Demographic Predictors and Biomarkers of Vascular Injury Associated with Human Cytomegalovirus Infection

Jennifer Nicole Styles

A Master's technical report submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Public Health in the Department of Environmental Sciences and Engineering in the Gillings School of Global Public Health.

Chapel Hill
2016

Approved by:

Leena Nylander-French

Timothy J. Wade

Jill Stewart

ABSTRACT

Jennifer Nicole Styles: Demographic Predictors and Biomarkers of Vascular Injury Associated with Human Cytomegalovirus Infection
(Under the direction of Leena Nylander-French)

Human Cytomegalovirus (HCMV) infects between 50-80% of the adult population in the United States (US). We investigated the demographic predictors of HCMV Immunoglobulin G (IgG) seropositivity and the potential of HCMV IgG seropositive status to predict increased levels of vascular injury biomarkers, using a cross-sectional study. Both female and male participants (n=694) were recruited from Chapel Hill, NC and the surrounding area. HCMV IgG and four biomarkers of vascular injury, serum amyloid A (SAA), C-reactive protein (CRP), vascular cell adhesion molecule 1 (VCAM-1), and intercellular adhesion molecule 1 (ICAM-1) were analyzed using commercial enzyme-linked immunosorbent and sandwich electrochemiluminesent assays. Of the participants, 56.6% were HCMV IgG seropositive. HCMV IgG seropositivity was associated with increased body mass index, increased age, female gender, non-white or Hispanic ethnicity, and a history of smoking. HCMV IgG seropositivity was significantly associated with increased levels of vascular injury biomarkers ICAM-1 ($p=0.01$) and VCAM-1 ($p=0.0004$).

ACKNOWLEDGEMENTS

I would first like to acknowledge my advisor at UNC Professor Leena Nylander-French, PhD and my mentor Dr. Tim Wade from the U.S. Environmental Protection Agency (U.S. EPA). Tim and Leena, along with Associate Professor Jill Stewart, PhD are members of my committee and I would like to acknowledge them for their hard work and patience throughout this process. I would also like to acknowledge Dr. Andrey Egorov, Jo Klein, and Elizabeth Sams for their work with me at the U.S. EPA, guiding my analysis process and helping me complete laboratory work. I would also like to acknowledge Dr. Reagan Converse for her previous work on the SAFE study. Lastly, I would like to acknowledge my funding source from the U.S. EPA, Office of Research and Development (ORD), National Health and Environmental Effects Research Laboratory (NHEERL), Environmental Public Health Division (EPHD), Epidemiology branch.

This study has been funded by the U.S. EPA-UNC-CH Cooperative Training Agreement CR-83591401-0, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BD	Becton, Dickinson and Company, Franklin Lakes, NJ
BMI	Body Mass Index
CO	Carbon Monoxide
CRP	C-reactive protein
CVD	Cardiovascular Disease
CVD-related	Cardiovascular Disease Related
CDC	Centers for Disease Control
CAD	Coronary Artery Disease
CMV	Cytomegalovirus
ELISA	Enzyme-Linked Immunosorbent Assay
H. pylori	*Helicobacter pylori*
HCMV	Human Cytomegalovirus
HSF	Human Studies Facility
IgG	Immunoglobulin G
IgM	Immunoglobulin M
ICAM-1	Intercellular Adhesion Molecule 1
MSD	Meso Scale Discovery
ng/mL	Nanograms per Milliliter
NHANES	National Health and Nutrition Examination Surveys
NIEHS	National Institute of Environmental Health Sciences
NO_2	Nitrogen Oxide
NC	North Carolina
NF-kB	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
OD	Optical Density
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter with aerodynamic particle size of 2.5 μm
Q-Q	Quantile-Quantile

SAFE	Salivary Assay Feasibility Evaluation
SAA	Serum Amyloid A
SST	Serum Separation Tubes
TMB substrate	3,3’,5,5’-Tetramethylbenzidine
T. gondii	*Toxoplasma gondii*
TNF-α	Tumor Necrosis Factor - alpha
U.S. EPA	United States Environmental Protection Agency
VCAM-1	Vascular Cell Adhesion Molecule 1
VIP2	Vascular Injury Panel 2
WOCBA	Women of Child Bearing Age
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

History of Human Cytomegalovirus

Human Cytomegalovirus (HCMV) is a long-term latent virus that has been classified and studied under various evolving nomenclature for over a century. In a review of HCMV history, HD Riley surmised from Ribbert's 1904 original German publication that in 1881, Ribbert recognized enlarged intranuclear inclusion HCMV cells, but misclassified them as 'protozoan-like' in his publication (Riley, 1997). Ho *et al.* describes Ribbert as finding these 'protozoan-like' cells in the lungs, kidneys, and liver of a stillborn child and described the enlarged cells as having a "central nuclear body" and a clear halo surrounding the nuclear body (Ho, 2008). As early as 1921, there was some notion by Goodpasture, Talbot, and Lipschutz that HCMV was viral in nature (Riley, 1997). Goodpasture began using the term 'cytomegalia' after he noticed the enlarged nature of the infected cells (Riley, 1997). In 1932, Farber found HCMV in the salivary glands of 26 out of 183 children examined after their deaths (Farber & Wolbach, 1932). The amount of children that Farber found to be infected demonstrated the commonality of HCMV infection in infants (Farber et al., 1932). In 1952, the first case of HCMV was diagnosed by Fetterman from the urine of a suspected case following the discovery in 1932 by Wyatt that cytomegalic cells were found in the renal tubes of all 25 infants with lethal HCMV (Ho, 2008). 'Salivary gland virus' was also another early name for HCMV but it was abandoned when it became associated with a separate disease isolated in bats (Riley, 1997). In 1955, Margaret Smith was the first to notice that HCMV could only be cultured *in vitro* in human cells and not

cells from other species, a characteristic better understood later that HCMV was species specific (M. Ho 1997). HCMV was isolated by three different laboratories in 1956 after the advancement of cytology methods (Riley, 1997). In 1960, the term cytomegalovirus (CMV) was first used by TH Weller and has been adopted as the current terminology (Weller et. al. 1962). Weller isolated the Kerr Strain of HCMV from a 14-day old infant diagnosed with 'cytomegalic inclusion disease' (CID), an early name for HCMV (Weller et al., 1962). The first HCMV cells recovered from a live patient, referred to as the 'Davis strain', were found by Weller during his attempts to isolate Toxoplasma (Riley, 1997). Once isolated from tissue cultures, antibodies to HCMV were produced and used to demonstrate potential risks HCMV poses to a fetus (Riley, 1997).

HCMV Transmission

HCMV is a member of the herpes simplex virus family (Cannon et al., 2010; Lanzieri et al., 2015). HCMV is transmitted through contact with infected body fluids such as, blood, saliva, breast milk, and urine, including transmission through sexual contact and organ transplantation (CDC, 2016). In 1983, G. Knox hypothesized that sexual transmission of HCMV was likely not a major mode of transmission but that other contributing factors such as blood transfusions, contact with infected blood, vertical transmission from mother to child, and organ transplants were more common causes of transmission (Knox, 1983). Because of the estimated cost, both in quality life years and economic cost that is placed on the developed world, HCMV has become a high priority for vaccine development, but efforts since 1984 have so far proved fruitless (Schleiss, 2008; Schleiss, 2005).

Congenital Infection

HCMV has a high rate of prevalence among different populations worldwide (45-100%) and is a very common congenital infection, affecting 0.5-2% or 20,000 - 40,000 births yearly (Bialas et al., 2015; Schleiss, 2016). With these high rates of prevalence and infection, infants and very young children infected with HCMV make up 25% of all childhood hearing loss and causes the largest amount of non-genetic childhood birth defects, approximately 8,000 cases of permanent disability per year (Bialas et al., 2015). There is estimated to be between a 1% and 7% seroconversion in pregnant women (Hyde et al., 2010). Symptoms of congenital infection in infants includes sensorineural hearing loss, visual impairment, mental retardation, and cognitive defects, on top of the estimated 4% of infants who do not survive due to infection (Prince & Lapé-Nixon, 2014). Congenital HCMV infection is transmitted at much higher rates (30-40%) in women who get a primary infection during pregnancy versus women who have reactivated infections during pregnancy (~1%) (Prince et al., 2014; Schleiss, 2016).

Immunocompromised Patients

HIV/AIDS infected Individuals

G. Knox hypothesized in 1983 that HCMV might play a role, perhaps even preceding or causing, what was later to be discovered was HIV/AIDS (Knox, 1983). In reality, persons with HIV/AIDS suffer from activation of the opportunistic pathogen HCMV, since the immune system of the coinfected individual is unable to maintain the latency of infection and an acute active infection will persist without treatment with antiretrovirals (Crough & Khanna, 2009). The primary symptom of HCMV coinfection in HIV/AIDS infected patients is retinitis, which can lead to detached retinas and blindness, and represents 85% of coinfection cases (Biron, 2006; Crough et al., 2009). Three promising antiretrovirals, ganciclovir, foscarnet, and

cidofovir, have been used to treat HCMV infection mainly in immunocompromised patients, but benefits of treating congenital infection has also been studied in clinical trials (Schleiss, 2005).

Transplant Recipients

Transplant recipients taking immune suppressing drugs are at risk of suffering from HCMV reactivation. After transplantation, HCMV and other latent infections may be able to take advantage of the host's immune suppression and, thus, cause an active infection (Crough et al., 2009). An active infection may lead to increased risk of morbidity and mortality after organ transplant, particularly if the recipient is uninfected with HCMV and receives an HCMV infected organ (Crough et al., 2009). In both heart transplant vasculopathy and end-stage renal disease, HCMV was found to be a predictor of both increased risk of cardiovascular disease development and mortality/morbidity from other complications, including graft rejection (Betjes et al., 2007; Crough et al., 2009; Fateh-Moghadam et al., 2003).

Biology of HCMV

HCMV is a virus characterized by enlarged cells (Riley, 1997). HCMV is the largest of the herpes viruses at ~235 kilobases in length of double stranded DNA and 200-300 nanometers in diameter (Crough et al., 2009). HCMV infects human cells by transfusion or uptake processes (Crough et al., 2009). Once HCMV enters the cell and the viral envelope breaks down, HCMV is able to enter the nucleus of the cell where it can then replicate (Crough et al., 2009). After replication the virus is then spread throughout the body when the virion is re-enveloped in the cytoplasm and released from the cell via exocytosis (Crough et al., 2009). Reactivation of HCMV is poorly understood, but tumor necrosis factor -alpha (TNF-α) is suspected to play an important role, engaging latent cells and activating protein kinase C and nuclear factor kappa-

light-chain-enhancer of activated B cells (NF-kB) which leads to replication of the virus. HCMV commonly infects and can be detected in endothelial, smooth muscle cells, monocytes, macrophages, lymphocytes, immature dendritic cells, and bone marrow cells (Crough et al., 2009). Once infected with HCMV, immunoglobulin M (IgM) mounts a response to the infection and within 3-6 months IgM can no longer be detected (Prince et al., 2014). During this 3-6 month period, the immunoglobulin G (IgG) response is being established which will remain high throughout the life-long latency of the HCMV infection, and can increase four fold during periods of reinfection or reactivation (Prince & Lapé-Nixon, 2014). Recent evidence suggests that tests for IgM are sensitive but have low sensitivity when used to detect primary infection, which is vital information for expectant mothers (Prince & Lapé-Nixon, 2014). IgG avidity testing is currently being suggested as a more sensitive test for primary HCMV infection (Prince & Lapé-Nixon, 2014).

Descriptive Epidemiology

HCMV serostatus is tested as part of the National Health and Nutrition Examination Surveys (NHANES) conducted by Centers for Disease Control and Prevention (CDC). NHANES is a program, used by the CDC, to assess the health of the U.S. population by conducting personal interviews and physical examinations of participants (CDC/National Center for Health Statistics, 2015). The examination results are used to determine prevalence of diseases in the U.S. and influence U.S. health policy (CDC/National Center for Health Statistics, 2015). In the NHANES study conducted from 1988-2004, HCMV IgG seroprevalence was found to be associated with non-Hispanic Black individuals and Mexican Americans, older age, female sex, foreign birthplace, low household income, high household crowding, and low

household education (Bate et al., 2010). In women of childbearing age (12-49 years old, as defined by NHANES), IgM seropositive status in an IgG seropositive population was predicted by increasing age, and unmarried status (Wang et al., 2016). A review of HCMV demographic predictors indicated that increasing age, non-white or Hispanic race/ethnicity, female gender, and low socioeconomic status consistently predicted HCMV IgG seropositivity (Cannon et al., 2010). There is debate about the causative nature of HCMV on diseases of inflammation such as cancer, autoimmune and vascular disease, that may be the result of the virus's cellular and immunological defenses (Söderberg-Nauclér, 2006).

Cardiovascular Injury

The American Heart Association (AHA) defines cardiovascular disease (CVD) as a heart and blood vessel disease that is the result of atherosclerosis, the buildup of plaque in the walls of the arteries (American Heart Association, 2014). The buildup of plaque in the walls of arteries leads to a narrowing of the arteries, reducing blood flow and increasing the risk for heart attack or stroke (American Heart Association, 2014). According to the CDC, heart attack and stroke are among the leading causes of death in the U.S., heart attack being number one on the list (CDC, 2016).

It has been hypothesized that cardiovascular disease develops in HCMV infected individuals when the lifelong-latent infection is reactivated through inflammation and spreads to other cells in the body interrupting normal cellular processes (Söderberg-Nauclér, 2006). The CD4+ T-cells that are found in atherosclerosis plaques could have large effects on the plaque, but it is possible that HCMV is able to sustain the inflammation during latency of the virus causing the cellular damage (Söderberg-Nauclér, 2006). A proposed mechanism of HCMV's

enhancement of cardiovascular disease suggests that HCMV increases the conversion of prothrombin to thrombin, a clot forming agent, either directly or indirectly through increased inflammation (Popović et al., 2012). Thrombin is proposed to enhance macrophage adhesion to endothelial cells that line the vascular walls resulting in blood clots and thrombosis leading to heart attack and stroke (Popović et al., 2012). In both avian and rat species, CMV infection was found to correlate with thrombosis and atherosclerosis (Fabricant & Fabricant, 1999; Span et al., 1992). HCMV has been known to infect endothelial cells, epithelial cells, smooth muscle cells, and fibroblasts (Popović et al., 2012). Vascular cell adhesion molecule 1 (VCAM-1), and intercellular adhesion molecule 1 (ICAM-1), both proinflammatory adhesion molecules used to measure cardiovascular disease, have been observed to activate when endothelial cells were infected with HCMV (Popović et al., 2012). Association with HCMV infection among transplant recipients and increased risk for cardiovascular symptoms, including atherosclerotic disease, increased plaque thickness and area, and increased risk of transplant vasculopathy has been reported (Betjes et al., 2007; Fateh-Moghadam et al., 2003). In patients with end-stage renal disease, cardiovascular disease and C-reactive protein (CRP) were significantly associated with HCMV seropositivity (Betjes et al., 2007). In a study using NHANES data from 1988-1994, all-cause mortality and cardiovascular disease related (CVD-related) mortality increased (all-cause: $p=0.0358$, CVD-related $p=0.1092$) for individuals seropositive for HCMV (Simanek et al., 2011). The association between HCMV seropositivity and all-cause and CVD-related mortality was more pronounced among individuals with high CRP levels (all-cause: $p<0.0001$, CVD-related $p=0.0040$) (Simanek et al., 2011).

