES) revealed that coffee consumption accounted for 75% of the United States population who consume caffeine, ages at and above 20, between 2003 to 2012. Tea consumption accounted for 20.8% of the United States population, ages at and above 20, between 2011 to 2016 (Rehm, Ratliff, Riedt, and Drewnowski, 2020).

However, caffeine's lesser-known aspect is the diuretic effect, reducing the amount of water reabsorbed into the body following renal filtration, thus increasing the urine expelled from the body. In doing so, it increases the number of unfiltered ions such as sodium and calcium to be unabsorbed (Peerapen and Thongboonkerd, 2018). The water retention effect and its possible consequences of caffeine are not completely understood. One such study (Fenton et al, 2015), attempted to identify a correlation between the exchanger isoform known as NHE3, which inhibits the Na+/H+ ion channels controlling of water reabsorption in mice. While the study concluded that caffeine was independent of its natural diuretic effect, it was noted that NHE3 was not the only form

for water reabsorption. Further studies attempted to understand why caffeine has this diuretic effect with one revolving around cultivated renal cells. This study investigated if kidney stone formation was possible due to the increased ion counts in urine following caffeine consumption and absorption (Peerapen and Thongboonkerd, 2016). While the study concluded with ionized crystals condensing around the renal cells, the researchers remained unconvinced due to the lack of a biological system filtering through the renal cells. These contributing factors, the water reabsorption loss and build up for kidney stones, illustrate a potential loss of renal function which may be linked with an increased consumption of caffeine.

One common trait with tracing caffeine correlation and renal function was the average age of fifty to sixty years, the ages at which lowered renal function begins to manifest. This study aids in a data gap for United States populations on the effects of caffeine. With increased awareness into a potential correlation between caffeine and lowered renal function, this thesis aims to act as a stringboard for better suited studies to investigate these effects. With a lack of data for the younger generation and under these known factors, further observations are necessary to understand a possible correlation between caffeine and lowered renal function.

Epidemiological Statistics to Lowered Renal Function

The major disease that contributes to lowered renal function is Chronic Kidney Disease (CKD). In the United States, roughly 15% or approximately 37 million people, are affected by this disease yearly. In most cases, adults with CKD would not know if they have it due to the lack of screening and detectable symptoms until the later stages of

the disease. Common with age groups above sixty-five years, lesser cases occur between age groups between forty-five and sixty-four, and eighteen and forty-four years old (Centers for Disease Control and Prevention, 2019). Resources are typically allocated towards the more vulnerable age group. According to the data charts from *Healthy People 2020* (Healthy People 2020, 2020), as well as their recorded data from previous years, CKD and general renal function cases increased over the last twenty years from 1% to 15% before a steady between the years of 2001 to 2012, between 14.8% and 14.6%. According to the World Health Organization estimations (Webster, Nagler, Morton, and Masson, 2017), chronic kidney disease is attributed for 1.5% of worldwide deaths. This marks CKD as the 14th leading cause of death, however the Global Health Observatory states that the death rate for CKD will increase. Despite these statistics, kidney disease's surveillance systems are neglected in favor for diseases such as cancers, cardiovascular diseases, and diabetes (Crews, Bello, and Saadi, 2019).

Increased Risk Factors to Reduced Renal Function

With the projected increase in CKD in the world, what propagates the disease depends on various factors. The most known risk factors to CKD and general lowered renal function are diabetes, obesity, and high blood pressure (CDC, 2019). Factors that are not biological may include environmental and sociodemographic factors, such as the availability of fast-foods in urban environments, the inability of obtaining healthier options due to prices, and dietary foods that are linked to specific cultures. According to one study (Crews, Bello, and Saadi, 2019), this becomes more apparent in higher-income countries, splitting between racial factors and socioeconomic factors. For the United

States, African Americans are at a higher risk of chronic kidney disease. Low-income persons are also at a higher risk of acute/chronic kidney injuries due to lifestyle choices and compromises, caused by economic barriers. These barriers may be disallowing access to healthier foods, alleviations to balance their unhealthy life-choices. These choices may include undernutrition or overnutrition, physical inactivity, and stress. Family genetics also play a role in contracting lowered renal function. Susceptibilities to kidney injury and the body's inability to recover from prolonged kidney damages all contribute to chronic kidney disease possibility being a genetic disease however it is noted that increased incidence and prevalence over this short time may also illustrate environmental factors as well (Webster, Nagler, Morton, and Masson, 2017). This trend follows with hypertension, causing increased strain on the glomerulus, causing a breakdown in the endothelial cells that are used in the Bowman's capsules.

Susceptible Populations to Lowered Renal Function

However, specific populations such as Mexican and Black American show an increase in chronic kidney disease cases, from 13.9% to 16.4%. Racial disparities are greater between these populations in comparison to their white counterparts, who did not show signs of improvement or decline. Other disparities such as people between 18 to 24 years old increasing in CKD prevalence from 5.6% to 7.4%, and 25 to 44 years old increased in CKD prevalence from 5.5% to 6.5%. These increasing trends are despite overall CKD prevalence remaining at 15% (Healthy People 2020, 2020). While this can be attributed to a lack of surveillance in these lower income regions, it should be noted that these regions have varying sources how they contract chronic kidney disease. For

Asia, India, and sub-Saharan Africa, the disease is attributed to glomerulonephritis, or the damaging of kidney tissues that block filtration. Despite lower income populations being at higher risk at contracting chronic kidney disease, countries with higher incomes suffer higher percentages of chronic kidney disease. For high-income countries, it is associated with worsening blood pressure (Webster, Nagler, Morton, and Masson, 2017).

As these populations grow older, there is the possibility that chronic kidney disease cases will rise and aggravate the response. This is caused by blood pressure and blood sugar reducing the effectiveness of the glomerulus, leading to deficient waste filtration and water retention. It is also possible that caffeine may be acting with other exposures to increase the risk of low renal function.

Known Effects of Caffeine Concentration in Cellular Biology

At high concentrations, 8-10 mg/L in the blood plasma, caffeine is metabolized in the liver to become metabolite byproducts while others will be excreted through urine. The overlooked issue that arises is then, an unknown limit for how much caffeine can be concentrated in the human body. Renal cell exposure to the caffeine has noted the diuretic effect of the molecule, pull water out of the cells along with a higher concentration of calcium and sodium ions, components of kidney stone formation. Kidney stones by themselves can be a marker potential kidney functionality. A meta-analysis on three cohort studies (Peerapen and Thongboonkerd, 2018), noted the possible effects of kidney stone formation in conjunction with caffeine drinks. While two of the studies stated that the kidney stone formation was reduced in prevalence, it was noted that the studies had restrictions in dietary oxalate, a primary component in caffeine and

kidney stones. The final study yielded similar results, especially in cases between caffeinated and decaffeinated coffees, however it was also stated that coffee or tea may have some bioactive components that act as protective agent against caffeine. In a previous study by the same authors (Peerapen and Thongboonkerd, 2016), a different form of renal function deficiency was observed. A possible link between the kidney stone formation compound known as calcium oxalate monohydrate (COM) was investigated for COM crystal binding proteins to aid their formation. The crystallization was generated through an oversaturated concentration of calcium and oxalate compounds in the blood fluid being deposited through urine in the bladder. The study noted this molecule compound was influenced by the blood flow rate, temperature, pH, and presence of other organic/inorganic components in the glomerulus. While the data supported the decrease of COM crystals, the size of said crystals increased. These crystals would form a "kidney stone" like formation, much like how the human body creates kidney stones. Kidney stones themselves reveal potential decreases in renal function, as the kidney reduces its ability to retain water and reabsorb the ions required to form kidney stones. That stated, while the study concluded with caffeine's commonly known biological effects, it had not been observed in the human body with complete functions of waste filtration, increased blood pressure, and water retention.

Epidemiologic Studies of Caffeine Intake and Renal Function

Caffeine through direct cellular studies highlight one point of the molecule. When caffeine is interfaced with people, other results manifest. A meta-analysis data taken from populations in Italy, Japan, and Korea stated their studies having the subject populations

between fifty and sixty years of age (Wijarnpreecha et al., 2017). While the study stated there may possibly be a link, it also stated that the source of the caffeine may include some interactions that could interfere with its effects. Another meta-analysis with a different population group for kidney stones formation based on caffeine illustrated similar results (Ferraro, Taylor, Gambaro, and Curhan, 2014). Given these two meta-analyses, it follows the general risk factors with chronic kidney disease and lowered renal function. While increased age has been known as an unmodifiable risk factor for lowered renal function as well as chronic kidney disease, the unknown protracted effects of caffeine may do more harm in the long-term following increased sources of caffeine and in different concentrations.

Additionally, there is the possibility that the effects of caffeine may also depend on co-exposures and in the way caffeine is consumed in the diet. As stated previously, coffee, tea, sodas, and energy drinks are some of the possible sources. These items are typically sweetened or artificially produced with other chemicals and dietary factors. High concentrations of sugar, and sodium, may influence a regularity protein for urination known as protein kinase C (Hypolite, and Malykhina, 2015). This protein serves as a biological signal, for smooth muscle contractions and bladder control. As it is influenced by calcium ion channels, molecules such as sugar and salt may influence its ability to function properly.

Study Purpose

The purpose of this study is to determine the correlation between high caffeine intake and lowered renal function in young adults.

CHAPTER 2

METHODS

Target Population and Study Population

The NHANES study is a national survey of individuals conducted on a continuous yearly basis across all states with approximately 8000 randomly selected residents in the United States and oversampling among Hispanic, Non-Hispanic Black, and Asian populations*. Additional oversampling was done in Non-Hispanic white and other persons at or below the 185 percent poverty guidelines as described by the Department of Health and Human Services (HHS) as well as persons ages 80 and older. The target study population for this investigation was aimed at young adults, between the ages eighteen to fifty-five. All data used in this thesis publically available and were obtained through the NHANES website. The provided datasets, with explanations on what and how the data was gathered in the sections below, were categorized by how this thesis used them.

NHANES survey eligibility was based on participants age and sex during the time of the survey in the mobile examination centers (MEC). Patients with known risk factors such as diabetes, special diets, reported kidney treatments, extreme obesity, and alcoholic intakes were excluded from the study. NHANES survey data for the years 2013-2014, 2015-2016, and 2017-2018 were chosen, being the most recent surveys.

The survey methods in collecting, assessing, and having the person come into the MEC has not changed through these three yearly cycles. The documentation instructions are provided to each selected survey interviewer on how to approach a potential candidate, based on location where the survey interviewer is scouting for survey takers.

Initial steps in sampling were done by identifying a dwelling unit (DU). This may include a single house, flats/apartments, mobile homes, seasonal dwellings, and housing on military bases. In order to identify these DU, segments were drawn by the survey giver. Once these segments were set, addresses were identified and checked if it could be identified as a DU. A listing sheet was created through an automated system to start their initial survey location. Each segment was assigned a stand number, segment number, and serial number. These numbers make up a Household ID, with the first three numbers indicating the stand number, identifying the area. The next one or two digits were the segment number, identifying the segment within the stand. The last digits were designated an unique DU in the sample.

Once the DU are identified, the survey giver arrives at the selected DU, with a brief summary of what they are accomplishing and whether or not the person at the DU wishes to take part in the survey. Several key items need to be addressed to the survey taker such as the purpose of the study, and confidentiality of the data collected. Survey takers may have additional questions that can be answered, however the responses must not be elaborated unless asked further. However, the eligibility of a person to respond to the survey giver must be at least 18 years of age or an emancipated minor. A written consent form is provided to the participants prior to continuing the interview. If the person declines, or there is no sample persons (SP) to enroll in the study, the interviewing process is marked as completed and no further interviews would be taken place at that DU.

The number of people in the household, as well as their ethnicity, gender, race, age, and income, were recorded by the survey interviewer on the first visit. Contact

information, and the relationship with the people also living in the DU were required for the continuation of the study. All items are questioned through a computer-assisted personal interviewing (CAPI) device. Following the initial interview, the SP was assigned a two digit MEC exam appointment scheduling code for setting up an appointment at the closet MEC. Age was measured as a continuous variable, however in order to determine differentiations in the thesis age range, it would be split between 18 to 40, and 40 to 55. The reasoning behind these differentiations is due to the biology of lowered eGFR values closely associated with older ages. Unless a person already has reduced renal functions below the age of 45, it is unexpected for said person to not have some form of kidney injury. This is further explained in the Exclusion Criteria section of this chapter.

The appointment interview at the MEC done at specific locations occupied by four trailers equipped and designed for the MEC testing. Physical measurements, such as weight, height, bone density, and blood work were done there. Questionnaire styled testing included dietary interviews, dental examination, and liver ultrasound. How these tests were performed in order to obtain specific parameters being used in the model and thesis are explained in the relevant sections below.

Lastly, a follow-up third interview was completed through the telephone. Named the Phone Follow-Up (PFU) interview, they were scheduled to three to ten days following the MEC interview and completed in a single phone-call unless there are interruptions. There are three sections for the PFU, including a second dietary recall, supplementary and antacid section, and the Post-Recall section. The system used for this phone interview was developed by Westat under contract from the Agricultural Research

Service (ARS) of the USDA. It should be noted that this particular interview is specifically meant to collect data on supplements, and antacids. Measurements used were done by a set of measuring tools, given through a Food Model Booklet (FMB) that have shapes and pictures of food sizes and meaasurements. Dietary behaviors such as participation in food assistance programs, type of water supply at home, history of weight, height, and anemia, and lifestyle habits were also gathered from this interview.

Quality control of the survey giver is done through field observations, field editing, field office review of cases for errors and discrepancies, and validation. Other methods for quality control include audio recordings of the sessions between survey giver and SP and key data item quality reviews. The recordings are checked over by supervisors to make sure that the interview is following protocols (National Health and Nutrition Examination Survey, 2017).

Exposure Measures – Caffeine Intake

Caffeine measurements were completed at the MEC, using a program by the United States Department of Agriculture (USDA) Automated Multiple Pass Method (AMPM) called “Blaise Data Entry”. That stated, there are several points to note that are specific for the predictor parameter, caffeine. Brand named products have information pertaining to their nutritional content and were encouraged by the survey input to recall if they had taken foods that were recognizable through said brand names. Measurements for cups and/or glasses was standardized based on models created by Anchor Hocking, Crate & Barrel, Food Wares, Carlisle, and Kroger. The amount of fluid was also standardized at scaled levels, unless specifically input into the system as brand name items with marketed

sizes. The scales for those models, what the models looked like, and how much would be a specifically filled levels are taken directly from the Measuring Guides for Dietary Recall Interview, 2002+ (CDC, 2015). Caffeine intake was estimated in milligrams per day, extrapolated from the various food items. This program and its functions are explained in further detail in the section Exposures, Characteristics, and Comorbidities.

Outcome Measures – Renal Function

A single serum creatinine measurement was taken during fasting for at least 12 hours to 16 hours to calculate estimated glomerular filtration rate (eGFR). During a fasting blood draw, a single 7 mL red tube was collected through venipuncture. Tubes were stored as quickly as possible to be refrigerated. Tubes were processed with a Beckman Coulter Allegra X-15R centrifuge at 2900 rpm for 15 minutes at 4-8°C. Serum creatinine was measured in mg/dL. The substance was generated through serum converted by reacting creatinine to form creatine, extracting sarcosine then reacted with hydrogen peroxide and then measured by a mass spectrometer. Serum creatinine was extracted into 2-mL vessels from the blood sample, labelled as ‘Biochem’.

These tests were done by the University of Minnesota, Advanced Research and Diagnostic Laboratory (ARDL). All laboratories are staffed with three American Society of Clinical Pathologists-certified medical technologists or laboratory technicians. Laboratory interference, where the calculation of the concentration would be off without calibration issues, would not occur up to 524 mg/L of creatinine levels or hemolyzed blood with HbF antibodies greater than 600 mg/dL. The standard limit of detection for serum creatinine was 0.10 mg/dL. Out of range results may be triggered by low-end or

high-end values limits were retested. High-limit values were diluted with saline and adjusted accordingly. Laboratory staff were blinded, as qualified by a hematology 5C survey challenge. The machine used for these measures was a Roche Cobas 6000 Chemistry Analyzer.

Quality of the testing and equipment was completed with a Roche Calibrator for Automated Systems, catalog #10759350190. Frequency calibration was performed on water, completing a two-point calibration before allowing the machines to be used, with validation every 6 months or after major maintenance/service procedures. A quality assurance log was created and maintained by laboratory staff (NHANES, 2020).

Estimated GFR (eGFR) is a continuous numerical value in milliliters per minute by 1.73 meters squared ($mL/min/1.73m^2$). The calculation of eGFR through the CKD-EPI formula is dependent on patient characteristics including gender, age, and ethnicity. This formula was chosen over the MDRD equation due to CKD-EPI providing accurate eGFR values for patients who's unknown if they already have lowered renal function (Matsushita et at, 2012). The equation for CKD-EPI is: $141 * [\min(S_{Cr}/K_1)^{\alpha}$ or $\max(S_{Cr}/K_1)^{-1.209}] * 0.993^{Age} * 1.018$ [if female] * 1.159 [if Black]; where S_{Cr} is the serum creatinine measure in mg/dL and K_1 is a gender variable, assigned 0.7 for females and 0.9 for males. Thus, female, and black individuals would have a higher eGFR value following calculation. In addition, for calculating if the person was considered 'Black' or not, individuals self-reported as "multi-racial" were classified as 'Non-Black'. Hispanic and non-Hispanic individuals were also classified as 'non-Black'.

Dietary Recall

Once a survey person has been identified, an advance letter was sent to last known address of the person, along with incentives to participate. Meal questionnaires were administered once a participant has been identified. The data collected was assisted with a Computer Assisted Personal Interview (CAPI) and done in person for general demographic, socioeconomic, dietary, and health-related questions. Once the in-person survey was complete, the participant would be invited to the MEC for further examination. Days when the survey taker could come to the MEC for completing the dietary recall was scheduled ahead of time, both for day one and day two dietary recall. Protocol states time separation from day two of a dietary intake is meant to be within three days after the first survey. Weighted data was adjusted for non-response for the second recall, and proportion intakes on weekends and weekdays. The full in-person questionnaire for dietary meals were completed at the MEC, broken down into three parts: the dietary recall, the supplement and antacid section, and the post-recall. The program used by USDA AMPM through a program called “Blaise Data Entry”.

The 24-hour dietary recall collected a list of all food and beverages within the 24-hour timespan, including: what time the food was eaten and the reasoning, descriptions of the food and the amount, where it was obtained, and where the food was eaten. The food is completed in a five-step process, starting with: a general quick list, of foods consumed the day before from midnight to the midnight of ‘today’; forgotten foods that may be consumed as snacks, or specific foods that are forgotten; what was the occasion for consuming the food; the detail and review cycle, detailing food amounts, their preparation, and where it was eaten; finally, to see if there was any other foods they may

have missed. This is done using the USDA ‘What We Eat in America’ survey. Combination foods such as cereals, ice creams, salads and sandwiches were noted in the survey study, estimated by survey taker estimates on how much of the food items was added. Nurtritional data was estimated on food composition data. This was used to obtain caffeine intake as well as water, sugar, alcohol, and sodium. Brand named products having information regarding their nutritional content were encouraged by the survey input. Likewise for fluid measurements, cups and/or glasses were standardized based on models created by Anchor Hocking, Crate & Barrel, Food Wares, Carlisle, and Kroger. These sizes and models were used to help the sample person estimate how much they had drank or eaten (NHANES, 2015). Water was measured in grams while sodium and caffeine were measured in milligrams. Water sources were categorized by bottled or tap water, with tap water stated as “community supply”. For the purposes of this study, water intake was averaged between plain and tap water for total water intake. For all variables obtained through the dietary recall, the data was defined as continuous numerical data. Caffeine in addition would be split into quartiles based on the total caffeine contained within the sample size ages between 18 to 55.

Quality control for the interviews were recorded, first informing the survey person if they agree. The recording starts at the beginning of the initial interview and stops at the end of the interview. Westat or National Center for Health Statistics (NCHS) staff quality control members were to observe the questionnaires were conducted. Interview data input into the MEC CAPI have internal error and consistency checks to ensure that all sections are filled out (NHANES, 2015).

Blood Pressure

While age, sex, and race/ethnicity were the primary factors accounted by the eGFR equation, additional data such as blood pressure, and body mass index were included as they relate to those three variables. Blood pressure was measured in millimeters of mercury (mmHg) using the systolic and diastolic measures. Blood pressure was measured in different methods, including which arm it was taken, the cuff size, and if they had an irregular heartrate or not. An Omron IntelliSense Blood Pressure Monitor was used to measure the blood pressure and provided the battery pack, air tubes, and cuffs. Limitations to the machine's ability to detect blood pressure were: Systolic blood pressure greater than 300 mmHg, or the pressure was calculated being lower than diastolic blood pressure; differences between systole and diastole cannot be less than 27 mmHg or greater than 96 mmHg; minimum and maximum systolic pressure cannot be greater than 30 mmHg; and the same follows for diastolic pressure.

There are also limits to systolic and diastolic measurements based on age ranges such that 18–39 years old have a range of 90 to 155 mmHg systolic pressure, and 51-104 diastolic pressure. Age ranges between 40-59 have a range between 91 to 179 mmHg systolic pressure, and 54 to 108 diastolic pressure. Daily quality control checks were done prior to use, including cleaning the machine and its equipment, and logged into the machine's quality control checks (NHANES, 2019).