Common biomarkers of vascular injury include serum amyloid A (SAA), CRP, VCAM-1, and ICAM-1. CRP is the most commonly studied biomarker of coronary artery disease

(CAD), one of the diseases encompassed by the cardiovascular disease classification, but is a non-specific marker of inflammation that may indicate other serious health conditions such as cancer or lupus (Kaptogeet al., 2012; Zakynthinos et al., 2009). SAA and CRP are positive markers of acute inflammation, their levels increasing rapidly after injury (Baumann et al., 2017; Zakynthinos et al., 2009). SAA is formed primarily in the liver, but is also formed extrahepatically by macrophages and endothelial cells, and has been recognized as a biomarker that predicts the development of cardiovascular disease (Baumann et al., 2017). Elevated CRP levels were found to be predictive of first cardiovascular events in individuals who were at intermediate risk, but CRP is a non-specific maker of inflammation and increased levels may be an indication of a disease such as cancer or lupus (Kaptoge et al., 2012; Zakynthinos et al., 2009). Soluble Cellular Adhesion Molecules (CAMs), VCAM-1 and ICAM-1, are makers of cellular adhesion (Blankenberg et al, 2003; Zakynthinos et al., 2009). ICAM-1 and VCAM-1 are both factors in firm cellular adhesion and are released when the cells are activated during adhesion (Blankenberg et al., 2003). ICAM-1 can be found in both the endothelial and leukocyte cells, but VCAM-1 can only be found in endothelial cells (Blankenberg et al., 2003). VCAM-1 and CRP were found to be predictive of cardiovascular mortality (Zakynthinos & Pappa, 2009). CRP an ICAM-1 were found to predict risk of cardiovascular events (Zakynthinos & Pappa, 2009). Observing levels of these proteins in patients' serum may provide information about the level of cardiovascular injury that is occurring in their body, quantifying the level of clotting and plaque primarily with Cellular Adhesion Molecules, and acute injury with CRP and SAA.

Environmental Relevance

Development of cardiovascular disease has been associated with exposure to environmental air pollution (Brook, 2008). Environmental air pollution such as particulate matter (PM), nitrogen oxide (NO_2), and carbon monoxide (CO) have been studied for their association with many cardiovascular diseases including, thrombosis, endothelial dysfunction, stroke, and CVD-related mortality (Bind et al., 2013; Brook, 2008). In a study of elderly veterans, both VCAM-1 and ICAM-1 were associated with various markers of air pollution including particulate matter with aerodynamic particle size of 2.5 μm ($PM_{2.5}$), NO_2, and particle number (Bind et al., 2013). The association between HCMV infection and CVD development needs further investigation so that the effects of this common lifelong infection can be controlled for in future CVD studies exploring environmental air pollution's causal effect on CVD.

SPECIFIC AIMS

My hypothesis for this study is that HCMV IgG seropositivity is associated with increased levels of vascular injury biomarkers. This is a specific hypothesis to help answer the broader question of HCMV infection's role in the development of cardiovascular disease. To test my hypothesis, the primary aims of this study were to determine demographic predictors of HCMV IgG seropositive status and to test for associations between HCMV IgG response and increased markers of vascular injury. Demographic predictors of HCMV IgG seropositive status were determined for study participants and compared to possible demographic predictors of infection using responses to survey questionnaires and chi-squared and Student's t-test analyses. To test for associations between HCMV IgG response and increased levels of vascular injury biomarkers, HCMV IgG seropositive status, HCMV IgG ratio to plate specific control and levels of four common biomarkers of vascular injury were tested in serum. Using linear regression analysis with correction for variables, associations between HCMV IgG response and markers of vascular injury were assessed.

CHAPTER 2: MATERIALS & METHODS

Study Population

The serum samples used in this analysis were collected in a previously conducted cross-sectional study called Salivary Assay Feasibility Evaluation (SAFE) in which serum and saliva samples were collected from 696 volunteer participants recruited from the area surrounding Chapel Hill, North Carolina (NC). The SAFE study was launched to develop methods of detecting common infections including *Helicobacter pylori* (*H. pylori*) and *Toxoplasma gondii* (*T. gondii*) in serum and saliva. Participants were recruited using flyers located in places identified as likely to be visited by individuals who were IgG seropositive for *H. pylori* and *T. gondii* such as veterinary clinics, where exposure to the feline reproducing *T. gondii* is likely. Participants were included in the study if they could provide a saliva sample, but did not necessary fill out a survey or provide a blood sample (processed to and tested as serum). Participants in the SAFE study who filled out a survey were given privacy in a room while they responded to questions using a computer at the U.S. Environmental Protection Agency (U.S. EPA) Human Studies Facility (HSF), located in Chapel Hill, NC. A subset of the individuals participating in the SAFE study (n= 351) were recruited and participated in the study through a partnership with the National Institute of Environmental Health Sciences (NIEHS) at Research Triangle Park, NC. These participants were not recruited based on their likelihood to be infected but because of their willingness to participate in a concurrent NIEHS study, i.e., a convenience sampling. These participants filled out an abbreviated questionnaire (see Appendices B and C).

Survey Questionnaire

Survey participants were asked to self-report their demographic information (gender, race, education, etc.) and medical history (smoking, diabetes, depression, asthma, etc.). Participants were asked to self-report exposure to certain environmental exposure sources that are known vectors of *H. pylori* and *T. gondii* (soil exposure, animal contact, consumption of undercooked meat, etc.). Survey questionnaires administered in SAFE study and for the NIEHS subgroup can be found in Appendices A, B, and C.

Survey variables were created using coded responses next to the response boxes on the survey (see Appendices A, B, and C). Survey variables race, ethnicity, current smoker, and ever smoker, diabetes, depression, and asthma were summarized into dichotomous variables. Age was self-reported in whole-year increments. For the purposes of this study, a subgroup of "women of childbearing age" (WOCBA) was created to investigate the risk factors in this at-risk population. Women of childbearing age were defined as 15-49 years of age by the World Health Organization (WHO) (WHO, 2016). Because participants in this study were required to be a minimum of 18 years of age, the WOCBA population was comprised of women 18 to 49 years old. Body-mass index (BMI) was calculated from height and weight data and was analyzed as both a continuous variable and categorized using the CDC designations underweight, healthy weight, overweight, or obese for demographic predictor analysis (CDC, 2010). Gender was reported as female (0), male (1), or no response (NA). Race and ethnicity information was dichotomized into white non-Hispanic (1), non-white or Hispanic (0), or decline to answer/don't know (NA). Smoking data was collected from two questions capturing both current and past smoking status. Current smokers were asked to describe their smoking behavior stating if they smoked daily (2), less than every day (1), not at all (0), or decline to answer/don't know (NA).

Participants were asked to state if they had ever smoked, responding in the affirmative (1), negative (0), or decline to answer/don't know (NA). Current smokers were further dichotomized into current smoker (response 1 or 2), not at all (0), or decline to response/don't know (NA). Diabetes, depression, and asthma were self-reported based on a physician's diagnosis and participants were asked to respond with no (0), yes (1), or decline to answer/don't know (NA) for each separate diagnosis. Education data was collected from SAFE participants surveyed at the U.S. EPA HSF but not from the NIEHS subgroup and could unfortunately not be used in this study.

Sample Collection

Saliva and serum samples were collected from 696 study participants for the purposes of the SAFE study at both U.S. EPA and NIEHS locations. Whole blood samples were collected by trained phlebotomists into BD Vacutainer serum separation tubes (SST; Becton, Dickinson and Company, Franklin Lakes, NJ) coated with silica particles to accelerate clotting and a serum separator gel (BD, 2016). Serum was separated the same day the samples were collected from participants following manufacturer instruction and stored in cryogenic vials (in 2013) at -80°C until analysis (in 2016).

Sample Analysis

H. pylori and T. gondii

As part of the original study protocol, samples were tested for *T. gondii* and *H. pylori*. Samples were tested using anti-*T. gondii* IgG (cat#: EG 127, Viro-Immun Diagnostics GmbH, Germany) and anti-*H. pylori* IgG (product #:HpKi-GB, Micro Detect, Inc., Tustin, California,

distributed by Bio-Rad Laboratories, Hercules, California) *in vitro* enzyme-linked immunosorbent assays (ELISA). For the purposes of this study, *H. pylori* and *T. gondii* results were dichotomized seropositive (1) or seronegative (0) according to manufacturer's instructions.

HCMV

To test for the presence of IgG HCMV serum antibodies, serum samples were assayed using anti-HCMV IgG ELISA kit (cat#ab108724, Abcam, Cambridge, MA). This kit utilizes horseradish peroxidase, 3,3',5,5'-Tetramethylbenzidine (TMB) substrate, and an acidic stop solution to produce a yellow color whose density is directly proportional to the amount of HCMV IgG antibodies captured. The yellow color density was read using a calibrated spectrophotometer (SpectraMax model #340PC384, Molecular Devices, Sunnyvale, CA). According to the manufacturer's instructions, HCMV IgG optical density (OD) values read at 450 nm were reference wavelength (620 nm) and blank subtracted. After correction, HCMV IgG OD values were divided by the plate specific cutoff value, referred to as 'HCMV ratio'. The HCMV ratio was used to determine if a sample was seropositive or seronegative for HCVM IgG antibodies. HCMV ratios higher than 1.1 were considered seropositive, ratios between 0.9 and 1.1 were indeterminate, and ratios below 0.9 were considered seronegative.

Vascular Injury Panel

To quantify vascular injury in the study participant's, SAA, CRP, VCAM-1, and ICAM-1 protein levels were measured in serum using an assay developed on the Meso Scale Discovery (MSD) platform using a 4-spot 96-well plate. Vascular Injury Panel 2 (VIP2, cat# K15198D, MSD, Rockville, Maryland) plate was read using a MSD QuickPlex SQ 120 instrument. Results

from the MSD QuickPlex SQ 120 were used along with a standard curve to calculate concentration of SAA, CRP, VCAM-1, and ICAM-1 in nanograms per milliliter (ng/mL). In this assay, increase of any of the four biomarkers of vascular injury indicates an increase in vascular injury in a dose-response fashion (Zakynthinos & Pappa, 2009).

Data Analysis

R statistical analysis software (version 3.3.2) along with Microsoft Excel (2016) were used to analyze the survey and sample data. Base-R and R-packages were used to perform chi-squared analyses on dichotomous variables as well as appropriate Student's *t*-tests and linear regression analysis on continuous variables (Walker & Braglia, 2017; Wickham, 2016; Wickham et al., 2017; Wickham & RStudio, 2016). Independent two-group Student's *t*-tests were performed between binomial dichotomous variables and continuous variables. Associations between HCMV seropositive status and the study groups SAFE (samples collected at U.S. EPA) and NIEHS (samples collected at NIEHS), age, BMI, gender, race, smoking status, diabetes, depression, asthma, *H. pylori*, and *T. gondii* were analyzed using chi-squared and Student's *t*-tests. Shapiro-Wilks test was used to test if the data for each continuous variable were normally distributed. Data were $\log_{10}$-transformed if the $\log_{10}$-transformed data were more normally distributed (p-value closer to 1 and more symmetric Quantile-Quantile (Q-Q) plots) than the untransformed data based on the Shapiro-Wilks p-value and confirmed with Q-Q plots. The association between HCMV and each of the four markers of vascular injury (SAA, CRP, VCAM-1, ICAM-1) were investigated using linear regression analysis accounting for possible sources of confounding, variables BMI ($\log_{10}$-transformed), race, age, current smoking status, and gender. HCMV association was tested against the four markers of vascular injury (SAA,

CRP, VCAM-1, ICAM-1) using HCMV dichotomized (seropositive and seronegative) results, HCMV OD values (log_{10}HCMV ratio), and tertiary [seronegative (0), seropositive <33%, 33-67%, >67%] categorization of HCMV OD results, to confirm our findings through multiple methods of analysis. To exclude cardiovascular injury of non-HCMV related origin, HCMV seronegative results were set to 0 and used as a reference category for the log_{10}HCMV ratio and tertiary analyses that used continuous HCMV OD values. Anti-HCMV IgG response has been reported to increase during active or reactivated HCMV infection that can be quantified in the OD, plate specific-cutoff adjusted, HCMV results (Mehta et al., 2014). The four markers of vascular injury (SAA, CRP, VCAM-1, ICAM-1) were log_{10}-transformed for the analysis.

Table 1 summarizes the results of Shapiro-Wilks tests for normality. All data that was log_{10} transformed had p-values closer to 1 and thus were more normally distributed after log_{10} transformation.

Table 1. Shapiro-Wilks test for normality with raw and log-transformed data.

Raw Data				Log -Transformed Data		
Names	W-statistic	P-value		Names	W-statistic	P-value
CRP	0.42	2.1E-42		log_{10} CRP	1.00	9.5E-02
SAA	0.15	4.4E-48		log_{10} SAA	0.97	2.2E-10
VCAM-1	0.76	9.6E-31		log_{10} VCAM-1	0.98	6.5E-08
ICAM-1	0.83	2.6E-26		log_{10} ICAM-1	0.99	7.4E-06
HCMV ratio+	0.98	3.0E-06		log_{10} HCMV ratio+	0.98	1.7E-05
HCMV ratio-	0.83	1.9E-17		log_{10} HCMV ratio-	0.95	6.7E-09
BMI	0.92	3.4E-19		log_{10} BMI	0.98	5.2E-07

Q-Q plots were used to confirm the findings of the Shapiro-Wilks test for normality. Results of Q-Q plots can be seen in Figure 1. Q-Q plots show a more linear distribution after

continuous biomarker data have been $\log_{10}$ transformed, indicating a more normal distribution of data after $\log_{10}$ transformations.

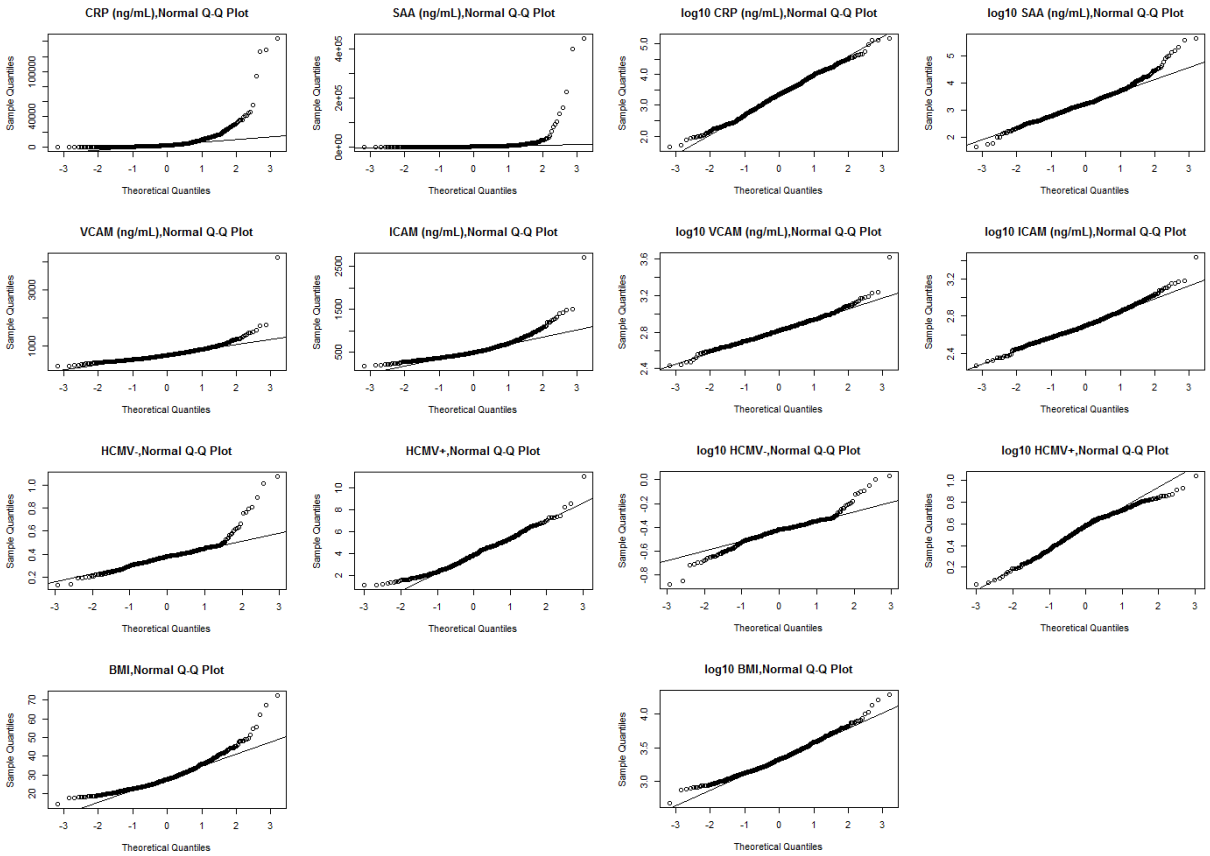


Figure 1. Q-Q Plots to confirm normal distribution as determined from the Shapiro-Wilks test for normality.