Body Mass Index

Body mass index (BMI) was measured in kilograms per meters squared (kg/m^2). Derived from weight and height, these two measures were calculated using an electronic

floor scale and stadiometer completed at a MEC. Measures were converted to metric system values, kilograms, and meters. For survey persons going to a MEC, a limit cap on weight above 450 lbs. was implemented and excluded from the entire body composition, with a maximum capacity up to 600 lbs. for the digital weight scale, and 440 lbs. for the portal scales in case the digital scale was malfunctioning. Likewise, a hard limit cap on height was implemented for people above 6'5". While the measurement of the weight and/or height was recorded, this was noted due to a mechanical issue that could occur with another procedure for the body composition measurements, unrelated to BMI.

Quality control for the BMI equipment was handled by daily checks prior to testing and by the staff technologist. The digital equipment for electronically reading the measurements was built by Mettler Toledo. Fifteen 10-Kg weights were used to calibrate the electronic scale, placed at top, middle, and bottom of the scale to ensure proper reading. A 100-cm metal rod was used for calibrating the stadiometer. Quality control was completed through a check box program to ensure functionality. This was also recorded for quality control. Major equipments such as the digital scale and stadiometer were custom-built with fail-safes in case of malfunctions. In the case of the electronic scale, a portable digital scale would be used. For the stadiometer, physically reading the notches for height measurements were done (NHANES, 2018).

Exclusion Criteria

Several exclusion criteria were set for this investigation. For the calculation done in this investigation, a limit was set for participants who's BMI score was lower than or equal to 40 kg/m^2. This is further explained in the Body Mass Index section of Chapter 2.

Another exclusion was done on the dietary data, specifically participants with specific kidney and renal diets. These were tailored for those with chronic kidney disease, limiting their foods with sodium, potassium, and phosphorus content in order to slow their CKD progression (Cleveland Clinic, 2017). As such, these were excluded from the study, due to targeted dietary needs and already having lowered renal function. On the NHANES dietary survey, questions were formatted to ask the sample person if they had any specific dietary needs or were following regiments. These were recorded on the survey as a separate entry. Due to the targeted nature of the diets, these sample persons were excluded from analysis (NHANES, 2017).

Pregnant women, diabetics, and those suffering from kidney related complications were also excluded from the studies due to excess strain on the bladders, the known effects of diabetes on those effected with CKD, and those who already have renal complications. NHANES listed on the Questionnaire data whether a sample person was pregnant or not as ‘Yes’, ‘No’, ‘Refused’, ‘Don’t know’. For diabetic persons, the question the sample person was asked if they have diabetes was given answers of ‘Yes’, ‘No’, ‘Borderline’, ‘Refused’, and ‘Don’t Know’. Those with kidney related issues were given the same type of answers. Questionnaires were done in person at a MEC, or at the sample person’s home. For people at the MEC, a computer screen was used to illustrate directly to the person, in a secluded room. Pre-recorded questions were given to the patient, allowing them to choose their answer. These questions could be administered in a variety of languages, such as English, Spanish, Mandarin Chinese, Korean, and Vietnamese. A known limitation of such a method was if the sample person was lying in the answer or misintrepretations and answers for the question being asked. For this study,

pregnant women who did not know if they were pregnant were classified as ‘pregnant’ and excluded from the study. Additionally, persons that were ‘Borderline’ and ‘Don’t Know’ with diabetes were excluded as well.

Quality Control of the machines were done with various checks and built-in error and consistency checking. Live interviewers served were recorded as the knowledge of the survey person to make sure the questions asked were not invasive. The program used for quality control and recording was CAPI, with an audio computer-assisted self-interviewing (ACASI) system used to pause, resume, and stop recording questions (NHANES, 2017).

Two other exclusions were formed after plotting and observing the statistical data provided by NHANES. The dataset from 2013-2014 reported a caffeine intake over 5000 mg. While plausible for a person could consume multiple sources of caffeine, in comparison to the other datasets used in the thesis, and some deliberation with the committee members, it was determined that said data point was to be considered an outliner. This was also shown when the combine datasets were plotted eGFR by caffeine intake. A limit of 2000 mg of caffeine consumed was set. Additionally, participants who’s caffeine intake was missing from the combined dataset were also excluded from the models. How caffeine was calculated was explained in the Dietary Recall section.

Confounding Factor Definition

This study controlled for Body Mass Index (BMI), hypertension, hyperuricemia, and alcohol that may weight into the potential correlation. These potential confounders were first identified by a review of the literature and determination of factors well-known

to be associated with eGFR. High BMI is generally associated with obesity, which is a known risk factor to chronic kidney disease (Johansen and Lee, 2015). Cases with high BMI or extreme obesity induce kidney strain, due to increased blood pressure. High BMI values were applied as value 40 and above. Similarly, hypertension was also controlled for due to its known association with chronic kidney disease as a risk factor (CDC, 2020). Hypertension and BMI calibration, calculation, and quality control were explained in Exposures, Characteristics, and Comorbidities. Alcohol intake was controlled for due to its diuretic effects, reducing the ion reabsorption and water retention through the nephron tubule (Epstein, 1997). Similarly, the calculation, calibration, and quality control of alcohol intake was explained in Exposures, Characteristics, and Comorbidities (NHANES, 2017).

Hyperuricemia was also investigated as a potential confounder (Johnson et al, 2018). This variable was measured in mg/dL. Hyperuricemia was defined as high uric acid concentration in the blood above 7 mg/dL. It was drawn from the blood at a MEC and stored in a 100 µL sample cup or a 2 mL microtube in the procedure done for serum creatinine. Samples were then sent to one of 35 laboratories in the United States, refrigerated or frozen before transport. Limitations of the sample were found with calcium dobesilate creating artificially low uric acid results along with purine derivative such as type IgM, having a similar effect. Laboratory interferences in calculating the uric acid concentration would not occur with the icteric index below 40, the hemolytic index below 1000, the lipemic index below 1500, and the ascorbic acid below 3 mg/dL. An automatic re-analysis was done if the limits were triggered at low or high ends.

The equipment used to draw the samples were a Roche Cat No. 03183807190, UA2 reagent kit and a Roche Cobas 6000 Chemistry analyzer and Millipore Elix Gulfstream Clinical 35 System. To run quality control and assurance, the Roche Calibrator for Automated Systems, catalog #10759350360 was used, stored until expiration date on the bottle at 2-8°C. This was done with 3.0 mL deionized water pipette. A serum control was used for daily calibration. Manual calibration was done due to the following reasons: reagent lot change not occurring over the last 6 months, after major service/repair, and if there was troubleshooting issues. The laboratory enlisted in the College of American Pathologist (CAP) linearity program for uric acid samples were shipped samples in a LN2 kit twice a year for checkups. Members were blinded during extraction and sampling (NHANES, 2020).

Members who's blood pressure could be considered hypertensive were also excluded from the investigation, meaning systolic blood pressure averaging above 130 mm Hg and diastolic blood pressure averaging above 80 mm Hg were excluded from the data. How blood pressure was calculated was explained in the Blood Pressure section.

Written Informed Consent

All participants used in the NHANES survey provided written informed consent at the time of the study as previously stated in the Target Population and Study Population section. For this thesis, the Temple University Institutional Review Board was informed of the present study and classified this thesis as exempt.

Search Strategy

The methods of searching for this thesis is meant as a guide to understanding how specific articles and studies were used in the Background section. In no form is this thesis meant to be a systematic review of known caffeine studies on lowered renal function. From initial search results, more specific key words were used to narrow studies that may have been performed previously. Search strings for this topic were used through PubMed and CINAHL for various pieces of literature such as: low renal function was defined by chronic kidney disease, urine flow, kidney stone formation, and possible biological correlation with caffeine to renal cells. Phrases such as 'renal function', 'kidney function', 'kidney disease', and 'caffeine' were used to identify various articles, studies, and research topics meant to serve as a baseline for analysis and data holes in caffeine correlation with renal function. In PubMed, there is a secondary function for filtering studies, used to filter for meta-analyses, and clinical trials. Several meta-analyses were used in this thesis: One meta-analysis consisted of three prospective cohort studies on caffeine's effects on kidney stone formation; Two meta-analyses of observational studies were used for the study, based on different caffeine products for an older population; and two animal experimental studies to understand the biological component of the study. Another meta-analysis reviewed were created by the same authors as a further investigation into caffeine concentration and its potential renal effects. In order to calculate the renal function, another meta-analysis consisting of 25 different general population cohorts, 7 high-risk cohorts, and 13 CKD cohorts was used to determine what mathematical function would be used to calculate renal function. Other retrospective

cohort studies included studies on specific populations, and disease analysis from caffeine consumption.

As noted, the lowered renal functions are categorized by a broad list of conditions and diseases, such as chronic kidney disease, kidney stone formation, and decreasing glomerular filtration rate. These conditions are typically associated with a middle-aged to older population. The Centers for Disease Control and Prevention states that the prevalence of CKD is higher among people above 65 years of age. While there has been noted cases between the lower age ranges, an overwhelming majority is with the older population (CDC, 2019). However, that does not imply that full resources should be solely focused on the older populations.

Statistical Analysis

Data collected for the investigation was evaluated and analyzed through the SAS programming language, the standard programming language used in epidemiological studies. Several linear regressions were performed, a simple regression model for the general correlation between eGFR and caffeine, an assessment regression model to check beta coefficients for understanding potential confounding, and a multivariable linear regression, with the stated comorbidities and exclusion criteria accounted for. Oversampling for Hispanic, non-white Black, and Asian populations was stated on the NHANES website, weighted by variable “WTDR2D”. However, as per NHANES data compiling, the weighted variables modified to account for different survey years, set as “MEC6YR” (NHANES, 2021a). All exclusion criteria was compiled into a single dummy variable to be used in the “domain” statement for all viable scripts.

A frequency table of the collected data years was done to measure how many subjects were within the age limits, with any stated levels of caffeine consumed. The data was then further split between sex, “male” and “female”. The means of the data followed next, split between sex and race. Exclusion criteria was applied to this section. A simple parameter regression was done with the simple variable regression between caffeine and eGFR. An assessment of confounders regression model followed, in order to determine if specific parameters were confounders with caffeine intake. Lastly, a multivariable regression was completed, with all parameters as explained in the previous chapter.

Similar to the simple linear regression model above, a second eGFR model was analyzed in order to understand if there were confounding parameters with caffeine intake. Parameters were paired with caffeine to see how the beta coefficient of caffeine changed. Percent changes over ten percent were considered to be confounding factors to caffeine intake.

CHAPTER 3

RESULTS

Sample Characteristics

The results of **Table 1** show the frequency of calculable caffeine for the age group between 18 to 55, separated between genders, and then races. A total number of 10560 participants were within this age range, prior to exclusion criteria and weighted values. Following exclusion criteria and caffeine intake who's value was not missing, only 4263 members had calculable caffeine values. Of the 4263 members, 2436 (57.14%) of the persons were female, while the remaining 1827 (42.86%) were male. As stated previously, Mexican American, and other Hispanic ethnic groups were over-sampled in the data collection, resulting with 732 (17.17%) of the sample population identifying as "Mexican American" and 489 (11.47%) as "Other Hispanic" ethnicities. The Non-Hispanic White population remained as most of the total identified populations with 1453 (34.08%) persons with Non-Hispanic Blacks as the other large sample population. Subsequent smaller groups included the Non-Hispanic Asian and multiracial groups.