CHAPTER 3: RESULTS

Demographics

The demographic breakdown of the whole study cohort (SAFE and NIEHS subgroup) is summarized in Table 2. The study cohort included a total of 710 participants, of whom 694 were tested for both HCMV and the four vascular injury biomarkers for this study. Of the participants in this study, 62.3% were female (n = 435) and 37.7% were male (n = 263). The mean and median age of the participants was 40.8 and 40 years, respectively. The majority of participants reported their race/ethnicity as white non-Hispanic (63.5%; n = 377) while 36.5% (n = 217) self-identified as non-white or of Hispanic descent. Participants reported having health conditions and deleterious health related behaviors including current smoking (23.1%; n = 160), if participants had ever smoked (39.3%; n = 271), diabetes (8.2%; n = 57), depression (20.4%; n = 141), and asthma (14.7%; n = 102). Of the participants, 13.7% (n = 96) tested *H. pylori* IgG seropositive, 9.3% (n = 65) *T. gondii* IgG seropositive, and 56.5% (n = 393) HCMV IgG seropositive. One member of the study population reported having heart disease, a type of cardiovascular disease hypothesized to be associated with HCMV (Betjes et al., 2007; Fateh-Moghadam et al., 2003; Popović et al., 2012; Söderberg-Nauclér, 2006).

HCMV IgG demographic predictors for the total population are summarized in Table 3. In the total population (n = 710), significant associations were found between HCMV seropositivity and age [two-sample *t*-test, p = 7.32E-4, confidence interval (CI) = -5.81; -1.55], BMI (two-sample *t*-test, p = 1.05E-4, CI = -3.12; -1.03), gender (χ^2, p = 1.01E-3), race/ethnicity

(χ^2, p = 9.08E-14), smoker now (χ^2, p = 7.44E-6), smoker ever (χ^2, p = 8.72E-3), *H. pylori* (χ^2, p = 4.39E-7), and *T. gondii* (χ^2, p = 1.10E-2).

Table 2. Summary of the demographic descriptors and reported health conditions analyzed in this study for the SAFE study population (n = 710).

Population Demographics				
		N of Demographic Category	Sample Size*	% Total Pop.
Total Population			**710**	
Age			695	
	18-29	204		29.4%
	30-39	138		19.9%
	40-49	131		18.8%
	50-59	154		22.2%
	60-69	52		7.5%
	70-85	16		2.3%
Median Age	40			
Mean Age	40.8			
BMI			698	
	< 18.5	7		1.0%
	18.5–24.9	227		32.5%
	25.0–29.9	209		29.9%
	> 30	255		36.5%
Gender			698	
	Male	263		37.7%
	Female	435		62.3%
Race/Ethnicity			594	
	White	377		63.5%
	Other	217		36.5%
Smoke				
	smoke now	160	692	23.1%
	smoke ever	271	689	39.3%
Diabetes		57	695	8.2%
Depression		141	691	20.4%
Asthma		102	694	14.7%
H. pylori		96	702	13.7%
T. gondii		65	702	9.3%
HCMV		393	696	56.5%

** Samples with missing data were excluded from sample size*

Table 3. Demographic predictors of HCMV in infected and uninfected individuals.

Total Population, HCMV IgG Seropositive Status									
		HCMV Serostatus						χ^2 p-value	Significance
		Positive % (N)		Negative % (N)		Total % (N)			
Overall						**696**			
	NIEHS	38%	(131)	62%	(215)	100%	(346)	3.08E-03	**
	SAFE	49%	(172)	51%	(177)	100%	(349)		
Age								(two sample t-test)	
	18-29	55%	(113)	44%	(89)	100%	(202)		
	30-39	41%	(56)	59%	(82)	100%	(138)		
	40-49	33%	(43)	67%	(88)	100%	(131)	7.32E-04	***
	50-59	39%	(60)	60%	(93)	100%	(153)		
	60-69	46%	(24)	54%	(28)	100%	(52)		
	70-85	31%	(5)	69%	(11)	100%	(16)		
Median Age	40								
Mean Age	40.8								
BMI								(two sample t-test)	
Underweight	< 18.5	57%	(3)	43%	(2)	100%	(5)		
Healthy	18.5–24.9	53%	(56)	46%	(50)	100%	(106)		
Overweight	25.0–29.9	44%	(37)	56%	(45)	100%	(82)	1.05E-04	***
Obese	>= 30	34%	(26)	65%	(75)	100%	(101)		
Gender									
	Female	39%	(167)	61%	(265)	100%	(432)	1.01E-03	**
	Male	52%	(136)	48%	(127)	100%	(263)		
Race/Ethnicity									
	Other	27%	(58)	73%	(157)	100%	(215)	9.08E-14	***
	White	59%	(223)	41%	(154)	100%	(377)		
Smoke									
smoke now	no	48%	(257)	52%	(274)	100%	(531)	7.44E-06	***
	yes	28%	(44)	72%	(114)	100%	(158)		
smoke ever	no	48%	(199)	52%	(218)	100%	(417)	8.27E-03	**
	yes	37%	(100)	63%	(169)	100%	(269)		
Diabetes									
	no	45%	(284)	55%	(351)	100%	(635)	1.28E-01	
	yes	33%	(19)	67%	(38)	100%	(57)		
Depression									
	no	43%	(236)	57%	(312)	100%	(548)	5.35E-01	
	yes	46%	(65)	54%	(75)	100%	(140)		
Asthma									
	no	44%	(261)	56%	(329)	100%	(590)	5.66E-01	
	yes	41%	(41)	59%	(60)	100%	(101)		
H. pylori									
	seronegative	47%	(285)	53%	(316)	100%	(601)	4.39E-07	***
	seropositive	19%	(18)	81%	(76)	100%	(94)		
T. gondii									
	seronegative	45%	(285)	55%	(345)	100%	(630)	1.10E-02	*
	seropositive	28%	(18)	72%	(47)	100%	(65)		
HCMV									
	seronegative	100%	(303)	0%	(0)	100%	(303)	NA	
	seropositive	0%	(0)	100%	(393)	100%	(393)		

** indicates significance of χ^2 p-value: '.' p<0.1; '*' p<0.05; '**' p<0.01; '***' p <0.001. P-value significant at α <0.05.*

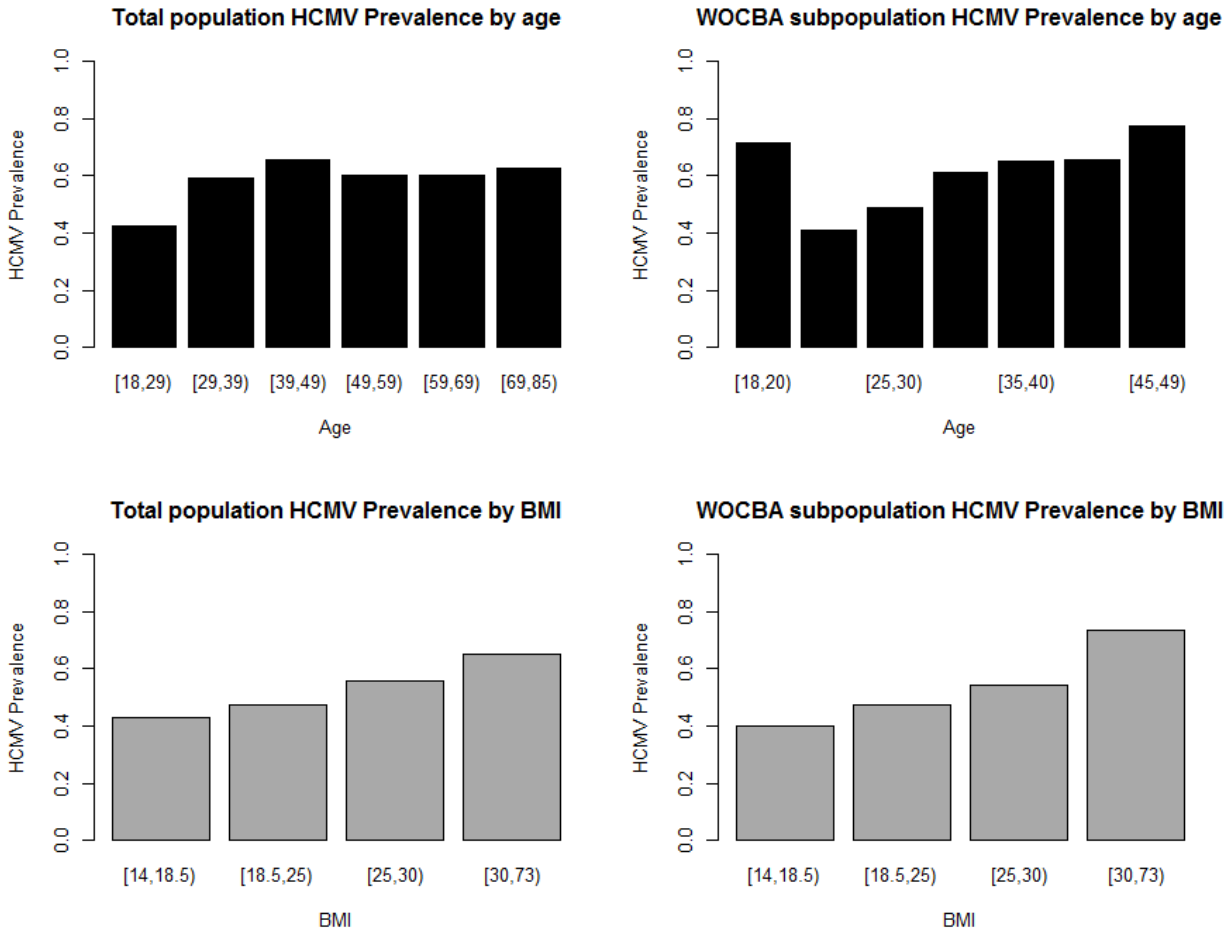
Demographic predictors of HCMV in the WOCBA subpopulation are summarized in Table 4. WOCBA subpopulation (n = 296) had similar significant associations with HCMV compared to the HCMV associations in the study population (n = 696) with the exception of *T. gondii*, which was not significantly associated with HCMV seropositive status ($p > 0.05$) in WOCBA. The strength of association for older age was higher with respect to HCMV seropositive status (study population $p = 7.3E-4$, WOCBA subpopulation $p = 6.4E-5$) between the study population (n = 696) and the WOCBA subpopulation (n = 296) (Tables 3 and 4).

Figure 2 shows the prevalence of HCMV seropositivity compared with age and BMI, both measured to be significant demographic predictors of HCMV infection (Tables 3 and 4). With increasing age or BMI, prevalence of HCMV infection increased (Figure 2).

Table 4. Demographic predictors of HCMV in infected and uninfected individuals for the WOCBA subpopulation.

Women of Child Bearing Age, HCMV IgG Seropositive Status									
		HCMV Serostatus							
		Positive % (N)		Negative % (N)		Total % (N)		χ^2 p-value	Significance
Overall					**296**				
	NIEHS	30%	(35)	70%	(81)	100%	(116)	2.21E-03	**
	SAFE	49%	(87)	51%	(91)	100%	(178)		
Age								(two sample t-test)	
	18-24	54%	(33)	44%	(27)	100%	(60)		
	25-29	50%	(35)	49%	(34)	100%	(69)		
	30-34	39%	(19)	61%	(30)	100%	(49)		
	35-39	35%	(13)	65%	(24)	100%	(37)	6.43E-05	***
	40-44	34%	(10)	66%	(19)	100%	(29)		
	45-49	23%	(11)	77%	(37)	100%	(48)		
Median Age	31								
Mean Age	32.4								
BMI								(two sample t-test)	
Underweight	< 18.5	44%	(3)	56%	(2)	100%	(5)		
healthy	18.5–24.9	53%	(56)	46%	(50)	100%	(106)		
overweight	25.0–29.9	44%	(37)	56%	(45)	100%	(82)	3.52E-04	***
obese	>= 30	34%	(26)	65%	(75)	100%	(101)		
Race/Ethnicity									
	Other	19.1%	(18)	80.9%	(76)	100%	(94)	1.27E-10	***
	White	61.8%	(97)	38.2%	(60)	100%	(157)		
Smoke									
smoke now	no	47.6%	(110)	52.4%	(121)	100%	(231)	5.67E-05	***
	yes	18.0%	(11)	82.0%	(50)	100%	(61)		
smoke ever	no	45.9%	(90)	54.1%	(106)	100%	(196)	4.70E-02	*
	yes	33.0%	(32)	67.0%	(65)	100%	(97)		
Diabetes									
	no	42.9%	(118)	57.1%	(157)	100%	(275)	1.03E-01	
	yes	21.1%	(4)	78.9%	(15)	100%	(19)		
Depression									
	no	40.4%	(92)	59.6%	(136)	100%	(228)	4.45E-01	
	yes	46.8%	(29)	53.2%	(33)	100%	(62)		
Asthma									
	no	42.7%	(103)	57.3%	(138)	100%	(241)	4.10E-01	
	yes	35.3%	(18)	64.7%	(33)	100%	(51)		
H. pylori									
	seronegative	45.8%	(119)	54.2%	(141)	100%	(260)	8.61E-05	***
	seropositive	8.8%	(3)	91.2%	(31)	100%	(34)		
T. gondii									
	seronegative	42.6%	(118)	57.4%	(159)	100%	(277)	1.95E-01	
	seropositive	23.5%	(4)	76.5%	(13)	100%	(17)		
HCMV									
	seronegative	100%	(122)	0%	(0)	100%	(122)	5.24E-65	***
	seropositive	0%	(0)	100%	(172)	100%	(172)		

** indicates significance of χ^2p-value: '.' p<0.1; '*' p<0.05; '**' p<0.01; '***' p <0.001. P-value significant at α <0.05.*



Study population (n = 696) is separated into 10-year age bins, while the WOCBA subpopulation is separated into 5-year age bins. Using BMI categories developed by the CDC as a measure of health, BMI values were categorized into underweight, healthy weight, overweight, and obese (CDC), 2010).

Figure 2. Age and BMI distributed prevalence of HCMV infection in the study population (n = 696) and in WOCBA subpopulation.

Vascular Injury and HCMV Biomarkers

Of the 710 samples from the SAFE population, 694 samples were tested for both HCMV IgG and biomarkers of vascular injury and thus comprise the study population for this part of the analysis. Linear regression models in Figure 3 indicate that without controlling for variables (log_{10}BMI, race/ethnicity, smoker, age, gender), there are positive correlations between log_{10}HCMV ratio to plate-specific cutoff and all four biomarkers of vascular injury. Scatter plots demonstrate that in seropositive individuals, as the unadjusted levels of log_{10}HCMV ratio increase, levels of log_{10}-adjusted biomarkers of vascular injury (CRP, SAA, VCAM-1, and ICAM-1) also increase (Figure 3). In seronegative individuals, the slope of the unadjusted association between log_{10}HCMV and biomarkers of vascular injury was not positive, or not as steep as the slopes among seropositive individuals (Figure 3).

HCMV seropositive subjects had significantly higher levels of ICAM-1 compared to their seronegative counterparts (Table 5). Dichotomized HCMV serostatus was a significant predictor of log_{10}-adjusted ICAM-1 but not log_{10}-adjusted CRP, SAA or VCAM-1 measured in serum (Table 5). Log_{10}BMI was the most significant independent variable for CRP and SAA ($p = 2.6E-37$ and $p = 7.7E-18$, respectively), race was the most significant independent variable for VCAM-1 ($p = 1.9E-7$), and age was the most significant independent variable for ICAM-1 ($p = 4.9E-5$) (Table 5).

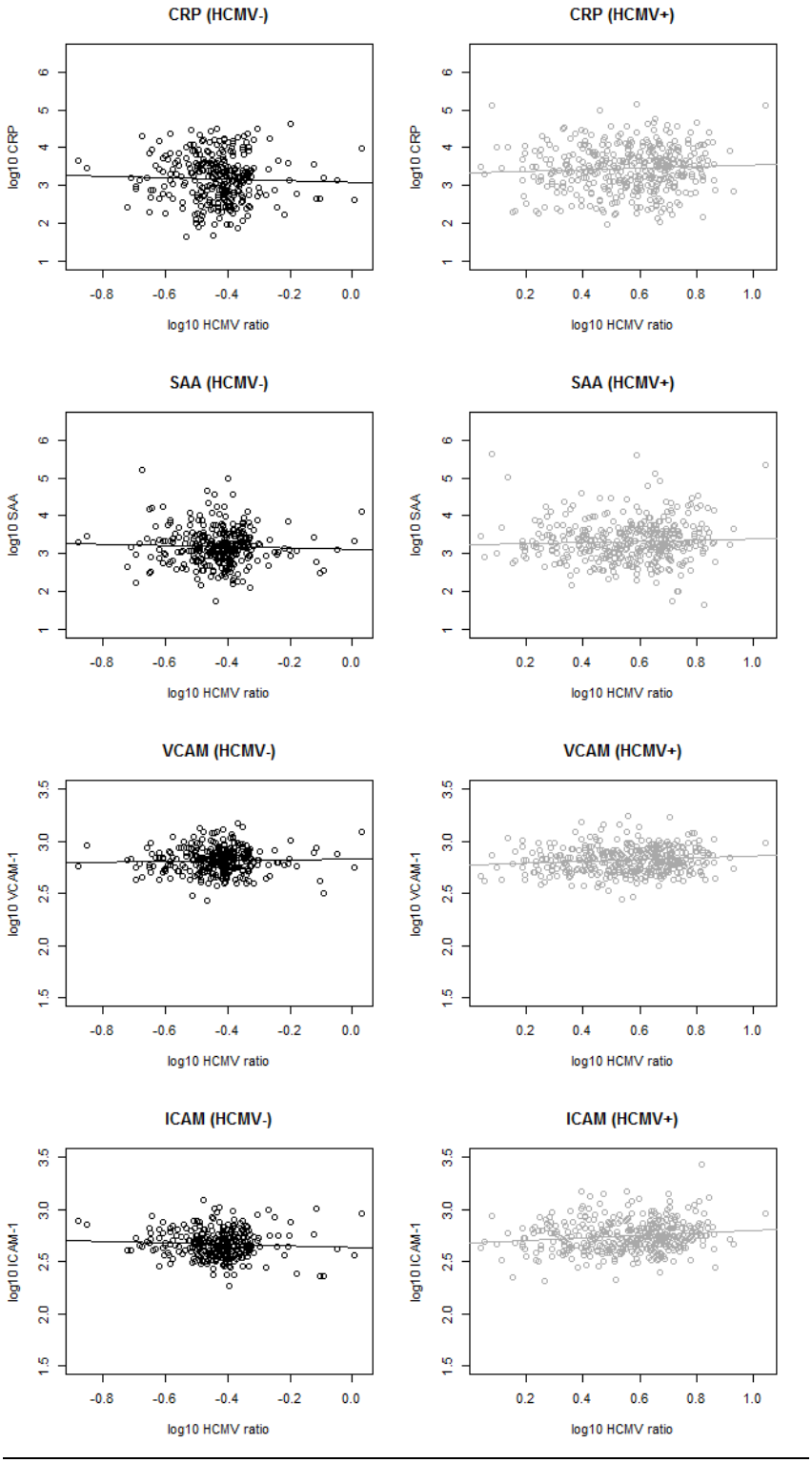


Figure 3. Scatter plots and fitted values from the simple linear regression model of $\log_{10}$HCMV ratio and CRP, SAA, VCAM-1, or ICAM-1. Linear regression models not controlled for variables $\log_{10}$BMI, race/ethnicity, smoker, age, gender.

Table 5. Linear regression models for the four biomarkers of vascular injury (CRP, SAA, VCAM-1, and ICAM-1) predicted by dichotomized seropositive and seronegative HCMV status.

CRP vs. HCMV Seropositive Status				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV	0.08	0.04	1.72	**0.086**
log_{10}BMI	1.31	0.10	13.73	2.6E-37
Race	-0.09	0.05	-2.01	0.045
Age	0.00	0.00	3.00	0.003
Smoke Current	0.09	0.04	2.06	0.040
Gender	-0.23	0.04	-5.34	1.3E-07

SAA vs. HCMV Seropositive Status				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV	0.03	0.04	0.67	**0.51**
log_{10}BMI	0.80	0.09	8.89	7.7E-18
Race	-0.03	0.04	-0.77	0.444
Age	0.00	0.00	2.47	0.014
Smoke Current	-0.02	0.04	-0.40	0.693
Gender	-0.15	0.04	-3.56	4.0E-04

VCAM-1 vs. HCMV Seropositive Status				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV	0.02	0.01	1.72	**0.086**
log_{10}BMI	0.05	0.02	2.40	0.017
Race	0.06	0.01	5.28	1.9E-07
Age	0.00	0.00	3.96	8.3E-05
Smoke Current	-0.01	0.01	-0.60	0.547
Gender	0.02	0.01	2.20	0.028

ICAM-1 vs. HCMV Seropositive Status				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV	0.04	0.01	3.09	**0.002**
log_{10}BMI	0.10	0.03	4.02	6.6E-05
Race	-0.04	0.01	-3.39	0.001
Age	0.00	0.00	4.09	4.9E-05
Smoke Current	0.03	0.01	3.00	0.003
Gender	0.02	0.01	1.46	0.146

P-value (Pr(<|t|)) significant at $\alpha<0.05$. Linear models were controlled for variables log_{10}BMI, race, age, current smoking, and gender. All controlled for variables were dichotomous except log_{10}BMI, which was continuous.

To investigate if higher IgG response to HCMV resulted in increased levels of vascular injury biomarkers, the continuous Log_{10}HCMV ratio used to calculate seropositivity or seronegativity was tested for associations with vascular injury biomarker levels. Seronegative results were set to 0 and used as a reference category for this analysis (Table 6). Among those seropositive for HCMV, Log_{10}HCMV ratio was a significantly associated with increased levels of Log_{10}VCAM-1 and Log_{10}ICAM-1 biomarkers (Table 6). Results measured in serum after controlling for variables log_{10}BMI, race, age, current smoking, and gender (Table 6). Log_{10}BMI was the most significant independent variable for CRP (p = 1.4E-37), SAA (p = 6.7E-18), and VCAM-1 (p = 6E-5) while race was the most significant independent variable for ICAM-1 (p = 4.8E-8) (Table 6).

Table 6. Linear regression models of biomarkers of vascular injury (CRP, SAA, VCAM-1, and ICAM-1) predicted by the log_{10}HCMV ratio to plate-specific cutoff (log_{10}HCMV ratio).

CRP vs. log_{10}HCMV ratio positive					SAA vs. log_{10}HCMV ratio positive				
	Estimate	Std. Error	t value	Pr(>|t|)		Estimate	Std. Error	t value	Pr(>|t|)
log_{10}HCMV ratio	0.07	0.08	0.93	**0.351**	**log_{10}HCMV ratio**	0.05	0.07	0.65	**0.513**
log_{10}BMI	1.32	0.10	13.79	1.4E-37	log_{10}BMI	0.80	0.09	8.91	6.7E-18
Race	-0.10	0.05	-2.23	0.026	Race	-0.03	0.04	-0.76	0.446
Age	0.00	0.00	2.98	0.003	Age	0.00	0.00	2.39	0.017
Smoke Current	0.09	0.04	2.08	0.038	Smoke Current	-0.02	0.04	-0.41	0.683
Gender	-0.23	0.04	-5.34	1.4E-07	Gender	-0.14	0.04	-3.51	4.9E-04
VCAM-1 vs. log_{10}HCMV ratio positive					**ICAM-1 vs. log_{10}HCMV ratio positive**				
	Estimate	Std. Error	t value	Pr(>|t|)		Estimate	Std. Error	t value	Pr(>|t|)
log_{10}HCMV ratio	0.07	0.02	3.55	**4.2E-04**	**log_{10}HCMV ratio**	0.05	0.02	2.52	**0.012**
log_{10}BMI	0.10	0.03	4.04	6.0E-05	log_{10}BMI	0.05	0.02	2.38	0.018
Race	-0.04	0.01	-3.23	0.001	Race	0.06	0.01	5.53	4.8E-08
Age	0.00	0.00	3.74	2.0E-04	Age	0.00	0.00	3.66	2.7E-04
Smoke Current	0.03	0.01	2.90	0.004	Smoke Current	-0.01	0.01	-0.71	0.481
Gender	0.02	0.01	1.68	0.094	Gender	0.02	0.01	2.41	0.016

P-value (Pr(<|t|)) significant at $\alpha<0.05$. Linear model controlled for variables log_{10}BMI, race, age, current smoking, and gender. All controlled for variables were dichotomous except log_{10}BMI, which was continuous.

In addition to testing dichotomized and continuous HCMV IgG responses, the continuous log_{10}HCMV ratio was categorized into tertiles and tested for association with vascular injury biomarker levels. Seronegative results were set to 0 and used as a reference category for this analysis (Table 7). Individuals in the highest HCMV ratio tertile had significantly higher levels of VCAM-1 (p = 0.031) and ICAM-1 (p = 0.005). In addition, individuals in the second HCMV tertile had significantly higher levels of ICAM-1 (p = 0.032) (Table 7). Log_{10}BMI was the most significant independent variable for CRP (p = 2.4E-37), SAA (p = 8.1E-18), and ICAM-1 (p = 6.4E-5) while race was the most significant independent variable for VCAM-1 (p = 1.1E-7) (Table 7).

The multiplicative adjusted median of VCAM-1 and ICAM-1 in seropositive individuals was significantly (VCAM-1: 19%, p = 4.2E-4; ICAM-1: 11%, p = 0.012) higher than the adjusted median in seronegative individuals (Table 8). The median percent change increased as HCMV tertile increased in both VCAM-1 (1^{st}: 2%, p = 0.484; 2^{nd}: 4%, p = 0.272; 3^{rd}: 8%, p = 0.031) and ICAM-1 (1^{st}: 7%, p = 0.051; 2^{nd}: 8%, p = 0.032; 3^{rd}: 12%, p = 0.005), demonstrating

that the association between HCMV and VCAM-1 and ICAM-1 increased with increasing HCMV IgG response (Table 8).

Table 7. Linear model of biomarkers of vascular injury (CRP, SAA, VCAM-1, and ICAM-1) predicted by the HCMV tertiles to plate-specific cutoff (HCMV ratio).

CRP vs. HCMV Tertiles				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV 1st Tertile	0.15	0.06	2.58	**0.010**
HCMV 2nd Tertile	0.03	0.06	0.44	**0.657**
HCMV 3rd Tertile	0.04	0.06	0.60	**0.550**
log_{10}BMI	1.31	0.10	13.74	2.4E-37
Race	-0.10	0.05	-2.15	0.032
Age	0.00	0.00	3.23	0.001
Smoke Current	0.10	0.04	2.21	0.028
Gender	-0.24	0.04	-5.55	4.3E-08

SAA vs. HCMV Tertiles				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV 1st Tertile	0.04	0.06	0.75	**0.456**
HCMV 2nd Tertile	0.00	0.06	-0.02	**0.981**
HCMV 3rd Tertile	0.05	0.06	0.76	**0.448**
log_{10}BMI	0.80	0.09	8.89	8.1E-18
Race	-0.03	0.04	-0.75	0.451
Age	0.00	0.00	2.42	0.016
Smoke Current	-0.02	0.04	-0.38	0.701
Gender	-0.15	0.04	-3.57	3.9E-04

VCAM-1 vs. HCMV Tertiles				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV 1st Tertile	0.01	0.01	0.70	**0.484**
HCMV 2nd Tertile	0.02	0.01	1.10	**0.272**
HCMV 3rd Tertile	0.03	0.01	2.16	**0.031**
log_{10}BMI	0.05	0.02	2.42	0.016
Race	0.06	0.01	5.38	1.1E-07
Age	0.00	0.00	3.69	2.5E-04
Smoke Current	-0.01	0.01	-0.70	0.484
Gender	0.02	0.01	2.31	0.021

ICAM-1 vs. HCMV Tertiles				
	Estimate	Std. Error	t value	Pr(>|t|)
HCMV 1st Tertile	0.03	0.02	1.96	**0.051**
HCMV 2nd Tertile	0.03	0.02	2.15	**0.032**
HCMV 3rd Tertile	0.05	0.02	2.81	**0.005**
log_{10}BMI	0.10	0.03	4.03	6.4E-05
Race	-0.04	0.01	-3.29	0.001
Age	0.00	0.00	3.88	1.1E-04
Smoke Current	0.03	0.01	2.93	0.004
Gender	0.02	0.01	1.52	0.128

P-value (Pr(<|t|)) significant at $\alpha<0.05$. Linear model controlled for variables log_{10}BMI, race, age, current smoking, and gender. All controlled for variables were dichotomous except log_{10}BMI, which was continuous.

Table 8. Adjusted multiplicative median percent change of vascular injury biomarkers

HCMV Variable	Log_{10}CRP Median % Change	Log_{10}SAA Median % Change	Log_{10}VCAM-1 Median % Change	Log_{10}ICAM-1 Median % Change
log_{10}HCMV ratio	19%	12%	19%	11%
HCMV 1st Tertile	42%	10%	2%	7%
HCMV 2nd Tertile	6%	0%	4%	8%
HCMV 3rd Tertile	9%	11%	8%	12%

Linear model controlled for variables, log_{10}BMI, race, age, current smoking, and gender. All controlled for variables were dichotomous except log_{10}BMI, which was continuous.

CHAPTER 4: DISCUSSION

Demographic Characteristics

We observed that HCMV IgG seropositivity was significantly associated with demographic predictors including age, BMI, gender, race, current smoking, past smoking, *H. pylori* IgG seropositivity, and *T. gondii* IgG seropositivity. The prevalence of HCMV IgG seropositivity increased with increasing age and BMI. HCMV IgG seropositivity was more common among reported females, non-white or Hispanic individuals, current smokers, and individuals who had ever smoked. Individuals who were infected with HCMV IgG were also more likely to test IgG seropositive with either *H. pylori* or *T. gondii*. HCMV IgG associations with increasing age, non-white or Hispanic race/ethnicity, and female gender are consistent with results published in the scientific peer-reviewed literature (Bate et al., 2010; Cannon et al., 2010) (Table 9). A history of smoking has not previously been associated with HCMV infection. However, smoking is known to be more prevalent with lower social economic status (Hiscock et al., 2012). In this study, the association with diabetes was not significant but literature suggests that HCMV is associated with type 1 diabetes (Pak et al., 1988). It is possible that no significant association was observed between HCMV IgG seropositivity and diabetes in this study because data on type of diabetes were not collected.

Table 9. The demographic characteristics in the NHANES (1999-2004) are similar with the results obtained in the SAFE study.

	Demographic Characteristic	SAFE 2013	NHANES 1999-2004
Gender	Female	61%	56%
	Male	48%	45%
Race/Ethnicity	non-white	73%	70.6%-76.9%
	White (non-Hispanic)	39%	40%
Age	SAFE (18-29) NHANES (20-29)	44%	50%
	30-39	59%	57%
	40-49	66%	58%

NHANES data analyzed and reported by Bate et al., 2010.