Caffeine Demographic Statistics Ages 18-55			
Gender	**Number of People**	**Percent**	**Total Sample**
Male	1827	42.86%	4263
Female	2436	57.14%	
Ethnicity	**Number of People**	**Percent**	**Total Sample**
Mexican American	732	17.17%	4263
Other Hispanic	489	11.47%	
Non-Hispanic White	1453	34.08%	
Non-Hispanic Black	796	18.67%	
Non-Hispanic Asian	594	13.93%	
Other Race (Multi-Racial)	199	4.67%	

Table 1: Demographic Statistics Ages 18-55

Other parameters to be used in the multivariable model following exclusion criteria were illustrated in **Table 2**. Systolic and diastolic blood pressures were separate parameters, while water consumption through tap water and bottled water were averaged together for the survey taker. Uric acid and BMI were measured as a singular measurement. Other parameters have smaller total sample sizes than caffeine concentration. This was due to missing data from the dataset. Following NHANES recommendations for calculating variance, the values were treated as "not missing completely at random" (NHANES, 2021b).

Variables Used in the Regression Model		
All Values Split by Quartiles		
Caffeine	**Number of People**	**Total Sample**
Up to 5	1267	4050
<5 to 83	1100	
<83 to 192	943	
>192+	740	
BMI (Kg/m^2)	**Number of People**	**Total Sample**
Up to 23	1045	4020
<23 to 26.5	976	
<26.5 to 30.9	1025	
>30.9+	974	
Sugar (grams)	**Number of People**	**Total Sample**
Up to 53.66	1023	4050
<53.66 to 89.73	1043	
<89.73 to 134.77	1004	
>134.77+	980	
Water (grams)	**Number of People**	**Total Sample**
Up to 253.5	1165	4050
<253.5 to 705	1035	
<705 to 1312.5	981	
>1312.5+	869	
Uric Acid (mg/dL)	**Number of People**	**Total Sample**
Up to 4.1	1067	4047
<4.1 to 5	1096	
<5 to 5.9	930	
>5.9+	954	
Sodium (mg)	**Number of People**	**Total Sample**
Up to 2169	1094	4050
<2169 to 3041	986	
<3041 to 4148	946	
>4148+	1024	
Systolic Pressure (mmHg)	**Number of People**	**Total Sample**
Up to 105.33	935	3946
<105.33 to 112	1133	
<112 to 118	971	
>118+	907	
Diastolic Pressure (mmHg)	**Number of People**	**Total Sample**
Up to 62.67	1208	3946
<62.67 to 68	958	
<68 to 73.33	809	
>73.33+	971	
eGFR (mL/min)	**Number of People**	**Total Sample**
Up to 93.92	814	4050
<93.92 to 108.86	929	
<108.86 to 126.27	1006	
>126.27+	1301	

Table 2: Parameter Table to be Used in The Regression Models. These values were split into quartiles based on the total number of variables that were within the age range.

Table 3 further illustrates the breakdown of those members with calculable caffeine. The first column indicated as “Caffeine Intake”, split into quartiles. These

quartiles are then used to illustrate the number of people within the specific age group in the row, separated by gender. The second column indicated as "eGFR" is derivative of the members who have recorded caffeine intake, and how many have calculable eGFR.

The while caffeine intake for the lower quartiles did not reveal a direct increase or decrease in consumption between the age groups, members of the fourth quartile in caffeine consumption experienced an increased trend as they grew older. As biologically expected, as age increased, eGFR decreased.

Caffeine Intake and Calculated eGFR By Age Groups and Gender

Age Groups	Gender	Caffeine Quartile Split (mg)	Caffeine Intake		eGFR
			Number of Persons	Mean (mg)	Mean (mL/min/1.73m^2)
18 up to 35	Male	Up to 5	453	0.57	115.87
		<5 to 83	347	42.88	118.02
		<83 to 192	263	134.84	113.68
		>192+	218	361.48	108.75
	Female	Up to 5	580	0.76	126.45
		<5 to 83	471	43.57	126.45
		<83 to 192	347	131.25	118.40
		>192+	191	319.89	113.13
<45 to 55	Male	Up to 5	82	0.51	100.91
		<5 to 83	95	45.71	96.46
		<83 to 192	135	139.48	94.88
		>192+	141	415.94	96.11
	Female	Up to 5	152	0.98	98.91
		<5 to 83	187	39.96	104.25
		<83 to 192	198	134.45	101.22
		>192+	190	370.82	98.38

Table 3: Age Stratification for Caffeine Intake. The total number of persons in the caffeine intake is the same total number who has calculable eGFR, following the exclusion of caffeine intake values that were missing.

Table 4 illustrates the means of the variables as well as the other parameters used in the model, along with the total number of survey persons, standard error of the mean,

and 95% confidence intervals from the mean. Like the previous table, some parameters to be used in the model have larger number of survey persons due to missing data.

General Statistical Analysis of Parameters and Model					
Parameters in the Model	Number of Survey Data Points	Mean	Standard Error for Mean	95% Confidence Interval for Mean	
				Lower Bounds	Upper Bounds
Caffeine Intake (mg)	4050	134.71	4.80	125.05	144.37
Sugar Intake (g)	4050	103.74	1.57	100.57	106.90
Water Intake (g)	4050	962.40	32.51	896.92	1027.87
BMI (Kg/m^2)	4020	27.10	0.15	26.79	27.40
Systolic Blood Pressure Average (mm Hg)	3946	111.58	0.22	111.14	112.01
Diastolic Blood Pressure Average (mm Hg)	3946	67.05	0.24	66.55	67.54
Uric Acid Concentration (mg/dL)	4047	5.06	0.03	5.00	5.12
Sodium (mg)	4050	3320.03	35.82	3247.89	3392.18
eGFR (mL/min/1.73m^2)	4050	111.85	0.87	110.09	113.60

Table 4: General Statistical Analysis of Parameters and Model.

Simple Linear Regression

The first linear regression model used was caffeine consumption correlating to eGFR. All exclusion criteria was applied. **Table 5** below illustrates with a simple linear regression model that a single milligram of caffeine would detract from a person's eGFR by 0.022 mL/min/1.73m^2 with 2.39% of the eGFR variability accounted for by caffeine intake alone.

eGFR Linear Regression Model Based on Caffeine Intake					**R-Squared Value**
Parameter	*Estimate*	*Standard Error*	*t-Value*	*p-Value*	0.02393
Intercept	114.7756	1.08445	105.84	<.0001	
Caffeine Intake (mg)	-0.021752	0.0047212	-4.61	<.0001	

Table 5: Simple Linear Regression Model Based on Caffeine Intake.

A scatter plot was generated to illustrate the correlation between caffeine and eGFR. As this is a raw data sample with no weights or exclusion criteria, it does not reflect was is stated on **Table 5**. As stated previously, a caffeine intake above 2000 mg was excluded. The plot of eGFR by daily dietary intake shows there is a much larger range of eGFR function values at low intake when compared to high caffeine intake.

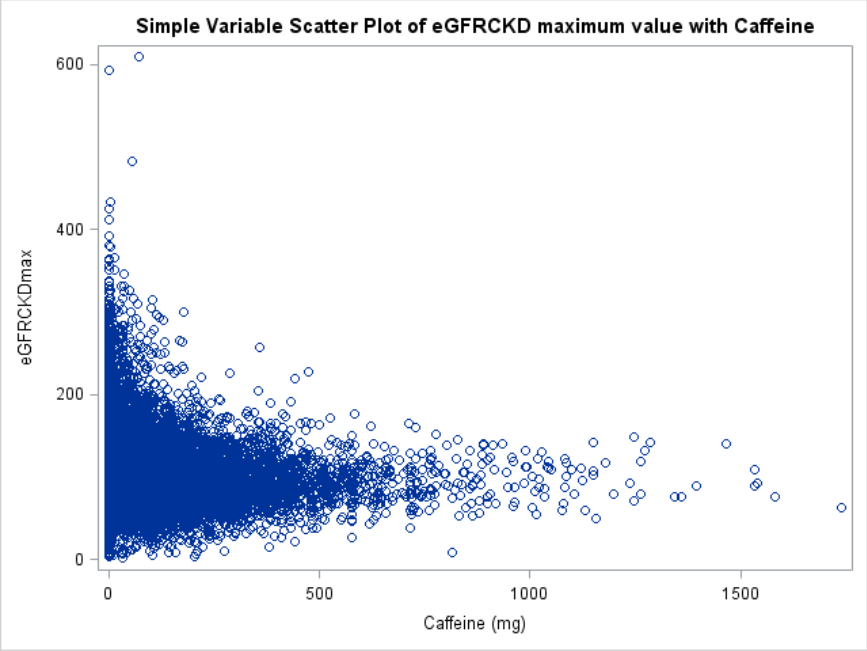

Figure 1: Simple Scatter Plot Between Caffeine Intake and eGFR.

Assessment of Confounding

Table 6 summarizes the beta coefficient for caffeine intake following adjustment for a single potential confounding variable and the percent change, when compared to the crude (unadjusted) simple regression model for caffeine. Based on the 10% change to caffeine intake, age was calculated as a confounder.

Caffeine Intake Beta Coefficient With Single Parameter Added to Model										
	Caffeine Intake (Reference)	Sugar Intake	Water Intake	Uric Acid	BMI	Systolic BP	Diastolic BP	Sodium	Age	Gender
Caffeine Intake	-0.021752	-0.021310	-0.022511	-0.021471	-0.020982	-0.020969	-0.019940	-0.021719	-0.007535	-0.021034
Absolute Percent Change Compared to Reference	Ref	2.03	3.49	1.29	3.54	3.60	8.33	0.15	65.36	3.30

Table 6: Assessment of Confounding Beta Coefficients.

Multivariable Linear Regression

Three multivariable linear regression models were created in order to assess the confounding variable of age, between age groups 18 to 40 and 40 to 55, and well limiting the caffeine intake to at or below 2000 mg. **Table 7** below is the multivariable linear regression model with the parameter's beta coefficients, standard error, t-value, p-value, and r-squared value. A single milligram of caffeine would reduce eGFR by 0.007 mL/min/1.73m^2. The estimated eGFR variability of 16.91% was accounted for by caffeine intake and age.

eGFR Multivariable Linear Regression				
Parameter	*Estimate*	*Standard Error*	*t-Value*	*p-Value*
Intercept	145.81855	1.5293933	95.34	<.0001
Caffeine Intake (mg)	-0.007535	0.0026219	-2.87	0.006
Age (years)	-0.950291	0.0399194	-23.81	<.0001
R-Squared Value	0.1691			

Table 7: eGFR Multivariable Linear Regression with Confounder.