WOCBA subpopulation

We observed that the women of childbearing age had similar demographic predictors when compared to the total study population. Similar to the total population, in the WOCBA subpopulation, an increase in prevalence of HCMV infection increased with increasing age and increasing BMI. The same self-reported demographic predictors observed in the total population were also observed in WOCBA, i.e., HCMV IgG seropositivity was associated with non-white or Hispanic ethnicity, and current or past history of smoking. Significant associations were also observed between HCMV IgG seropositivity and *H. pylori* IgG seropositivity, but not among *T. gondii* IgG seropositivity, as was seen in the total population. *T. gondii* and HCMV both cause congenital birth defects, suggesting that having both diseases at the same time could lead to a higher risk of birth defects following initial infection or reactivation of either long-term latent infection in the mother during pregnancy. Associations between HCMV IgG seropositivity and increasing age and non-white or Hispanic ethnicity observed in this study are consistent with HCMV IgM seroprevalence findings reported in previous peer-reviewed literature where HCMV IgM seroprevalence in an HCMV IgG seropositive population was examined (Wang et al., 2016). High levels of HCMV IgM, HCMV IgM seropositivity, occurs during primary HCMV infection, reinfection or reactivation (Wang,et al., 2016). We did not test HCMV IgM levels of

the participant's serum, but associations observed in peer-reviewed literature between both HCMV IgG and IgM, and individuals of older age or non-white or Hispanic ethnicity suggest that these populations should be priority targets for intervention (Bate et al., 2010; Cannon et al., 2010; Terrazzini, Bajwa, Thomas, Smith, & Kern, 2014; Wang et al., 2016). Because we demonstrated a large change in prevalence of HCMV IgG seropositivity from younger to older age in WOCBA in this study, a significant increased risk exists for a woman in reproductive years to become infected with HCMV. It is imperative that women with greater risk, including non-white or Hispanic women who smoke, are older, and/or have increased BMI, be targeted for education about the risk factors of HCMV infection during pregnancy. The targeted intervention should also include sexual partners, particularly in the most concerning cases where the mother is seronegative and their partner is seropositive.

Vascular Injury

We observed positive associations between HCMV IgG seropositive status and increased levels of vascular injury biomarkers ICAM-1 and VCAM-1. However, only ICAM-1 was significant (seropositive, $p = 0.002$; Log_{10}HCMV IgG ratio, $p = 0.012$; 2nd tertile HCMV IgG, $p = 0.032$; 3rd tertile HCMV IgG, $p = 0.005$) and consistent in all three regression model analyses. Significant positive associations were also observed between Log_{10}HCMV IgG ratio ($p = 4.2E\text{-}4$) and the 3rd tertile of HCMV IgG ($p = 0.031$) and increased levels of VCAM-1. To my knowledge, this is the first study to focus on associations between HCMV infection and adhesion molecules VCAM-1 and ICAM-1 in a human cohort. The majority of previous studies only focused on CRP or studied adhesion molecules *in vitro*. Betjes et al. only focused on CRP as a biomarker of vascular injury and observed no association with HCMV IgG seropositive status

(Betjes et al., 2007) while in other studies associations with CRP or CRP-enhanced associations between vascular injury and HCMV IgG seropositive status were observed (Popović et al., 2012; Simanek et al., 2011; Terrazzini et al., 2014). These conflicting results may be due to the fact that the pathway for the development of CVD is not directly influenced by HCMV, or that CRP is an acute non-specific biomarker. The vascular injury biomarker CRP measures acute damage, levels rising quickly after injury, identifying those at immediate risk of a first cardiac event, but increased levels of CRP can also indicate the presence of other diseases such as cancer or lupus (Kaptoge et al., 2012; Zakynthinos et al., 2009). If HCMV infection is leading to the development of CVD, as we have hypothesized, a non-specific biomarker like CRP would be confounded by non-CVD related disease. Additionally, CRP's ability to detect acute vascular injury would not capture individuals who are in the process of developing CVD (i.e. plaque buildup), but only individuals who have already developed CVD, possibly explaining the observed lack of a statistically significant association between HCMV infection and increased CRP. Adhesion molecules are released much earlier during the development of CVD when plaque buildup begins to occur compared to when the non-specific acute inflammation biomarker CRP is expressed. VCAM-1 and ICAM-1 have the most biological relevance for their role in the vascular endothelium, where HCMV has been observed to target endothelial cells (Terrazzini et al., 2014; Zakynthinos et al., 2009). VCAM-1 and ICAM-1 are released from endothelial cells as markers of an increased inflammatory response (Popović et al., 2012). Based on the associations between HCMV IgG seropositivity and specific markers of coagulation and plaque buildup, VCAM-1 and ICAM-1, it is likely that the mechanism specific to coagulation pathways proposed by Popović et al. explains the associations observed in our study (Popović et al., 2012). The vascular inflammation is a result of coagulation occurring after the disruption of endothelial

processes by HCMV infections, allowing more thrombin to be generated (Popović et al., 2012). As a result of CVD, and more specifically, VCAM-1 and ICAM-1's association with environmental air pollution, the associations observed in this study indicates that HCMV infection may lead to susceptibility to environmental challenges in infected individuals (Bind et al., 2013).

Limitations

A limitation of this study was that household income could not be ascertained and may confound the results of the analyses for HCMV IgG seropositivity. Education data (that can be used as a proxy for income, by assuming more education results in higher socioeconomic status) was only recorded for the participants whose samples were collected at the U.S. EPA HSF facility (n = 349), and thus we could not use this variable when investigating the total study population (n = 694). It is likely that no significant association was found between HCMV IgG seropositivity and diabetes because information on type 1 or type 2 diabetes was not collected in this study. HCMV IgG seropositivity was observed in peer-reviewed literature to be associated with autoimmune type 1 diabetes, but not type 2 diabetes (Pak et al., 1988). In the NHANES III study from 1988-1994, a significant predictor of HCMV IgM seropositivity among the HCMV IgG seropositive women was family size (Wang et al., 2016). This information was not collected from our study participants and thus, we could not assess this variable. Not testing for HCMV IgM seropositivity limited the scope of our study as we could not examine the effects of primary infection, reactivation, or reinfection. However, a four-fold increase in HCMV IgG response is indicative that a reactivation or reinfection has occurred (Prince et al., 2014) and, thus for future studies, we could measure HCMV IgG response over time to look for a four-fold increase

indicative of reactivation or reinfection of HCMV infection. Despite evidence among transplant patients, some animal experiments, and the biological plausibility of HCMV leading to vascular injury and CVD, it is possible that the inflammation caused by vascular damage from something other than HCMV may lead to a reactivation of the long-term latent infection during a period of stress (Betjes et al., 2007; Fateh-Moghadam et al., 2003; Mehta et al., 2014; Popović et al., 2012). Because samples were not tested using a medical diagnostic test, it is possible that misclassification of HCMV IgG results could account for the increased association between HCMV and VCAM-1 and ICAM-1. It is less likely that samples would be misclassified at high HMCV IgG seropositive response than at low HCMV IgG seropositive response, particularly near the indeterminate range. If misclassification were to occur at a higher rate in the samples with low HCMV IgG seropositive response, it could explain why the observed association between HCMV IgG, and VCAM-1 and ICAM-1 was significant in the 3rd tertile ($p = 0.031$ and 0.005 respectively) but weaker and not significant in the 1st tertile ($p = 0.484$ and 0.051 respectively).

Next Steps/Future Studies

A study of NHANES III data collected from 1988-1994 indicates that increased CRP level was predictive of CVD-related mortality but not all-cause mortality, and that HCMV alone was not predictive of CVD-related mortality (Simanek et al., 2011). This data suggests that performing a follow-up study to investigate participants' cause of death, specifically related to CVD-related mortality, could provide more insight into the possible association between HCMV IgG serostatus as a predictor of CVD-related mortality. Because HCMV IgM is only increased during primary infection, reactivation, or reinfection, it would be warranted in a future study to include testing for HCMV IgM seropositivity to see if the vascular injury is being caused by

acute (HCMV IgM seropositive) or latent infection (HCMV IgG seropositive, IgM seronegative) (Prince et al., 2014). By studying acute infection of HCMV, it might be possible to determine if active (primary, reactivation, reinfection) infection is causing vascular injury or if latent infection is the likely culprit of HCMV associated vascular injury. A study in which HCMV infected patients are followed for a period of a year and the levels of HCMV IgG, HCMV IgM, CRP, SAA, ICAM-1, and VCAM-1 are measured every month could increase our understanding of the relationship between HCMV and CVD. For future study of environmental air pollutants being studied for their potential to exacerbate CVD, I would recommend considering HCMV infection as a possible confounder in the development of CVD in response to environmental air pollution.

CHAPTER 5: CONCLUSIONS

Analysis of this study data revealed evidence that HCMV IgG seropositive status was associated with increased levels of biomarkers of vascular injury, specifically VCAM-1 and ICAM-1. Significant demographic predictors observed in this study were the same as reported in the peer-reviewed scientific literature for HCMV IgG seropositivity, namely increased age and BMI, as well as female gender, non-white ethnicity, and a history of smoking. These demographic predictors were consistent between the total population and the subpopulation of women of childbearing age.

APPENDIX A: SAFE QUESTIONNAIRE

SAFE

Salivary Assay Feasibility Evaluation

Education	
1. What is the highest level of education you have attained? **[**Check corresponding box below**]**	
Did not graduate high school	☐ 0
High school graduate	☐ 1
Some college, no degree	☐ 2
Associate degree	☐ 3
Bachelor's degree (EXAMPLE: BA, AB, BS, BBA)	☐ 4
Post baccalaureate degree	☐ 5
Decline to answer	☐ 888
Don't know	☐ 999

2. What do you consider your race to be? **[**Check all that apply**]**	
White	☐ 0
Black or African American	☐ 1
American Indian or Alaska Native	☐ 2
Asian or Pacific Islander	☐ 3
Other	☐ 4
Decline to answer	☐ 888
Don't know	☐ 999

Ethnicity		
3. Do you consider yourself to be Latino or Hispanic?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer

Gender	
4. What is your gender?	
Female	☐ 0
Male	☐ 1

The following questions are about cats [Check corresponding box below]		
5.a. Have you <u>ever</u> had cats in your house or at your residence?	☐ 0	No **(Skip to 6.a.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
5.b. How many cats <u>currently</u> live in your house or at your residence?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
5.c. What is the greatest number of cats that have lived in your house or at your residence at one time?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
5.d. <u>In total</u>, how many years have cats lived in your house or at your residence?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
5.e. Have <u>all</u> the cats that you lived with been only indoor cats? (this means they never go outside)	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
5.f. Have <u>all</u> your cats been treated for worms (current and past cats)?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
5.g. Have <u>all</u> of your cats been to a veterinarian?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

Other CAT care

[Check corresponding box below]

6.a. Which of the following best describes how often you currently touch cats?	☐ 1	Daily
	☐ 2	Weekly
	☐ 3	Monthly
	☐ 4	Annually
	☐ 5	Less than once a year
	☐ 0	Never
6.b. Have you ever been responsible for cleaning the litter box?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
6.c. Do you pet sit cats in your residence or at someone's home?	☐ 0	No
	☐ 1	Once per year
	☐ 2	Twice per year
	☐ 3	Three times per year
	☐ 4	More than three times per year
6.d. Have you ever touched a stray cat, foster cat, a cat staying at an animal shelter?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
6.e. When was the last time you touched any cat? (estimate year)		Year _____
	☐ 888	Never
	☐ 999	Don't know

The following questions are about dogs [Check corresponding box below]		
7.a. Have you <u>ever</u> had dogs in your house or at your residence?	☐ 0	No **(Skip to 8.a.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
7.b. How many dogs do you <u>currently</u> live in your house or at your residence?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
7.c. What is the greatest number of dogs you had in your house or at your residence at one time?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
7.d. <u>In total</u>, how many years have dogs lived at your residence?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more
7.e. Are any dogs in your house or at your residence ever allowed inside?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
7.f. Have <u>all</u> of your dogs been treated for worms (current and past dogs)?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
7.g. Have <u>all</u> of your dogs been to a veterinarian?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

Other DOG care

[Check corresponding box below]

8.a. Which of the following best describes how often you currently touch dogs?	☐ 1	Daily
	☐ 2	Weekly
	☐ 3	Monthly
	☐ 4	Annually
	☐ 5	Less than once a year
	☐ 0	Never
8.b. Do you pet sit dogs in your residence or at someone's home?	☐ 0	No
	☐ 1	Once per year
	☐ 2	Twice per year
	☐ 3	Three times per year
	☐ 4	More than three times per year
8.c. Have you ever touched a stray dog, foster dog or a dog staying at an animal shelter?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
8.d. When was the last time you touched any dog? (estimate year)		Year _____
	☐ 888	Never
	☐ 999	Don't know

FOOD, WATER & ENVIRONMENT [Check corresponding box below]		
9.a. Have you eaten any beef in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.b. Have you eaten any pork in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.c. Have you eaten any chicken in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.d. Have you eaten any lamb in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.e. Have you eaten any goat in the in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.f. Have you eaten any venison (deer) in the in the last 3 months that was raw or not cooked all the way through?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.g. Have you consumed raw goat's milk in the last 3 months?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
9.h. What is the source of your drinking water?	☐ 0 ☐ 1 ☐ 2 ☐ 999	Private well water Municipal city or county water Commercially bottled water Other

9.j. Have you ever lived on a farm?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
9.k. Which of the following best describes how often you currently handle soil with your bare hands?	☐ 1	Daily
	☐ 2	Weekly
	☐ 3	Monthly
	☐ 4	Annually
	☐ 5	Less than once a year
	☐ 0	Never

General health information [Check corresponding box below]	
10.a. How would you rank your general health?	☐ 0 Excellent ☐ 1 Very good ☐ 2 Good ☐ 3 Fair ☐ 4 Poor ☐ 888 Decline to Answer ☐ 999 Don't know
10.b. Do you wear dentures?	☐ 0 No ☐ 1 Yes ☐ 888 Decline to answer ☐ 999 Don't know
10.c. Do you now smoke cigarettes or other types of tobacco (cigar, pipe, etc)?	☐ 0 Not at all ☐ 1 Yes, less than every day ☐ 2 Yes, daily ☐ 888 Decline to answer ☐ 999 Don't know
10.d. Did you ever smoke cigarettes or other types of tobacco at least once a week in the past?	☐ 0 No ☐ 1 Yes ☐ 888 Decline to answer ☐ 999 Don't know
10.e. How many alcohol drinks do you consume in a week?	☐ 0 None ☐ 1 1 ☐ 2 2-7 ☐ 3 8-14 ☐ 4 15+ ☐ 888 Decline to answer ☐ 999 Don't know

10.f. Which of the following are you allergic to? [Check "No" or "Yes" for **all]**

No	Yes	
☐ 0	☐ 1	Drug allergies
☐ 0	☐ 1	Animal dander
☐ 0	☐ 1	Dust
☐ 0	☐ 1	Food
☐ 0	☐ 1	Mold
☐ 0	☐ 1	Plants
☐ 0	☐ 1	Pollen
☐ 0	☐ 1	Smoke
	☐ 888	Other

11. Have you EVER been told by a physician or health professional that you have any of the following illnesses or conditions? [Check corresponding box below**]**

11.a. Diabetes	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.b. Kidney disease	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.c. Organ transplant	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.d. Liver disease	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

11.e. Cancer, other than skin cancer	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.f. HIV	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.g. Ulcers	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.h. Inflammatory bowel disease (IBS)	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.i. Dyspepsia	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.j. Chronic obstructive pulmonary disease (COPD)	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.k. Kidney Disease	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.l. Heart Disease	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.m. Arthritis	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

14.n. Schizophrenia	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.o. Asthma	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.p. Epilepsy	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
11.q. Depression	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