Table 8 is the multivariable linear regression model with age as the confounder, however the value for age was split from 18-40 years old. The linear regression model illustrates with a single milligram decreased eGFR by 0.009 mL/min/1.73m^2. With the parameters, 5.69% of the variability of eGFR was accounted for the simple linear model. In parallel with **Table 9,** another linear regression model for participants ages 40 to 55

illustrated that a single milligram increase of caffeine would reduce eGFR by 0.006 mL/min/1.73m^2 with 4.38% accounting for the eGFR variability. Finally, in **Table 10**, as previously discussed, some participants caffeine intake ranged into 5000 mg. This only appeared in the 2013-2014 NHANES dietary survey data. In order to determine whether these extreme values influenced model parameters, an additional model limiting caffeine intake to 2000 mg/day was created. When a single milligram of caffeine is consumed, eGFR is reduced by 0.009 mL/min/1.73m^2, and 16.92% accounting for eGFR variability by caffeine intake and age.

eGFR Multivariable Linear Regression for Ages 18-40				
Parameter	Estimate	Standard Error	t-Value	p-Value
Intercept	145.02098	2.281368	63.57	<.0001
Caffeine Intake (mg)	-0.009384	0.0043015	-2.18	0.034
Age (years)	-0.91237	0.0963133	-9.47	<.0001
R-Squared Value	0.0569			

Table 8: eGFR Multivariable Linear Regression for Participants Between Ages 18-40.

eGFR Multivariable Linear Regression for Ages 40-55				
Parameter	Estimate	Standard Error	t-Value	p-Value
Intercept	145.48705	6.8278685	21.31	<.0001
Caffeine Intake (mg)	-0.006459	0.0025447	-2.54	0.015
Age (years)	-0.949919	0.1421541	-6.68	<.0001
R-Squared Value	0.04376			

Table 9: eGFR Multivariable Linear Regression for Participants Between Ages 40-55.

eGFR Multivariable Linear Regression for Caffeine Intake Below 2000 mg				
Parameter	Estimate	Standard Error	t-Value	p-Value
Intercept	145.82338	1.5295472	95.34	<.0001
Caffeine Intake (mg)	-0.008974	0.0027545	-3.26	0.002
Age (years)	-0.945132	0.0401663	-23.53	<.0001
R-Squared Value	0.1692			

Table 10: eGFR Multivariable Linear Regression for Participants with Caffeine Intake Below 2000 mg.

CHAPTER 4

DISCUSSION

This thesis was meant to evaluate the potential correlation between caffeine intake and lowered renal function measured through eGFR. Review of the data showed that daily dietary caffeine intake has a strong association with lowered eGFR. From the single variable to multivariable regression models, the association was similar in both crude and adjusted models. However, the correlation and strength of the effecting overall eGFR does change with increased parameters.

While the caffeine parameter is significant, a better illustration would be from examples explained through the thesis as well as one particular real-life example. From the first single parameter linear regression model, the beta coefficient for caffeine was -0.021752. If a theorical person was to consume to the beginning of the fourth quartile (193 mg), this would in turn reduce eGFR by 4.20 mL/min/1.73m^2. In direct comparison, the *Coca-Cola Company's* nutritional fact webpage stated for a single 20 oz. bottle of Coca-Cola, there was 57 mg (Coca-Cola Company, 2021). Drinking one bottle would reduce eGFR by 1.24 mL/min/1.73m^2.

As there are four different multivariable linear regression models, **Table 11** below illustrates the eGFR reduction by each model and the caffeine intake based on the low end of the fourth quartile and a single 20 oz. bottle of Coca-Cola.

eGFR Reduction by Caffeine Intake (mL/min/1.73m^2)			
Model	Caffeine Beta Coefficient of the Model	Coca Cola 20 oz. bottle (57 mg)	Lowest Fourth Quartile (193 mg)
Simple Linear Model	-0.021752	-1.24	-4.20
Linear Model with Age Parameter	-0.007535	-0.43	-1.45
Linear Model with Age Parameter (Ages 18-40)	-0.009384	-0.53	-1.81
Linear Model with Age Parameter (Ages 40-55)	-0.006459	-0.37	-1.25
Linear Model with Age Parameter (Caffeine Intake =<2000 mg)	-0.008974	-0.51	-1.73

Table 11: eGFR Reduction by Caffeine Intake.

The scatter plot revealed that caffeine consumption most concentrated below 500 mg per day of intake. While the quartile data spread of caffeine consumption illustrated caffeine consumption remaining below 500 mg, the mean fell within the third quartile. In comparative terms, the average person may consume over two 20 oz. bottles of Coca-Cola in caffeine. As expected with increased age, the filtration rate of the kidney's decreases. One of the surprising outcome was how gender was not as potentially confounding with caffeine intake correlation with eGFR as thought, as illustrated by **Table 6**. This calculation was despite gender being a component in the CKD-EPI equation, splitting male and female.

In the assessment for parameter confounding, three parameters adjusted caffeine intake's beta coefficient stronger than the other potential confounders: systolic and diastolic blood pressures, and age. While systolic blood pressure's change was compariable to the other parameters, diastolic blood pressure had a stronger change. However it should be noted that having hypertension would increase strain on the

kidneys, reducing eGFR. These two parameters are biologically tied to one another, thus this change was not unexpected.

Age was the other confounding parameter and was included into the multivariable model. As expected, age contributed a large part in eGFR's calculation. This outcome fell inline with why there is limited data on CKD for people under the age of 45, as this population only makes up approximately 6% of CKD cases (CDC, 2019). Due to the significantly smaller cases in this age range, there is little need to investigate these potential effects.

One of the other parameters explained in potential confounding effects was uric acid, previously explained in Chapter 2 as a potential biomarker for kidney and cardiovascular disease (Johnson et al, 2018). However, the assessment for confounding did not illustrate confounding effects with respect to daily dietary caffeine intake. Similarly stated in the previous study (Johnson et al, 2018), the potentional of uric acid being a biomarker for kidney and cardiovascular disease exists, however the methods and power that this biochemical was extracted are still in current debate.

Strengths

One of the strengths of this thesis lays on the data structure, availability, and accessibility of obtaining the datasets from NHANES. Each consecutive year, the NHANES survey updates, taking the survey data obtained through interviews and MECs. This data is updated constantly, with instructions and manuals on how the data was collected, how it was measured, and what weights were used. Changes to the measurement of how the data was collected is listed on the NHANES website, explaining

differences. Additionally, NHANES also provides a free-user SAS download for people that do not have SAS programming language if they wish to see the data. This ability to access free information on the United States and its demographics provides others the ability to test and examine the data themselves.

Another strength of the thesis was the sample size of the persons used. The sample size of the dataset was over 4000 members following exclusion criteria between the age ranges of 18 to 55, spread across three different data-cycles. These members ranged from different ethnicities including an over-sampled Hispanic population in the United States. While the age range does include a small range where observable lowered renal function may occur, its inclusion illustrates the caffeine consumption for the older generation, regardless of caffeine source.

One final strength of this study was that it relied upon data intended to be generalizable to the US population.

Limitations

NHANES data is a cross-sectional survey. The chief limitation of cross-sectional designs is the order of the exposure and the outcome. Blood was drawn immediately within a few days after answering the dietary intake questionnaire, increasing the likelihood that recall of recent intake could more accurately, reflect a subsequent response in the serum creatinine biomarker upon which the eGFR calculation relies. While one can infer those dietary behaviors may not change day to day for members, there are other disadvantages to this approach. When a survey person attended the MEC, there was separation between which days they were able to go. People may experience

different behavioral patterns depending on weekdays, weekends, or individual days. A direct temporal link between caffeine and serum creatinine leading to lowered eGFR is difficult to establish as caffeine would not stay in the body for a longer period of time. Additionally, as NHANES dietary recalls were based on estimates provided at the survey person's own recollection.

The statement of estimated values was also affected by the nature of a person remembering what they had eaten. The NHANES protocol included prompts to help recall of dietary information. A review (Archer, Pavela, and Lavie, 2015) called into question the validity of a person's memory for dietary meals and eaten foods, specifically for NHANES dietary recall. The possibility of false reporting what the survey person ate was noted, done through the Deese-Roediger and McDermott (DRM) paradigm, where a person was to repeat food items that were read or spoken to them minutes prior.

There was initially a step to possibly separate specific types of caffeine sources. A major difference between coffee and energy drinks is the presence of acrylamide in coffee (El-Zakhem Naous et al., 2018). At the time of the study, the acrylamide concentration data for the 2017-2018 cycle had not been released. This gap in data was unacceptable for the model regression and thus was left out. A future prospect would to be run the model check using acrylamide, if it would lower eGFR.

Conclusion

In conclusion, while this thesis has illustrated association between high caffeine concentrations and lowered eGFR, it should not be taken as definitive proof. Further

examination and studies must be done to allow greater accuracy in determining caffeine's negative effects on lowered renal function.

BIBLIOGRAPHY

Al Yacoub, R., Luczkiewicz, D., & Kerr, C. (2020). Acute kidney injury and hepatitis associated with energy drink consumption: a case report. Journal of medical case reports, 14(1), 23. https://doi-org.libproxy.temple.edu/10.1186/s13256-019-2340-0

Archer, E., Pavela, G., & Lavie, C. J. (2015). The Inadmissibility of What We Eat in America and NHANES Dietary Data in Nutrition and Obesity Research and the Scientific Formulation of National Dietary Guidelines. Mayo Clinic proceedings, 90(7), 911–926. https://doi-org.libproxy.temple.edu/10.1016/j.mayocp.2015.04.009

Belibi, F. A., Wallace, D. P., Yamaguchi, T., Christensen, M., Reif, G., & Grantham, J. J. (2002). The effect of caffeine on renal epithelial cells from patients with autosomal dominant polycystic kidney disease. Journal of the American Society of Nephrology : JASN, 13(11), 2723–2729. https://doi-org.libproxy.temple.edu/10.1097/01.asn.0000025282.48298.7b

Cappelletti, S., Piacentino, D., Sani, G., & Aromatario, M. (2015). Caffeine: cognitive and physical performance enhancer or psychoactive drug?. Current neuropharmacology, 13(1), 71–88. https://doi-org.libproxy.temple.edu/10.2174/1570159X13666141210215655

Centers for Disease Control and Prevention. (2015). Measuring Guides for the Dietary Recall Interview. National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/measuring_guides_dri/measuringguides.htm

Centers for Disease Control and Prevention. (2020). Chronic Kidney Disease Basics. Chronic Kidney Disease Initiative. https://www.cdc.gov/kidneydisease/basics.html

Centers for Disease Control and Prevention. (2019). Chronic Kidney Disease in the United States, 2019. Centers for Disease Control and Prevention. https://www.cdc.gov/kidneydisease/publications-resources/2019-national-facts.html

Clark, W. F., Sontrop, J. M., Huang, S. H., Gallo, K., Moist, L., House, A. A., Cuerden, M. S., Weir, M. A., Bagga, A., Brimble, S., Burke, A., Muirhead, N., Pandeya, S., & Garg, A. X. (2018). Effect of Coaching to Increase Water Intake on Kidney Function Decline in Adults With Chronic Kidney Disease: The CKD WIT Randomized Clinical Trial. JAMA, 319(18), 1870–1879. https://doi-org.libproxy.temple.edu/10.1001/jama.2018.4930

Cleveland Clinic. (n.d.). Renal Diet Basics. Clevland Clinic. https://my.clevelandclinic.org/health/articles/15641-renal-diet-basics

Coca-Cola Company (2021). Coca-Cola® Original. Our Products. https://us.coca-cola.com/products/coca-cola/original

Crews, D. C., Bello, A. K., & Saadi, G. (2019). Burden, access, and disparities in kidney disease. Clinical nephrology, 91(3), 129–137. https://doi-org.libproxy.temple.edu/10.5414/CN91WKDEDI

Epstein M. (1997). Alcohol's impact on kidney function. Alcohol health and research world, 21(1), 84–92.