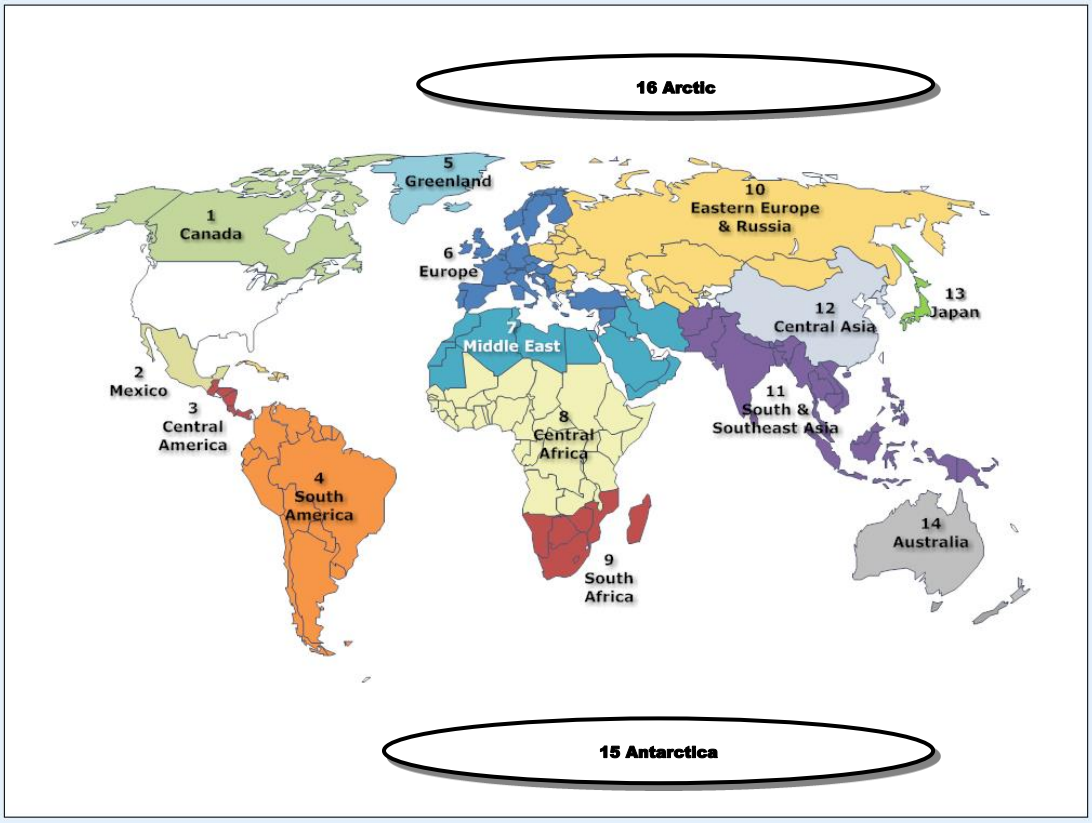
Symptoms [Check corresponding box below]		
12.a. Have you had a fever above 100.3 degrees (Fahrenheit) in the past 3 months?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No Yes Decline to answer Don't know
12.b. In the past 3 months, have you had diarrhea?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No **(Skip to 12.d.)** Yes Decline to answer Don't know
12.c. If yes, how many days altogether did you have diarrhea?	☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	0 1 2 3 4 5 or more
12.d. In the past 3months, have you experienced any vomiting?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No **(Skip to 12.f.)** Yes Decline to answer Don't know
12.e. If yes, how many days altogether did you experience any vomiting?	☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	0 1 2 3 4 5 or more
12.f. In the past month, have you experienced any wheezing?	☐ 0 ☐ 1 ☐ 888 ☐ 999	No **(Skip to 12.h.)** Yes Decline to answer Don't know
12.g. If yes, how many days altogether did you experience any wheezing?	☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5	0 1 2 3 4 5 or more

12.h. In the past month, have you had a cough?	☐ 0	No **(Skip to 13.a.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
12.i. If yes, how many days altogether did you have a cough?	☐ 0	0
	☐ 1	1
	☐ 2	2
	☐ 3	3
	☐ 4	4
	☐ 5	5 or more

Diagnosis/Treatment [Check corresponding box below]		
13.a. Have you ever been diagnosed with toxoplasmosis by a physician?	☐ 0	No **(Skip to 13.c.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
13.b. If yes, did you receive treatment for this infection?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
13.c. Have you ever been diagnosed with *Helicobacter pylori* infection by a physician?	☐ 0	No **(Skip to 13.e.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
13.d. If yes, did you receive treatment for this infection?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
13.e. Have you ever been diagnosed with toxocariasis by a physician?	☐ 0	No **(Skip to 14.)**
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know
13.f. If yes, did you receive treatment for this infection?	☐ 0	No
	☐ 1	Yes
	☐ 888	Decline to answer
	☐ 999	Don't know

14. IN THE PAST 3 months, did you take or receive any of the following drugs or medications for any reason or condition?[Check corresponding box below**]**	
14.a. Antibiotics	☐ 0 No ☐ 1 Yes ☐ 888 Decline to answer ☐ 999 Don't know
14.b. Chemotherapy	☐ 0 No ☐ 1 Yes ☐ 888 Decline to answer ☐ 999 Don't know
14.c. Steroids	☐ 0 No ☐ 1 Yes ☐ 888 Decline to answer ☐ 999 Don't know

15. If you have lived or traveled outside the United States, please choose which regions you have been to? [Check corresponding box below]	
Have not been outside United States	☐ 0
Canada	☐ 1
Mexico	☐ 2
Central America and Caribbean Islands	☐ 3
South America	☐ 4
Greenland	☐ 5
Europe	☐ 6
Middle East	☐ 7
Central Africa	☐ 8
South Africa	☐ 9
Eastern Europe & Russia	☐ 10
South & Southeast Asia	☐ 11
Central Asia	☐ 12
Japan	☐ 13
New Zealand	☐ 14
Antarctica	☐ 15
Arctic	☐ 16
Decline to answer	☐ 888
Don't know	☐ 999


16 Arctic
5
Greenland
1
Canada
10
Eastern Europe
& Russia
6
Europe
13
Japan
12
Central Asia
7
Middle East
2
Mexico
3
Central
America
11
South &
Southeast Asia
8
Central
Africa
4
South
America
14
Australia
9
South
Africa
15 Antarctica

APPENDIX B: NIEHS MEDICAL QUESTIONNAIRE

Sample Collection Registry Protocol Questionnaire

To be administered by study staff prior to sample collection. Please make notation on paper questionnaire and in the electronic questionnaire for any question not answered by the study participant.

Date_____ Time_____

Patient Initials _____ Patient # _____

Birth Date _____

Race
- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White

Ethnicity
- ☐ Hispanic/Latino
- ☐ Not Hispanic/Latino

Gender
- ☐ Male
- ☐ Female

Height _____inches Weight _____lbs.

*IF MALE, THEN START AT Q3:

1. Have you started or gone through menopause? ☐ YES

☐ NO

☐ NOT SURE

2. Do you currently take any type of hormone replacement therapy such as estrogen, progesterone, or prempro?
   ☐ YES
   ☐ NO

3. Have you had any alcoholic drinks during the past 24 hours?
   ☐ YES
   ☐ NO

(IF NO, GO TO Q4)

How many alcoholic drinks have you had in the past 24 hours?_____

4. Do you currently smoke (past 24 hours)?
   ☐ YES
   ☐ NO

If YES:

How many cigarettes do you usually smoke per day?_____

How many cigarettes have you smoked in the past 24 hours? _____

If NO:

When did you last smoke cigarettes? ___ /___ /____ or N/A ☐

MM DD YYYY

5. Have you smoked more than 100 cigarettes in your entire life? (do NOT include cigars, pipe, marijuana, chewing tobacco) (If no, skip other questions)
   ☐ YES
   ☐ NO

(IF NO, GO TO Q6)

If YES:

When did you start smoking regularly? ___ /___ /____

MM DD YYYY

6. Did you eat or drink anything other than water in the 8 hours prior to your blood draw? ☐ YES ☐ NO

(IF NO, GO TO Q7)

If YES:

What food or drink did you have?

☐ TEA, COFFEE, DIET SODA

☐ JUICE OR MILK

☐ REGULAR SODA

☐ SNACK

☐ FULL MEAL

What time did you last have something to eat or drink? AM/PM

_____:_____

HH MM

7. Did you take any medications in the last 24 hours? (includes all vitamins and supplements or over-the-counter medications AND all prescribed

medications including pills, patches, liquids, injections, inhalers, creams, etc).

☐ YES

☐ NO

If YES:

Please record the names of your medication bottles or packages in the space below. List those medications taken in the past 24 hours as well as those you take regularly along with the primary reason you take the medication.

Please do not include dosage or frequency of use.

1._____

2._____

3. _____

4._____

5. _____

6. _____

7. _____

8. _____

9. _____

10. _____

Medical and Exposure History

8. Has a doctor ever told you that you have or have had any of the following:

Asthma Yes No

If YES:

- In the past 12 months, have you had wheezing or whistling in your chest?
  Yes No
- In the past 12 months, have you taken any medication, prescribed by a doctor, for wheezing or whistling?
  Yes No
- Complete the NHANES Questionnaire for asthmatics

High Blood Pressure Yes No

If YES:

Because of your high blood pressure/hypertension, have you ever been told to take prescribed medicine?

Heart attack	Yes	No
Cancer (any type)	Yes	No
Diabetes	Yes	No
High Cholesterol	Yes	No
Anxiety/Depression	Yes	No

Allergies

Seasonal	Yes	No
Year Round	Yes	No
Food/ Medication	Yes	No

If YES:

During the past 12 months, have you had an episode of hay fever?

Other physician-diagnosed diseases/conditions (specify):

_____

_____

For sperm collection (males only)

9. When was the last intercourse/ejaculation prior to collection?____/____/_____

MM DD YYYY

For collection at home, please provide exact time of collection: AM/PM

_____:_____

HH MM

If Pulmonary Function Test (PFT) was conducted, please complete the following:

10. FVC _____liters

11. FVC _____ (% predicted)

12. FEV_1 _____liters

13. FEV_1 _____(% predicted)

14. FEV_1/FVC _____(%)

15. FEF25-75 _____liters/sec

16. FEF25-75 _____(% predicted)

17. Physician spirometry assessment:

- Unreliable measurements ☐
- Normal spirometry ☐
- Obstruction
  - Mild (FEV1/FVC<80%, FEV1=70-79%) ☐
  - Moderate (FEV1/FVC<80%, FEV1=50-69%) ☐
  - Severe (FEV1/FVC<80%, FEV1=30-49%) ☐

Very severe (FEV1/FVC<80%, FEV1<30%) ☐

- Possible restriction (FEV1/FVC≥80%, FVC<80%) ☐
- Possible restriction or air trapping (FEV1/FVC<80%, FVC<80%) ☐
- Small airways disease (FEV1/FVC≥80%, FEF25-75<70%) ☐
- Normal FV loop ☐
- FV loop indicates upper airway obstruction ☐
- FV loop indicates lower airway obstruction ☐

Other observations.....

APPENDIX C: NIEHS ENVIRONMENTAL EXPOSURE QUESTIONNAIRE

1. Have you ever had cats in your house or at your residence?
   No **(Skip to Q4.)**
   Yes
   Don't know

2. What is the greatest number of cats that have lived in your house or at your residence at one time?
   0
   1
   2
   3
   4
   5 or more

3. Have all the cats that you lived with been only indoor cats? (this means they never go outside)
   No
   Yes
   Don't know

4. Have you ever had dogs in your house or at your residence?
   No **(Skip to Q6.)**
   Yes
   Don't know

5. What is the greatest number of dogs you had in your house or at your residence at one time?
   0
   1
   2
   3
   4
   5 or more

6. Have you ever been diagnosed with or received treatment for toxoplasmosis by a physician?
   No
   Yes
   Don't know

7. Have you ever been diagnosed with or received treatment for *Helicobacter pylori* infection by a physician?
   No
   Yes
   Don't know

8. Have you eaten any beef or pork in the last 3 months that was raw or not cooked all the way through?
   No
   Yes
   Don't know

9. What is the source of your drinking water?
   Private well water
   Municipal city or county water
   Commercially bottled water
   Other
   Don't know

10. Which of the following best describes how often you currently handle soil with your bare hands?
    None
    daily
    weekly
    monthly
    annually
    less than once a year