El-Zakhem Naous, G., Merhi, A., Abboud, M. I., Mroueh, M., & Taleb, R. I. (2018). Carcinogenic and neurotoxic risks of acrylamide consumed through caffeinated beverages among the lebanese population. Chemosphere, 208, 352–357. https://doi-org.libproxy.temple.edu/10.1016/j.chemosphere.2018.05.185

Fenton, R. A., Poulsen, S. B., de la Mora Chavez, S., Soleimani, M., Busslinger, M., Dominguez Rieg, J. A., & Rieg, T. (2015). Caffeine-induced diuresis and natriuresis is independent of renal tubular NHE3. American journal of physiology. Renal physiology, 308(12), F1409–F1420. https://doi-org.libproxy.temple.edu/10.1152/ajprenal.00129.2015

Ferraro, P. M., Taylor, E. N., Gambaro, G., & Curhan, G. C. (2014). Caffeine intake and the risk of kidney stones. The American journal of clinical nutrition, 100(6), 1596–1603. https://doi-org.libproxy.temple.edu/10.3945/ajcn.114.089987

Harer, M. W., Askenazi, D. J., Boohaker, L. J., Carmody, J. B., Griffin, R. L., Guillet, R., Selewski, D. T., Swanson, J. R., Charlton, J. R., & Neonatal Kidney Collaborative (NKC) (2018). Association Between Early Caffeine Citrate Administration and Risk of Acute Kidney Injury in Preterm Neonates: Results From the AWAKEN Study. JAMA pediatrics, 172(6), e180322. https://doi-org.libproxy.temple.edu/10.1001/jamapediatrics.2018.0322

Hartley, T. R., Lovallo, W. R., Whitsett, T. L., Sung, B. H., & Wilson, M. F. (2001). Caffeine and stress: implications for risk, assessment, and management of hypertension. Journal of clinical hypertension (Greenwich, Conn.), 3(6), 354–361. https://doi-org.libproxy.temple.edu/10.1111/j.1524-6175.2001.00478.x

Hashibe, M., Galeone, C., Buys, S. S., Gren, L., Boffetta, P., Zhang, Z. F., & La Vecchia, C. (2015). Coffee, tea, caffeine intake, and the risk of cancer in the PLCO cohort. British journal of cancer, 113(5), 809–816. https://doi-org.libproxy.temple.edu/10.1038/bjc.2015.276

Healthy People 2020 (2020). Search the DATA: Healthy People 2020. Retrieved February 03, 2021, from https://www.healthypeople.gov/2020/data-search/Search-the-Data#objid=4076

Hypolite, J. A., & Malykhina, A. P. (2015). Regulation of urinary bladder function by protein kinase C in physiology and pathophysiology. BMC urology, 15, 110. https://doi-org.libproxy.temple.edu/10.1186/s12894-015-0106-6

Jhee, J. H., Nam, K. H., An, S. Y., Cha, M. U., Lee, M., Park, S., Kim, H., Yun, H. R., Kee, Y. K., Park, J. T., Chang, T. I., Kang, E. W., Yoo, T. H., Kang, S. W., & Han, S. H. (2018). Effects of Coffee Intake on Incident Chronic Kidney Disease: A Community-Based Prospective Cohort Study. The American journal of medicine, 131(12), 1482–1490.e3. https://doi-org.libproxy.temple.edu/10.1016/j.amjmed.2018.05.021

Johansen, K. L., & Lee, C. (2015). Body composition in chronic kidney disease. Current opinion in nephrology and hypertension, 24(3), 268–275. https://doi-org.libproxy.temple.edu/10.1097/MNH.0000000000000120

Johnson, R. J., Bakris, G. L., Borghi, C., Chonchol, M. B., Feldman, D., Lanaspa, M. A., Merriman, T. R., Moe, O. W., Mount, D. B., Sanchez Lozada, L. G., Stahl, E., Weiner, D. E., & Chertow, G. M. (2018). Hyperuricemia, Acute and Chronic Kidney Disease, Hypertension, and Cardiovascular Disease: Report of a Scientific Workshop Organized by the National Kidney Foundation. American journal of kidney diseases : the official journal of the National Kidney Foundation, 71(6), 851–865. https://doi-org.libproxy.temple.edu/10.1053/j.ajkd.2017.12.009

Kramer, H., & Shoham, D. (2019). The Millennial Physician and the Obesity Epidemic: A Tale of Sugar-Sweetened Beverages. Clinical journal of the American Society of Nephrology : CJASN, 14(1), 4–6. https://doi-org.libproxy.temple.edu/10.2215/CJN.13851118

Lew, Q.-L. J., Jafar, T. H., Jin, A., Yuan, J.-M., & Koh, W.-P. (2018). Consumption of Coffee but Not of Other Caffeine-Containing Beverages Reduces the Risk of End-Stage Renal Disease in the Singapore Chinese Health Study. Journal of Nutrition, 148(8), 1315–1322. https://doi-org.libproxy.temple.edu/10.1093/jn/nxy075

Massey, L. K., & Sutton, R. A. (2004). Acute caffeine effects on urine composition and calcium kidney stone risk in calcium stone formers. The Journal of urology, 172(2), 555–558. https://doi-org.libproxy.temple.edu/10.1097/01.ju.0000129413.87024.5c

Matsushita K, Mahmoodi BK, Woodward M, Emberson JR, Jafar TH, Jee SH, Polkinghorne KR, Shankar A, Smith DH, Tonelli M, Warnock DG, Wen CP, Coresh J, Gansevoort RT, Hemmelgarn BR, Levey AS, Matsushita, K., Mahmoodi, B. K., Woodward, M., & Emberson, J. R. (2012). Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. JAMA: Journal of the American Medical Association, 307(18), 1941–1951. https://doi-org.libproxy.temple.edu/10.1001/jama.2012.3954

Michels WM, Grootendorst DC, Verduijn M, Elliott EG, Dekker FW, & Krediet RT. (2010). Performance of the Cockcroft-Gault, MDRD, and new CKD-EPI formulas in relation to GFR, age, and body size. Clinical Journal of the American Society of Nephrology, 5(6), 1003–1009. https://doi-org.libproxy.temple.edu/10.2215/CJN.06870909

McKenzie, K. A., El Ters, M., Torres, V. E., Harris, P. C., Chapman, A. B., Mrug, M., Rahbari-Oskoui, F. F., Bae, K. T., Landsittel, D. P., Bennett, W. M., Yu, A., & Mahnken, J. D. (2018). Relationship between caffeine intake and autosomal dominant polycystic kidney disease progression: a retrospective analysis using the CRISP cohort. BMC nephrology, 19(1), 378. https://doi-org.libproxy.temple.edu/10.1186/s12882-018-1182-0

National Health and Nutrition Examination Survey. (2015). Measuring Guides for the Dietary Recall Interview. https://www.cdc.gov/nchs/nhanes/measuring_guides_dri/measuringguides.htm

National Health and Nutrition Examination Survey. (2017). NHANES 2017-2018 Questionnaire Instruments. https://wwwn.cdc.gov/Nchs/Nhanes/continuousnhanes/questionnaires.aspx?BeginYear=2017

National Health and Nutrition Examination Survey. (2018). Body Composition

Procedures Manual. https://wwwn.cdc.gov/nchs/data/nhanes/2017-2018/manuals/Body_Composition_Procedures_Manual_2018.pdf

National Health and Nutrition Examination Survey. (2020). Laboratory Procedure Manual Blood Pressure Procedures Manual. https://wwwn.cdc.gov/nchs/data/nhanes/2019-2020/manuals/2019-Blood-Pressure-Procedures-Manual-508.pdf

National Health and Nutrition Examination Survey. (2020). Laboratory Procedure Manual. https://wwwn.cdc.gov/nchs/data/nhanes/2017-2018/labmethods/BIOPRO-J-MET-Creatinine-508.pdf

National Health and Nutrition Examination Survey. (2021a). Module 3: Weighting. https://wwwn.cdc.gov/nchs/nhanes/tutorials/module3.aspx

National Health and Nutrition Examination Survey. (2021b). Module 4: Variance Estimation. https://wwwn.cdc.gov/nchs/nhanes/tutorials/module4.aspx

Nilnumkhum, A., Kanlaya, R., Yoodee, S., & Thongboonkerd, V. (2019). Caffeine inhibits hypoxia-induced renal fibroblast activation by antioxidant mechanism. Cell adhesion & migration, 13(1), 260–272. https://doi-org.libproxy.temple.edu/10.1080/19336918.2019.1638691

Peerapen, P., & Thongboonkerd, V. (2016). Caffeine prevents kidney stone formation by translocation of apical surface annexin A1 crystal-binding protein into cytoplasm: In vitro evidence. Scientific reports, 6, 38536. https://doi-org.libproxy.temple.edu/10.1038/srep38536

Peerapen, P., & Thongboonkerd, V. (2018). Caffeine in Kidney Stone Disease: Risk or Benefit?. Advances in nutrition (Bethesda, Md.), 9(4), 419–424. https://doi-org.libproxy.temple.edu/10.1093/advances/nmy016

Rebholz, C. M., Young, B. A., Katz, R., Tucker, K. L., Carithers, T. C., Norwood, A. F., & Correa, A. (2019). Patterns of Beverages Consumed and Risk of Incident Kidney Disease. Clinical journal of the American Society of Nephrology : CJASN, 14(1), 49–56. https://doi-org.libproxy.temple.edu/10.2215/CJN.06380518

Rehm, C. D., Ratliff, J. C., Riedt, C. S., & Drewnowski, A. (2020). Coffee Consumption among Adults in the United States by Demographic Variables and Purchase Location: Analyses of NHANES 2011-2016 Data. Nutrients, 12(8), 2463. https://doi.org/10.3390/nu12082463