11. Have you ever lived or travelled outside of the United States?
    No
    Yes
    Don't know

12. Have you ever lived on a farm?
    No
    Yes
    Don't know

APPENDIX D: R CODE

```
#.libPaths("C:/Program Files/R/R-3.1.2/library")
library(readxl)
library(plyr)
library(openxlsx)

#Load Data
setwd("C:/Users/jnsty_000/Documents/Grad School Work/Masters Thesis/Data")
all2<-read_excel("cmvALL.xlsx",sheet=1,col_names=TRUE,col_types=NULL,skip=0)
setwd("C:/Users/jnsty_000/Documents/Grad School Work/Masters Thesis/Conferences/Poster/Data")
pip<-read_excel("cmv_all_dat_JS 2017-06-09.xlsx",sheet=1,
            col_names=TRUE,col_types=NULL,skip=0)
pip<-pip[,c(1,145:148,176,288,289,204:207,285)]
colnames(pip)[1]<-"ID"
all2<-all2[,c(1,22:59)]
all<-Reduce(function(x,y)merge(x,y,by.x="ID",by.y="ID",all=TRUE,sort=TRUE),
          list(all2,pip))

mean(all$cmv,na.rm=TRUE)

all$bmicata<-ifelse(all$BMI<=18.5,0,
                ifelse(all$BMI>=18.5&all$BMI<=24.9,1,
                     ifelse(all$BMI>=30,3,2)))

#women of child bearing age
wocba<-all[ which(all$gender==0 & all$Age>11 & all$Age<50),]

############Prevalence###############
#summary of women of child bearing age
(table(wocba$cmv))
wocba.p<-wocba[ which(wocba$cmv==1),]
wocba.n<-wocba[ which(wocba$cmv==0),]

############ Functions ##################
#t-test
ttest<-function(x){
  t.test((x==1),(all$cmv==1),paired=TRUE)
}
#Chi-Squared
chisq<-function(x,y){
  chi<-table(x,y)
  chisq.test(chi)
  ctab<-prop.table(chi)
  print(ctab)
  chi2<-table(x,y)
  ptab<-prop.table(chi2, margin=1)
  print(ptab)
```

```
  cbind(ctab,ptab)
}

#Chi-Squared Table
runit<-function(x,y){
  study<-chisq(x$safe,y)
  gen<-chisq(x$gender,y)
  race<-chisq(x$racecat,y)
  smnow<-chisq(x$smokenow,y)
  smever<-chisq(x$smokever,y)
  dia<-chisq(x$diabetes,y)
  dep<-chisq(x$depression,y)
  ast<-chisq(x$asthma,y)
  hp<-chisq(x$SerumHpylori,y)
  tg<-chisq(x$SerumTgondii,y)
  cmv<-chisq(x$cmv,y)

  calcs<-data.frame(names=c('study','study','gender','gender','race','race',
                      'smnow','smnow','smever','smever','dia','dia','dep','dep',
                      'ast','ast','hp','hp','tg','tg','cmv','cmv'),
                amounts=rbind.data.frame(study,gen,race,smnow,smever,
                                         dia,dep,ast,hp,tg,cmv))
  View(calcs)

  study.q<-chisq.test(x$safe,y)
  gen.q<-chisq.test(x$gender,y)
  race.q<-chisq.test(x$racecat,y)
  smnow.q<-chisq.test(x$smokenow,y)
  smever.q<-chisq.test(x$smokever,y)
  dia.q<-chisq.test(x$diabetes,y)
  dep.q<-chisq.test(x$depression,y)
  ast.q<-chisq.test(x$asthma,y)
  hp.q<-chisq.test(x$SerumHpylori,y)
  tg.q<-chisq.test(x$SerumTgondii,y)
  cmv.q<-chisq.test(x$cmv,y)

  ttestAge<-t.test(x$Age~y)
  print(ttestAge)
  ttestBMI<-t.test(x$BMI~y)
  print(ttestBMI)

  pval.q<-rbind(study.q$p.value,ttestAge$p.value,ttestBMI$p.value,gen.q$p.value,race.q$p.value,
              smnow.q$p.value,smever.q$p.value,dia.q$p.value,dep.q$p.value,
              ast.q$p.value,hp.q$p.value,tg.q$p.value,cmv.q$p.value)
  pval.p<-data.frame(names=c('study','Age','BMI','gender','race','smnow','smever','dia','dep',
                       'ast','hp','tg','cmv'),
                  amounts=pval.q)
  View(pval.p)
```

```
#BMI
x$bmicata<-ifelse(x$BMI<18.5,0,
                  ifelse(x$BMI>=18.5&x$BMI<30,1,
                         ifelse(x$BMI>=30,3,2)))

table(x$bmicata)
underweight<-length(which(x$bmicata==0 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
normal<-length(which(x$bmicata==1 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
overweight<-length(which(x$bmicata==2 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
obese<-length(which(x$bmicata==3 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
p.underweight<-length(which(x$bmicata==0 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.normal<-length(which(x$bmicata==1 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.overweight<-length(which(x$bmicata==2 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.obese<-length(which(x$bmicata==3 & y==1))/length(x$bmicata[!is.na(x$bmicata)])

underweight2<-length(which(x$bmicata==0 & y==0))/length(which(x$bmicata==0))
normal2<-length(which(x$bmicata==1 & y==0))/length(which(x$bmicata==1))
overweight2<-length(which(x$bmicata==2 & y==0))/length(which(x$bmicata==2))
obese2<-length(which(x$bmicata==3 & y==0))/length(which(x$bmicata==3))
p.underweight2<-length(which(x$bmicata==0 & y==1))/length(which(x$bmicata==0))
p.normal2<-length(which(x$bmicata==1 & y==1))/length(which(x$bmicata==1))
p.overweight2<-length(which(x$bmicata==2 & y==1))/length(which(x$bmicata==2))
p.obese2<-length(which(x$bmicata==3 & y==1))/length(which(x$bmicata==3))
bmicategory<-data.frame(names=c('underweight','normal','overweight','obese'),
                        prevalence1.n=c(underweight,normal,overweight,obese),
                        prevalence1.p=c(p.underweight,p.normal,p.overweight,p.obese),
                       prevalence2.n=c(underweight2,normal2,overweight2,obese2),
                       prevalence2.p=c(p.underweight2,p.normal2,p.overweight2,p.obese2))
View(bmicategory)

#AGE
x$agecata<-cut(x$Age,c(18,29,39,49,59,69,85))
one<-length(which(x$agecata=='(18,29]' & y==0))/length(x$agecata[!is.na(x$agecata)])
two<-length(which(x$agecata=='(29,39]' & y==0))/length(x$agecata[!is.na(x$agecata)])
three<-length(which(x$agecata=='(39,49]' & y==0))/length(x$agecata[!is.na(x$agecata)])
four<-length(which(x$agecata=='(49,59]' & y==0))/length(x$agecata[!is.na(x$agecata)])
five<-length(which(x$agecata=='(59,69]' & y==0))/length(x$agecata[!is.na(x$agecata)])
six<-length(which(x$agecata=='(69,85]' & y==0))/length(x$agecata[!is.na(x$agecata)])

p.one<-length(which(x$agecata=='(18,29]' & y==1))/length(x$agecata[!is.na(x$agecata)])
p.two<-length(which(x$agecata=='(29,39]' & y==1))/length(x$agecata[!is.na(x$agecata)])
p.three<-length(which(x$agecata=='(39,49]' & y==1))/length(x$agecata[!is.na(x$agecata)])
p.four<-length(which(x$agecata=='(49,59]' & y==1))/length(x$agecata[!is.na(x$agecata)])
p.five<-length(which(x$agecata=='(59,69]' & y==1))/length(x$agecata[!is.na(x$agecata)])
p.six<-length(which(x$agecata=='(69,85]' & y==1))/length(x$agecata[!is.na(x$agecata)])


one2<-length(which(x$agecata=='(18,29]' & y==0))/length(which(x$agecata=='(18,29]'))
```

```
 two2<-length(which(x$agecata=='(29,39]' & y==0))/length(which(x$agecata=='(29,39]'))
 three2<-length(which(x$agecata=='(39,49]' & y==0))/length(which(x$agecata=='(39,49]'))
 four2<-length(which(x$agecata=='(49,59]' & y==0))/length(which(x$agecata=='(49,59]'))
 five2<-length(which(x$agecata=='(59,69]' & y==0))/length(which(x$agecata=='(59,69]'))
 six2<-length(which(x$agecata=='(69,85]' & y==0))/length(which(x$agecata=='(69,85]'))

 p.one2<-length(which(x$agecata=='(18,29]' & y==1))/length(which(x$agecata=='(18,29]'))
 p.two2<-length(which(x$agecata=='(29,39]' & y==1))/length(which(x$agecata=='(29,39]'))
 p.three2<-length(which(x$agecata=='(39,49]' & y==1))/length(which(x$agecata=='(39,49]'))
 p.four2<-length(which(x$agecata=='(49,59]' & y==1))/length(which(x$agecata=='(49,59]'))
 p.five2<-length(which(x$agecata=='(59,69]' & y==1))/length(which(x$agecata=='(59,69]'))
 p.six2<-length(which(x$agecata=='(69,85]' & y==1))/length(which(x$agecata=='(69,85]'))

 agecategory<-data.frame(age=c('18-29','30-39','40-49','50-59','60-69','70-85'),
                  prevalence1.n=c(one,two,three,four,five,six),
                  prevalence1.p=c(p.one,p.two,p.three,p.four,p.five,p.six),
                  prevalence2.n=c(one2,two2,three2,four2,five2,six2),
                  prevalence2.p=c(p.one2,p.two2,p.three2,p.four2,p.five2,p.six2))
 View(agecategory)
}

#WOCBA- runit function without Gender
w.runit<-function(x,y){
 study<-chisq(x$safe,y)
 race<-chisq(x$racecat,y)
 smnow<-chisq(x$smokenow,y)
 smever<-chisq(x$smokever,y)
 dia<-chisq(x$diabetes,y)
 dep<-chisq(x$depression,y)
 ast<-chisq(x$asthma,y)
 hp<-chisq(x$SerumHpylori,y)
 tg<-chisq(x$SerumTgondii,y)
 cmv<-chisq(x$cmv,y)

 calcs<-data.frame(names=c('study','study','race','race',
                   'smnow','smnow','smever','smever','dia','dia','dep','dep',
                   'ast','ast','hp','hp','tg','tg','cmv','cmv'),
              amounts=rbind.data.frame(study,race,smnow,smever,
                              dia,dep,ast,hp,tg,cmv))
 View(calcs)

 study.q<-chisq.test(x$safe,y)
 race.q<-chisq.test(x$racecat,y)
 smnow.q<-chisq.test(x$smokenow,y)
 smever.q<-chisq.test(x$smokever,y)
 dia.q<-chisq.test(x$diabetes,y)
 dep.q<-chisq.test(x$depression,y)
 ast.q<-chisq.test(x$asthma,y)
```

```
hp.q<-chisq.test(x$SerumHpylori,y)
tg.q<-chisq.test(x$SerumTgondii,y)
cmv.q<-chisq.test(x$cmv,y)

ttestAge<-t.test(x$Age~y)
print(ttestAge)
ttestBMI<-t.test(x$BMI~y)
print(ttestBMI)

pval.q<-rbind(study.q$p.value,ttestAge$p.value,ttestBMI$p.value,race.q$p.value,smnow.q$p.value,
          smever.q$p.value,dia.q$p.value,dep.q$p.value,
          ast.q$p.value,hp.q$p.value,tg.q$p.value,cmv.q$p.value)
pval.p<-data.frame(names=c('study','Age','BMI','race','smnow','smever','dia','dep',
                   'ast','hp','tg','cmv'),
              amounts=pval.q)
View(pval.p)

#BMI
x$bmicata<-ifelse(x$BMI<18.5,0,
             ifelse(x$BMI>=18.5&x$BMI<30,1,
                  ifelse(x$BMI>=30,3,2)))

table(x$bmicata)
underweight<-length(which(x$bmicata==0 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
normal<-length(which(x$bmicata==1 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
overweight<-length(which(x$bmicata==2 & y==0))/length(x$bmicata[!is.na(x$bmicata)])
obese<-length(which(x$bmicata==3 & y==0))/length(x$bmicata[!is.na(x$bmicata)])

p.underweight<-length(which(x$bmicata==0 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.normal<-length(which(x$bmicata==1 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.overweight<-length(which(x$bmicata==2 & y==1))/length(x$bmicata[!is.na(x$bmicata)])
p.obese<-length(which(x$bmicata==3 & y==1))/length(x$bmicata[!is.na(x$bmicata)])

underweight2<-length(which(x$bmicata==0 & y==0))/length(which(x$bmicata==0))
normal2<-length(which(x$bmicata==1 & y==0))/length(which(x$bmicata==1))
overweight2<-length(which(x$bmicata==2 & y==0))/length(which(x$bmicata==2))
obese2<-length(which(x$bmicata==3 & y==0))/length(which(x$bmicata==3))

p.underweight2<-length(which(x$bmicata==0 & y==1))/length(which(x$bmicata==0))
p.normal2<-length(which(x$bmicata==1 & y==1))/length(which(x$bmicata==1))
p.overweight2<-length(which(x$bmicata==2 & y==1))/length(which(x$bmicata==2))
p.obese2<-length(which(x$bmicata==3 & y==1))/length(which(x$bmicata==3))
bmicategory<-data.frame(names=c('overweight','normal','overweight','obese'),
                  prevalence.n=c(overweight,normal,overweight,obese),
                  prevalence.p=c(p.overweight,p.normal,p.overweight,p.obese),
                  prevalence.n=c(overweight2,normal2,overweight2,obese2),
                  Prevalence.p=c(p.overweight2,p.normal2,p.overweight2,p.obese2))
View(bmicategory)
```

```
 #AGE
 x$agecata<-cut(x$Age,c(18,24,29,34,39,44,49))
 one<-length(which(x$agecata=='(18,24]' & y==0))/length(x$agecata[!is.na(x$agecata)])
 two<-length(which(x$agecata=='(24,29]' & y==0))/length(x$agecata[!is.na(x$agecata)])
 three<-length(which(x$agecata=='(29,34]' & y==0))/length(x$agecata[!is.na(x$agecata)])
 four<-length(which(x$agecata=='(34,39]' & y==0))/length(x$agecata[!is.na(x$agecata)])
 five<-length(which(x$agecata=='(39,44]' & y==0))/length(x$agecata[!is.na(x$agecata)])
 six<-length(which(x$agecata=='(44,49]' & y==0))/length(x$agecata[!is.na(x$agecata)])

 p.one<-length(which(x$agecata=='(18,24]' & y==1))/length(x$agecata[!is.na(x$agecata)])
 p.two<-length(which(x$agecata=='(24,29]' & y==1))/length(x$agecata[!is.na(x$agecata)])
 p.three<-length(which(x$agecata=='(29,34]' & y==1))/length(x$agecata[!is.na(x$agecata)])
 p.four<-length(which(x$agecata=='(34,39]' & y==1))/length(x$agecata[!is.na(x$agecata)])
 p.five<-length(which(x$agecata=='(39,44]' & y==1))/length(x$agecata[!is.na(x$agecata)])
 p.six<-length(which(x$agecata=='(44,49]' & y==1))/length(x$agecata[!is.na(x$agecata)])


 one2<-length(which(x$agecata=='(18,24]' & y==0))/length(which(x$agecata=='(18,24]'))
 two2<-length(which(x$agecata=='(24,29]' & y==0))/length(which(x$agecata=='(24,29]'))
 three2<-length(which(x$agecata=='(29,34]' & y==0))/length(which(x$agecata=='(29,34]'))
 four2<-length(which(x$agecata=='(34,39]' & y==0))/length(which(x$agecata=='(34,39]'))
 five2<-length(which(x$agecata=='(39,44]' & y==0))/length(which(x$agecata=='(39,44]'))
 six2<-length(which(x$agecata=='(44,49]' & y==0))/length(which(x$agecata=='(44,49]'))

 p.one2<-length(which(x$agecata=='(18,24]' & y==1))/length(which(x$agecata=='(18,24]'))
 p.two2<-length(which(x$agecata=='(24,29]' & y==1))/length(which(x$agecata=='(24,29]'))
 p.three2<-length(which(x$agecata=='(29,34]' & y==1))/length(which(x$agecata=='(29,34]'))
 p.four2<-length(which(x$agecata=='(34,39]' & y==1))/length(which(x$agecata=='(34,39]'))
 p.five2<-length(which(x$agecata=='(39,44]' & y==1))/length(which(x$agecata=='(39,44]'))
 p.six2<-length(which(x$agecata=='(44,49]' & y==1))/length(which(x$agecata=='(44,49]'))

 agecategory<-data.frame(age=c('18-24','24-29','29-34','34-39','39-44','44-49'),
                         prevalence.n=c(one,two,three,four,five,six),
                         prevalence.p=c(p.one,p.two,p.three,p.four,p.five,p.six),
                         prevalence.n=c(one2,two2,three2,four2,five2,six2),