Rooney, A. A., Boyles, A. L., Wolfe, M. S., Bucher, J. R., & Thayer, K. A. (2014). Systematic review and evidence integration for literature-based environmental health science assessments. Environmental health perspectives, 122(7), 711–718. https://doi-org.libproxy.temple.edu/10.1289/ehp.1307972

Silva Junior, G. B., Bentes, A. C., Daher, E. F., & Matos, S. M. (2017). Obesity and kidney disease. Jornal brasileiro de nefrologia : 'orgao oficial de Sociedades Brasileira e Latino-Americana de Nefrologia, 39(1), 65–69. https://doi-org.libproxy.temple.edu/10.5935/0101-2800.20170011

Strippoli, G. F., Craig, J. C., Rochtchina, E., Flood, V. M., Wang, J. J., & Mitchell, P. (2011). Fluid and nutrient intake and risk of chronic kidney disease. Nephrology (Carlton, Vic.), 16(3), 326–334. https://doi-org.libproxy.temple.edu/10.1111/j.1440-1797.2010.01415.x

Sun, Y., Wang, D., & Zhou, Q. (2020). Caffeine intake and the risk of recurrent kidney stones in adults, an analysis of 2007-2014 National Health and Nutrition Examination Surveys. European journal of nutrition, 59(6), 2683–2692. https://doi-org.libproxy.temple.edu/10.1007/s00394-019-02115-0

Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. Lancet (London, England), 389(10075), 1238–1252. https://doi-org.libproxy.temple.edu/10.1016/S0140-6736(16)32064-5

Wijarnpreecha, K., Thongprayoon, C., Thamcharoen, N., Panjawatanan, P., & Cheungpasitporn, W. (2017). Association of coffee consumption and chronic kidney

disease: A meta-analysis. International journal of clinical practice, 71(1), 10.1111/ijcp.12919. https://doi-org.libproxy.temple.edu/10.1111/ijcp.12919

APPENDIX

2013-2014 NHANES Dataset SAS Code

```
*****************************************************************;
*File name: Data_Test_2013_2014.sas                              ;
*Created by: Hiroshi James Palomares Inuzuka                     ;
*Created on: 1/9/2021                                           ;
*Last Modified by: 6/15/2021                                    ;
*Last Modified: Hiroshi James Palomares Inuzuka                  ;
*Purpose: Testing data from NNS
          ;
*Datasets used: NHANES data for caffeine                        ;
*****************************************************************;

/* Assigning libnames in SAS */
libname Test3 'G:\College\Graduate School\Thesis\Datasets\Data 2013-2014';

/* Convert XPT file to SAS */
libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\DR2TOT_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;
```

```
libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\DEMO_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\BIOPRO_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\DIQ_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\BMX_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\BPX_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
```

```
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\RHQ_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2013-
2014\KIQ_U_H.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test3;
run;

/* Creating SAS Datasets */
data Test3.Dr2tot;
        set Test3.DR2TOT_H;
run;

data Test3.Demo;
        set Test3.DEMO_H;
run;

data Test3.Biopro;
        set Test3.BIOPRO_H;
```

```
run;

data Test3.DIQ;
        set Test3.DIQ_H;
run;

data Test3.BMX;
        set Test3.BMX_H;
run;

data Test3.BPX;
        set Test3.BPX_H;
run;

data Test3.RHQ;
        set Test3.RHQ_H;
run;

data Test3.KIQ;
        set Test3.KIQ_U_H;
run;

/* Merging SAS Datasets based on Dietary information */
```

```
data Test3.Combic;
	merge Test3.Dr2tot(in=a) Test3.Demo(in=b) Test3.Biopro(in=c) Test3.KIQ(in=d)
			Test3.DIQ(in=e) Test3.BMX_H(in=f) Test3.BPX_H(in=g)
Test3.RHQ(in=h);
	by seqn;
	if a then
		output;
run;

/* Setting Domains for the Age groups necessary in data */
data Test3.Combic;
	set Test3.Combic;
	if ridageyr<18 then age=0;
		else if 18=<ridageyr=<55 then age=1;
		else if ridageyr>55 then age=0;
	label age='Age Group Splitting';
run;

/* Creating eGFR values in combined NHANES Data */
data Test3.Combic;
	set Test3.Combic;
		if RIAGENDR=2 then
			do;
```

```
            a=0.742;
            e=1.018;
        end;
else
        do;
            a=1;
            e=1;
end;
if RIDRETH3=4 then
        do;
            b=1.212;
            f=1.159;
        end;
else
        do;
            b=1;
            f=1;
end;
if RIAGENDR=2 then
        do;
            c=0.7;
            d=-0.329;
        end;
```

```
else
        do;
                c=0.9;
                d=-0.411;
        end;
if RIDEXPRG='1' or RIDEXPRG='3' then
        do;
                h=0;
        end;
else
        do;
                h=1;
        end;
if DIQ010='1' or DIQ010='3' or DIQ010='7' then
        do;
                i=0;
        end;
else
        do;
                i=1;
        end;
if BMXBMI>='40' then
        do;
```

```
                g=0;
            end;
        else
            do;
                g=1;
            end;

data Test3.Combic;
    set Test3.Combic;
        eGFRCKDmin = 141 * ((LBXSCR)/c)** d * (0.993) ** (RIDAGEYR) *
e * f;
        eGFRCKDmax = 141 * ((LBXSCR)/c) ** (-1.209) * (0.993) **
(RIDAGEYR) * e * f;
run;
```

2015-2016 NHANES Dataset SAS Code

```
**************************************************************;
*File name: Data_Test_2015_2016.sas                          ;
*Created by: Hiroshi James Palomares Inuzuka                 ;
*Created on: 1/9/2021                                       ;
*Last Modified by: 6/15/21                                  ;
*Last Modified: Hiroshi James Palomares Inuzuka              ;
```

```
*Purpose: Testing data from NNS
          ;
*Datasets used: NHANES data for caffeine                    ;
*************************************************************;

/* Assigning libnames in SAS */
libname Test2 'G:\College\Graduate School\Thesis\Datasets\Data 2015-2016';

/* Convert XPT file to SAS */
libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\DR2TOT_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\DEMO_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\BIOPRO_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;
```

```
libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\DIQ_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\BMX_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\BPX_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\RHQ_I.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test2;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2015-
2016\KIQ_U_I.xpt' access=readonly;
```

```
proc copy inlib=xpttest outlib=Test2;
run;

/* Creating SAS Datasets */
data Test2.Dr2tot;
        set Test2.DR2TOT_I;
run;

data Test2.Demo;
        set Test2.DEMO_I;
run;

data Test2.Biopro;
        set Test2.BIOPRO_I;
run;

data Test2.DIQ;
        set Test2.DIQ_I;
run;

data Test2.BMX;
        set Test2.BMX_I;
run;
```

```
data Test2.BPX;
        set Test2.BPX_I;
run;

data Test2.RHQ;
        set Test2.RHQ_I;
run;

data Test2.KIQ;
        set Test2.KIQ_U_I;
run;

/* Merging SAS Datasets based on Dietary information */
data Test2.Combib;
        merge Test2.Dr2tot(in=a) Test2.Demo(in=b) Test2.Biopro(in=c)
Test2.KIQ_U_I(in=d)
                        Test2.DIQ(in=e) Test2.BMX_I(in=f) Test2.BPX_I(in=g)
Test2.RHQ(in=h);
        by seqn;
        if a then
                output;
run;
```

```
/* Setting Domains for the Age groups necessary in data */
data Test2.Combib;
	set Test2.Combib;
	if ridageyr<18 then age=0;
		else if 18=<ridageyr=<55 then age=1;
		else if ridageyr>55 then age=0;
	label age='Age Group Splitting';
run;

/* Creating eGFR values in combined NHANES Data */
data Test2.Combib;
	set Test2.Combib;
		if RIAGENDR=2 then
			do;
				a=0.742;
				e=1.018;
			end;
		else
			do;
				a=1;
				e=1;
		end;
```

```
if RIDRETH3=4 then
        do;
                b=1.212;
                f=1.159;
        end;
else
        do;
                b=1;
                f=1;
end;
if RIAGENDR=2 then
        do;
                c=0.7;
                d=-0.329;
        end;
else
        do;
                c=0.9;
                d=-0.411;
        end;
if RIDEXPRG='1' or RIDEXPRG='3' then
        do;
                h=0;
```

```
        end;
    else
        do;
            h=1;
        end;
    if DIQ010='1' or DIQ010='3' or DIQ010='7' then
        do;
            i=0;
        end;
    else
        do;
            i=1;
        end;
    if BMXBMI>='40' then
        do;
            g=0;
        end;
    else
        do;
            g=1;
        end;

data Test2.Combib;
```

```
        set Test2.Combib;
                eGFRCKDmin = 141 * ((LBXSCR)/c)** d * (0.993) ** (RIDAGEYR) *
e * f;
                eGFRCKDmax = 141 * ((LBXSCR)/c) ** (-1.209) * (0.993) **
(RIDAGEYR) * e * f;
run;
```

2017-2018 NHANES Dataset SAS Code

```
*****************************************************************;
*File name: Data_Test_2017_2018.sas                    ;
*Created by: Hiroshi James Palomares Inuzuka           ;
*Created on: 09/23/2020                               ;
*Last Modified by: 6/15/2021                          ;
*Last Modified: Hiroshi James Palomares Inuzuka        ;
*Purpose: Testing data from NNS
        ;
*Datasets used: NHANES data for caffeine                        ;
*****************************************************************;

/* Assigning libnames in SAS */
libname Test1 'G:\College\Graduate School\Thesis\Datasets\Data 2017-2018';

/* Convert XPT file to SAS */
```

```
libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\DR2TOT_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\DEMO_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\BIOPRO_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\DIQ_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\BMX_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
```

```
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\BPX_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\RHQ_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

libname xpttest xport 'G:\College\Graduate School\Thesis\Datasets\Data 2017-
2018\KIQ_U_J.xpt' access=readonly;
proc copy inlib=xpttest outlib=Test1;
run;

/* Creating SAS dataset for analysis */
data Test1.dr2tot;
        set Test1.DR2TOT_J;
run;

data Test1.Demo;
```

```
        set Test1.DEMO_J;
run;

data Test1.Biopro;
        set Test1.BIOPRO_J;
run;

data Test1.DIQ;
        set Test1.DIQ_J;
run;

data Test1.BMX;
        set Test1.BMX_J;
run;

data Test1.BPX;
        set Test1.BPX_J;
run;

data Test1.RHQ;
        set Test1.RHQ_J;
run;
```

```
data Test1.KIQ;
        set Test1.KIQ_U_J;
run;

/* Merging Datasets */
data Test1.Combi;
        merge Test1.dr2tot(in=a) Test1.Demo(in=b) Test1.Biopro(in=c) Test1.DIQ(in=d)
                        Test1.BMX(in=e) Test1.BPX(in=f) Test1.RHQ(in=g)
Test1.KIQ(in=h);
        by seqn;
        if a then
                output;
run;

/* Setting Domains for the Age groups necessary in data */
data Test1.Combi;
        set Test1.Combi;
        if ridageyr<18 then age=0;
                else if 18=<ridageyr=<55 then age=1;
                else if ridageyr>55 then age=0;
        label age='Age Group Splitting';
run;
```

```
/* Creating eGFR values in combined NHANES Data */
data Test1.Combi;
	set Test1.Combi;
		if RIAGENDR=2 then
			do;
				a=0.742;
				e=1.018;
			end;
		else
			do;
				a=1;
				e=1;
		end;
		if RIDRETH3=4 then
			do;
				b=1.212;
				f=1.159;
			end;
		else
			do;
				b=1;
				f=1;
		end;
```

```
if RIAGENDR=2 then
        do;
                c=0.7;
                d=-0.329;
        end;
else
        do;
                c=0.9;
                d=-0.411;
        end;
if RIDEXPRG='1' or RIDEXPRG='3' then
        do;
                h=0;
        end;
else
        do;
                h=1;
        end;
if DIQ010='1' or DIQ010='3' or DIQ010='7' then
        do;
                i=0;
        end;
else
```

```
                do;
                        i=1;
                end;
        if BMXBMI>='40' then
                do;
                        g=0;
                end;
        else
                do;
                        g=1;
                end;

data Test1.Combi;
        set Test1.Combi;
                eGFRCKDmin = 141 * (((LBXSCR)/c)** d) * (0.993) ** (RIDAGEYR) *
e * f;
                eGFRCKDmax = 141 * (((LBXSCR)/c) ** (-1.209)) * (0.993) **
(RIDAGEYR) * e * f;
run;
```

Final Regression Dataset SAS Code

```
**************************************************************;
*File name: Data_Regression.sas                              ;
```

```
*Created by: Hiroshi James Palomares Inuzuka                ;
*Created on: 12/20/2020                                     ;
*Last Modified by: 7/7/2021                                 ;
*Last Modified: Hiroshi James Palomares Inuzuka             ;
*Purpose: Data from NHANES for Analysis
          ;
*Datasets used: NHANES data for caffeine                    ;
*****************************************************************;

/* Assigning libnames in SAS */
libname Caffeine 'G:\College\Graduate School\Thesis\Datasets\Final Dataset';
libname First 'G:\College\Graduate School\Thesis\Datasets\Data 2017-2018';
libname Second 'G:\College\Graduate School\Thesis\Datasets\Data 2015-2016';
libname Third 'G:\College\Graduate School\Thesis\Datasets\Data 2013-2014';

/* Merging Datasets */
data Analysis;
          merge First.combi Second.combib Third.combic;
          by seqn;
run;

/* Averaging values */
data Analysis;
```

```
        set Analysis;
                WTR2TOT = mean(DR2_320Z,DR2_330Z);
                BPXSYAvg = mean(BPXSY1,BPXSY2,BPXSY3,BPXSY4);
                BPXDIAvg = mean(BPXDI1,BPXDI2,BPXDI3,BPXDI4);
run;

/* Excluding specific values for total */
/* Values for exclusion include age(for 18-55), RIDEXPRG(h)(pregnancy),
DIQ010(i)(diabetes), BMXBMI(g)(BMI over 40) */
data Analysis;
        set Analysis;
                if DR2TALCO>0 then /* DR2TALCO is for alcohol intake */
                        do;
                                j=0;
                        end;
                else
                        do;
                                j=1;
                        end;
                if KIQ022=1 then /* KIQ022 is for survey person being told they have a
weakened/failing kidney */
                        do;
                                k=0;
```

```
        end;
    else
        do;
            k=1;
        end;
    if BPXSYAvg>=130 then
        do;
            SYS=0;
        end;
    else
        do;
            SYS=1;
        end;
    if BPXDIAvg>=80 then
        do;
            DIA=0;
        end;
    else
        do;
            DIA=1;
        end;

/* Creating new weight from combine data */
```

```
data Analysis;
	set Analysis;
		if sddsrvyr in (8,9,10) then MEC6YR = 1/3 * WTMEC2YR;
run;

/* Double Checking the Total Value in the 18-55 Age Range */
proc freq data=Analysis;
	table Include;
run;

/* Creating the Surveyreg Domain Variable */
data Analysis;
	set Analysis;
		if age=1 and h=1 and i=1 and g=1 and j=1 and k=1 and SYS=1 and
DIA=1 and DR2TCAFF>=0 and eGFRCKDmax>=0
		then Include=1;
		else Include=0;
run;

/* Univariate used for finding Quartile data within the 18-55 Age Range */
proc univariate data=Analysis;
	weight MEC6YR;
	var eGFRCKDmax;
```

```
        where include=1;
        title 'Univariate of Variable needed';
run;

/* Stratifying the data from their Quartiles or separate Age groups by factors of 5 */
data Analysis;
        set Analysis;
                if 0 <= DR2TCAFF =< 5  then l=1;
                        else if 5 < DR2TCAFF =< 83 then l=2;
                        else if 83 < DR2TCAFF =< 192 then l=3;
                        else if DR2TCAFF > 192 then l=4;
                if 0 <= BMXBMI =< 23 then BMIQuat=1;
                        else if 23 < BMXBMI =< 26.5 then BMIQuat=2;
                        else if 26.5 < BMXBMI =< 30.9 then BMIQuat=3;
                        else if BMXBMI > 30.9 then BMIQuat=4;
                if 0 <= WTR2TOT =< 253.5 then WTRCot=1;
                        else if 253.5 < WTR2TOT =< 705 then WTRCot=2;
                        else if 705 < WTR2TOT =< 1312.5 then WTRCot=3;
                        else if WTR2TOT > 1312.5 then WTRCot=4;
                if 0 <= DR2TSUGR =< 53.66 then SUGQuart=1;
                        else if 53.66 < DR2TSUGR =< 89.73 then SUGQuart=2;
                        else if 89.73 < DR2TSUGR =< 134.77 then SUGQuart=3;
                        else if DR2TSUGR > 134.77 then SUGQuart=4;
```

```
if 0 <= LBXSUA =< 4.1 then UR=1;
        else if 4.1 < LBXSUA =< 5 then UR=2;
        else if 5 < LBXSUA =< 5.9 then UR=3;
        else if LBXSUA > 5.9 then UR=4;
if 0 =< DR2TSODI =< 2169 then SODI=1;
        else if 2169 < DR2TSODI =< 3041 then SODI=2;
        else if 3041 < DR2TSODI =< 4148 then SODI=3;
        else if DR2TSODI > 4148 then SODI=4;
if 0 =< BPXSYAvg =< 105.33 then SYSL=1;
        else if 105.33 < BPXSYAvg =< 112 then SYSL=2;
        else if 112 < BPXSYAvg =< 118 then SYSL=3;
        else if BPXSYAvg > 118 then SYSL=4;
if 0 =< BPXDIAvg =< 62.67 then DIAL=1;
        else if 62.67 < BPXDIAvg =< 68 then DIAL=2;
        else if 68 < BPXDIAvg =< 73.33 then DIAL=3;
        else if BPXDIAvg > 73.33 then DIAL=4;
if 0 =< eGFRCKDmax =< 93.92 then eGFRmax=1;
        else if 93.92 < eGFRCKDmax =< 108.86 then eGFRmax=2;
        else if 108.86 < eGFRCKDmax =< 126.27 then eGFRmax=3;
        else if eGFRCKDmax > 126.27 then eGFRmax=4;
if 18 =< RIDAGEYR =<40 then Split=1;
        else if 40 < RIDAGEYR =< 55 then Split=2;
run;
```

```
/* Standard Surveymeans with all boxplot values */
proc surveymeans data=Analysis varmethod=taylor nomcar;
        strata sdmvstra;
        cluster sdmvpsu;
        domain Include*Split*RIAGENDR*l;
        var  /*DR2TCAFF BMXBMI DR2TSUGR WTR2TOT LBXSUA DR2TSODI
BPXSYAvg BPXDIAvg*/ eGFRCKDmax;
        weight MEC6YR;
        title 'Survey Means Data from Merged Dataset';
run;

/* Surveygreg for correlation between Caffeine eGFR and other values */
/* Simple Variable Regression */
proc surveyreg data=Analysis;
        domain Include;
        weight MEC6YR;
        cluster sdmvpsu;
        model eGFRCKDmax = DR2TCAFF / solution;
        strata sdmvstra;
        title ‘Simple Variable Regression of eGFRCKD maximum value with Caffeine';
run;
```

```
/* Scatter Plot */
proc sgplot data=Analysis;
	scatter y=eGFRCKDmax x=DR2TCAFF;
	where DR2TCAFF<=2000;
	title 'Simple Variable Scatter Plot of eGFRCKD maximum value with Caffeine';
run;

/* Assessment Regression, Swapping omitted parameters to check for Covariance with
Beta Coefficient */
proc surveyreg data=Analysis;
	domain Include;
	weight MEC6YR;
	cluster sdmvpsu;
	model eGFRCKDmax = DR2TCAFF /*DR2TSUGR WTR2TOT LBXSUA
BMXBMI BPXSYAvg BPXDIAvg DR2TSODI ridageyr*/ RIAGENDR / solution;
	strata sdmvstra;
	title ‘Assessment Regression of eGFRCKD maximum value with Caffeine';
run;

/* Multivariate Regressions */
/* surveyreg for correlation between Caffeine eGFR and other values */
proc surveyreg data=Analysis;
	domain Include;
```

```
        weight MEC6YR;
        strata sdmvstra;
        cluster sdmvpsu;
        model eGFRCKDmax = DR2TCAFF ridageyr / solution;
        title 'Linear Regression Data from Merged Dataset eGFR Max with Confounding
Variables';
run;

proc surveyreg data=Analysis;
        domain Include;
        weight MEC6YR;
        strata sdmvstra;
        cluster sdmvpsu;
        model eGFRCKDmax = DR2TCAFF ridageyr / solution;
        where ridageyr>=40;
        title 'Linear Regression Data from Merged Dataset eGFR Max with Confounding
Variables where the Age is 40-55';
run;

proc surveyreg data=Analysis;
        domain Include;
        weight MEC6YR;
        strata sdmvstra;
```

```
        cluster sdmvpsu;
        model eGFRCKDmax = DR2TCAFF ridageyr / solution;
        where ridageyr<40;
        title 'Linear Regression Data from Merged Dataset eGFR Max with Confounding
Variables where the Age is 18-40';
run;


proc surveyreg data=Analysis;
        domain Include;
        weight MEC6YR;
        strata sdmvstra;
        cluster sdmvpsu;
        model eGFRCKDmax = DR2TCAFF ridageyr / solution;
        where DR2TCAFF<2000;
        title 'Linear Regression Data from Merged Dataset eGFR Max with Confounding
Variable where Caffeine Consumption is below
2000 mg';
run;
```