                         Prevalence.p=c(p.one2,p.two2,p.three2,p.four2,p.five2,p.six2))

 View(agecategory)
}

#Find n values for the chi.squared tests
w.aly<-function(x,y){
 a<-length(which(x$cmv==0 & y==0))
 b<-length(which(x$cmv==0 & y==1))
 c<-length(which(x$cmv==1 & y==0))
 d<-length(which(x$cmv==1 & y==1))
 w.cmv<-data.frame(names=c('-','+'),
```

```
            cmvn=c(a,b),
            cmvp=c(c,d))
  View(w.cmv)
}

#n for BMI
w.bmi<-function(x,y){
  a<-length(which(x$cmv==0 & y==0))
  b<-length(which(x$cmv==0 & y==1))
  c<-length(which(x$cmv==0 & y==2))
  d<-length(which(x$cmv==0 & y==3))
  e<-length(which(x$cmv==1 & y==0))
  f<-length(which(x$cmv==1 & y==1))
  g<-length(which(x$cmv==1 & y==2))
  h<-length(which(x$cmv==1 & y==3))
  w.bmi<-data.frame(names=c('underweight','normal','overweight','obese'),
            cmvn=c(a,b,c,d),
            cmvp=c(e,f,g,h))
  View(w.bmi)
}

#Histogram Prevalence
# Age
hist.all.p<-function(x,y){
  duration = x
  breaks = c(18,29,39,49,59,69,85)
  duration.cut = cut(duration, breaks, right=FALSE)
  duration.freq = table(duration.cut)
  duration2 = y
  breaks2 = c(18,29,39,49,59,69,85)
  duration.cut2 = cut(duration2, breaks2, right=FALSE)
  duration.freq2 = table(duration.cut2)

  prev = duration.freq2/duration.freq
}
# AGE-WOCBA
hist.wocba.p<-function(x,y){
  duration = x
  breaks = c(18,20,25,30,35,40,45,49)
  duration.cut = cut(duration, breaks, right=FALSE)
  duration.freq = table(duration.cut)
  duration2 = y
  breaks2 = c(18,20,25,30,35,40,45,49)
  duration.cut2 = cut(duration2, breaks2, right=FALSE)
  duration.freq2 = table(duration.cut2)

  prev = duration.freq2/duration.freq
}
```

```
# BMI
hist.all.bmi<-function(x,y){
  duration = x
  breaks = c(14,18.5,25,30,73)
  duration.cut = cut(duration, breaks, right=FALSE)
  duration.freq = table(duration.cut)
  duration2 = y
  breaks2 = c(14,18.5,25,30,73)
  duration.cut2 = cut(duration2, breaks2, right=FALSE)
  duration.freq2 = table(duration.cut2)

  prev = duration.freq2/duration.freq
}

#n for Age
w.age<-function(x,y){
  a<-length(which(x$cmv==0 & y=='(18,29]'))
  b<-length(which(x$cmv==0 & y=='(29,39]'))
  c<-length(which(x$cmv==0 & y=='(39,49]'))
  d<-length(which(x$cmv==0 & y=='(49,59]'))
  e<-length(which(x$cmv==0 & y=='(59,69]'))
  f<-length(which(x$cmv==0 & y=='(69,85]'))
  g<-length(which(x$cmv==1 & y=='(18,29]'))
  h<-length(which(x$cmv==1 & y=='(29,39]'))
  i<-length(which(x$cmv==1 & y=='(39,49]'))
  j<-length(which(x$cmv==1 & y=='(49,59]'))
  k<-length(which(x$cmv==1 & y=='(59,69]'))
  l<-length(which(x$cmv==1 & y=='(69,85]'))
  w.age<-data.frame(names=c('18-29','30-39','40-49','50-59','60-69','70-85'),
                    cmvn=c(a,b,c,d,e,f),
                    cmvp=c(g,h,i,j,k,l))
  View(w.age)
}

ww.age<-function(x,y){
  a<-length(which(x$cmv==0 & y=='(18,24]'))
  b<-length(which(x$cmv==0 & y=='(24,29]'))
  c<-length(which(x$cmv==0 & y=='(29,34]'))
  d<-length(which(x$cmv==0 & y=='(34,39]'))
  e<-length(which(x$cmv==0 & y=='(39,44]'))
  f<-length(which(x$cmv==0 & y=='(44,49]'))
  g<-length(which(x$cmv==1 & y=='(18,24]'))
  h<-length(which(x$cmv==1 & y=='(24,29]'))
  i<-length(which(x$cmv==1 & y=='(29,34]'))
  j<-length(which(x$cmv==1 & y=='(34,39]'))
  k<-length(which(x$cmv==1 & y=='(39,44]'))
  l<-length(which(x$cmv==1 & y=='(44,49]'))
```

```
  ww.age<-data.frame(names=c('18-24','25-29','30-34','35-39','40-44','45-49'),
                cmvn=c(a,b,c,d,e,f),
                cmvp=c(g,h,i,j,k,l))
  View(ww.age)
}

############ Processing prior to Analysis ##############
safe<-all
vip2<-safe
vip2$SAA_rawcata<-cut(vip2$saa_ngml,breaks=4)
vip2$CRP_rawcata<-cut(vip2$crp_ngml,breaks=4)
vip2$VCAM_rawcata<-cut(vip2$vcam_ngml,breaks=4)
vip2$ICAM_rawcata<-cut(vip2$icam_ngml,breaks=4)
saa<-data.frame(table(vip2$SAA_rawcata))
crp<-data.frame(table(vip2$CRP_rawcata))
vcam<-data.frame(table(vip2$VCAM_rawcata))
icam<-data.frame(table(vip2$ICAM_rawcata))

vip2$lbmi<-log(vip2$BMI)

#VIP2- control variables
vip2$smoker<-ifelse(vip2$smokenow==1 | vip2$smokever==1,1,0)
ageV2<-vip2$Age
gendV2<-vip2$gender
raceV2<-vip2$racecat
smokeV2<-vip2$smoker
eduV2<-vip2$edcat
lbmiV2<-vip2$lbmi

#VIP 3- only positive
vip3<-vip2[ which(vip2$cmv==1),]
#vip4-only negative
vip4<-vip2[ which(vip2$cmv==0),]

vip2$cmv_cat2<-ifelse(vip2$cmv_cat=="Negative",0,vip2$cmv_cat)

#means of Age and BMI-Chi-squared analysis
vip2$age_cat2<-cut(vip2$Age,breaks=c(18,29,39,49,59,69,85))
vip2$BMI_cat2<-cut(vip2$BMI,breaks=c(14.5,18.5,25,30,75))
chisq.test(vip2$cmv,vip2$age_cat2)
chisq.test(vip2$cmv,vip2$BMI_cat2)

################## Analysis ########################
#WOCBA Analysis
w.runit(wocba,wocba$cmv)

mean(wocba$Age)
median(wocba$Age)
```

```
#WOCBA, CMV n values
w.aly(wocba,wocba$safe)
w.aly(wocba,wocba$racecat)
w.aly(wocba,wocba$smokenow)
w.aly(wocba,wocba$smokever)
w.aly(wocba,wocba$diabetes)
w.aly(wocba,wocba$depression)
w.aly(wocba,wocba$asthma)
w.aly(wocba,wocba$SerumHpylori)
w.aly(wocba,wocba$SerumTgondii)
w.aly(wocba,wocba$cmv)

w.aly(all,all$safe)
w.aly(all,all$gender)
w.aly(all,all$racecat)
w.aly(all,all$smokenow)
w.aly(all,all$smokever)
w.aly(all,all$diabetes)
w.aly(all,all$depression)
w.aly(all,all$asthma)
w.aly(all,all$SerumHpylori)
w.aly(all,all$SerumTgondii)
w.aly(all,all$cmv)

w.bmi(all,all$bmicata)
w.bmi(wocba,wocba$bmicata)

wocba$age_cat2<-cut(wocba$Age,c(18,24,29,34,39,44,49))
ww.age(wocba,wocba$age_cat2)


#All Analysis
runit(all,all$cmv)

############ Histogram ##############
#Histogram CMV- Age & BMI prevalence
par(mfrow=c(2,2))
cmv.prev<-hist.all.p(all$Age,cmv.p$Age)
barplot(cmv.prev,col="black",main="Total population HCMV Prevalence by
age",ylim=c(0,1),xlab="Age",ylab="HCMV Prevalence")
cmv.prev.w<-hist.wocba.p(wocba$Age,cmv.w$Age)
barplot(cmv.prev.w,col="black",main="WOCBA subpopulation HCMV Prevalence by
age",ylim=c(0,1),xlab="Age",ylab="HCMV Prevalence")
cmv.bmi<-hist.all.bmi(all$BMI,cmv.p$BMI)
barplot(cmv.bmi,col="dark gray",main="Total population HCMV Prevalence by
BMI",ylim=c(0,1),xlab="BMI",ylab="HCMV Prevalence")
cmv.bmi.w<-hist.all.bmi(wocba$BMI,cmv.w$BMI)
```

```
barplot(cmv.bmi.w,col="dark gray",main="WOCBA subpopulation HCMV Prevalence by
BMI",ylim=c(0,1),xlab="BMI",ylab="HCMV Prevalence")


############ Tests for Normality ###########
#Test for normality
shapiro.crp<-shapiro.test(vip2$crp_ngml)
shapiro.saa<-shapiro.test(vip2$saa_ngml)
shapiro.vcam<-shapiro.test(vip2$vcam_ngml)
shapiro.icam<-shapiro.test(vip2$icam_ngml)
shapiro.cmv.p<-shapiro.test(vip3$cmv_rat)
shapiro.cmv.n<-shapiro.test(vip4$cmv_rat)
shapiro.bmi<-shapiro.test(vip2$BMI)
shapiro<-data.frame(names=c("crp","saa","vcam","icam","cmv+","cmv-","bmi"),
   W.statistic=c(shapiro.crp$statistic,shapiro.saa$statistic,shapiro.vcam$statistic,
          shapiro.icam$statistic,shapiro.cmv.p$statistic,shapiro.cmv.n$statistic,
          shapiro.bmi$statistic),
   p.value=c(shapiro.crp$p.value,shapiro.saa$p.value,shapiro.vcam$p.value,
         shapiro.icam$p.value,shapiro.cmv.p$p.value,shapiro.cmv.n$p.value,
         shapiro.bmi$p.value))
View(shapiro)

#Q-Q plot to verify normality
par(mfrow=c(4,2))
qqnorm(vip2$crp_ngml,main="CRP (ng/mL),Normal Q-Q Plot")
qqline(vip2$crp_ngml)
qqnorm(vip2$saa_ngml,main="SAA (ng/mL),Normal Q-Q Plot")
qqline(vip2$saa_ngml)
qqnorm(vip2$vcam_ngml,main="VCAM (ng/mL),Normal Q-Q Plot")
qqline(vip2$vcam_ngml)
qqnorm(vip2$icam_ngml,main="ICAM (ng/mL),Normal Q-Q Plot")
qqline(vip2$icam_ngml)
qqnorm(vip4$cmv_rat,main="HCMV-,Normal Q-Q Plot")
qqline(vip4$cmv_rat)
qqnorm(vip3$cmv_rat,main="HCMV+,Normal Q-Q Plot")
qqline(vip3$cmv_rat)
qqnorm(vip2$BMI,main="BMI,Normal Q-Q Plot")
qqline(vip2$BMI)

#Log10 transformed Normality
#Test for normality
shapiro.lcrp<-shapiro.test(vip2$lcrp)
shapiro.lsaa<-shapiro.test(vip2$lsaa)
shapiro.lvcam<-shapiro.test(vip2$lvcam)
shapiro.licam<-shapiro.test(vip2$licam)
shapiro.lcmv.p<-shapiro.test(vip3$lcmv_rat)
shapiro.lcmv.n<-shapiro.test(vip4$lcmv_rat)
shapiro.lbmi<-shapiro.test(vip2$lbmi)
```

```
shapirolog<-data.frame(names=c("log10 crp","log10 saa","log10 vcam","log10 icam","log cmv+","log
cmv-","log bmi"),
                W.statistic=c(shapiro.lcrp$statistic,shapiro.lsaa$statistic,shapiro.lvcam$statistic,
                        shapiro.licam$statistic,shapiro.lcmv.p$statistic,shapiro.lcmv.n$statistic,
                        shapiro.lbmi$statistic),
                p.value=c(shapiro.lcrp$p.value,shapiro.lsaa$p.value,shapiro.lvcam$p.value,
                      shapiro.licam$p.value,shapiro.lcmv.p$p.value,shapiro.lcmv.n$p.value,
                      shapiro.lbmi$p.value))
View(shapirolog)
#Q-Q plot to verify normality
par(mfrow=c(4,2))
qqnorm(vip2$lcrp,main="log10 CRP (ng/mL),Normal Q-Q Plot")
qqline(vip2$lcrp)
qqnorm(vip2$lsaa,main="log10 SAA (ng/mL),Normal Q-Q Plot")
qqline(vip2$lsaa)
qqnorm(vip2$lvcam,main="log10 VCAM (ng/mL),Normal Q-Q Plot")
qqline(vip2$lvcam)
qqnorm(vip2$licam,main="log10 ICAM (ng/mL),Normal Q-Q Plot")
qqline(vip2$licam)
qqnorm(vip4$lcmv_rat,main="log10 HCMV-,Normal Q-Q Plot")
qqline(vip4$lcmv_rat)
qqnorm(vip3$lcmv_rat,main="log10 HCMV+,Normal Q-Q Plot")
qqline(vip3$lcmv_rat)
qqnorm(vip2$lbmi,main="log10 BMI,Normal Q-Q Plot")
qqline(vip2$lbmi)


#VIP-LM Analysis
#correct for age, race, bmi, gender,smoke
#binomial CMV
lm.cmv.crpD<-summary(lm(vip2$lcrp~vip2$cmv+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.cmv.crpD$coefficients)
lm.cmv.saaD<-summary(lm(vip2$lsaa~vip2$cmv+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.cmv.saaD$coefficients)
lm.cmv.vcamD<-summary(lm(vip2$lvcam~vip2$cmv+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.cmv.vcamD$coefficients)
lm.cmv.icamD<-summary(lm(vip2$licam~vip2$cmv+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.cmv.icamD$coefficients)

#Positive ratio, negative set to 0
lm.crp.cl<-summary(lm(vip2$lcrp~vip2$lcmv_rat_pos+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.crp.cl$coefficients)
lm.saa.cl<-summary(lm(vip2$lsaa~vip2$lcmv_rat_pos+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.saa.cl$coefficients)
lm.vcam.cl<-summary(lm(vip2$lvcam~vip2$lcmv_rat_pos+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.vcam.cl$coefficients)
lm.icam.cl<-summary(lm(vip2$licam~vip2$lcmv_rat_pos+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.icam.cl$coefficients)
```

```
#Tertiary Analysis CMV
lm.crp.tert<-summary(lm(vip2$lcrp~vip2$cmv_cat2+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.crp.tert$coefficients)
lm.saa.tert<-summary(lm(vip2$lsaa~vip2$cmv_cat2+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.saa.tert$coefficients)
lm.vcam.tert<-summary(lm(vip2$lvcam~vip2$cmv_cat2+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.vcam.tert$coefficients)
lm.icam.tert<-summary(lm(vip2$licam~vip2$cmv_cat2+lbmiV2+raceV2+ageV2+smokeV2+gendV2))
View(lm.icam.tert$coefficients)




########## piecewise plots #########
par(mfrow=c(4,2))
#CRP
plot(vip2$lcrp~vip2$lcmv_rat,subset=vip2$cmv==0,ylim=c(1,6.5),main='CRP (HCMV-)',xlab="log10
HCMV ratio",ylab="log10 CRP",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lcrp~vip2$lcmv_rat,subset=vip2$cmv==0),col="black")
plot(vip2$lcrp~vip2$lcmv_rat,subset=vip2$cmv==1,ylim=c(1,6.5),main='CRP (HCMV+)',xlab="log10
HCMV ratio",ylab="log10 CRP",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lcrp~vip2$lcmv_rat,subset=vip2$cmv==1),col="dark gray")
#SAA
plot(vip2$lsaa~vip2$lcmv_rat,subset=vip2$cmv==0,ylim=c(1,6.5),main='SAA (HCMV-)',xlab="log10
HCMV ratio",ylab="log10 SAA",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lsaa~vip2$lcmv_rat,subset=vip2$cmv==0),col="black")
plot(vip2$lsaa~vip2$lcmv_rat,subset=vip2$cmv==1,ylim=c(1,6.5),main='SAA (HCMV+)',xlab="log10
HCMV ratio",ylab="log10 SAA",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lsaa~vip2$lcmv_rat,subset=vip2$cmv==1),col="dark gray")
#VCAM
plot(vip2$lvcam~vip2$lcmv_rat,subset=vip2$cmv==0,ylim=c(1.5,3.5),main='VCAM (HCMV-)',xlab="log10
HCMV ratio",ylab="log10 VCAM-1",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lvcam~vip2$lcmv_rat,subset=vip2$cmv==0),col="black")
plot(vip2$lvcam~vip2$lcmv_rat,subset=vip2$cmv==1,ylim=c(1.5,3.5),main='VCAM
(HCMV+)',xlab="log10 HCMV ratio",ylab="log10 VCAM-1",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$lvcam~vip2$lcmv_rat,subset=vip2$cmv==1),col="dark gray")
#ICAM
plot(vip2$licam~vip2$lcmv_rat,subset=vip2$cmv==0,ylim=c(1.5,3.5),main='ICAM (HCMV-)',xlab="log10
HCMV ratio",ylab="log10 ICAM-1",
```

```
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$licam~vip2$lcmv_rat,subset=vip2$cmv==0),col="black")
plot(vip2$licam~vip2$lcmv_rat,subset=vip2$cmv==1,ylim=c(1.5,3.5),main='ICAM (HCMV+)',xlab="log10
HCMV ratio",ylab="log10 ICAM-1",
    col=ifelse(vip2$cmv==1,"dark gray","black"))
abline(lm(vip2$licam~vip2$lcmv_rat,subset=vip2$cmv==1),col="dark gray")
```

APPENDIX E: PRACTICUM REPORT

Fish Exposure Data Imputation
Jennifer Styles' Practicum
Preceptor: Joachim D. Pleil, Ph.D. US EPA

During this Practicum, I analyzed fish exposure data measured in the fatty tissues of fish from streams around the country. Specifically, I focused on the concentrations of pesticides and their residuals that were measured in the fish tissue. This project sought to impute (a process that substitutes missing data with calculated values) missing fish exposure to pesticides data. While this type of data does not replace measured data, it helps fill in missing data that is necessary for robust analysis. Throughout this practicum, excellent oral and written communication skills were required to work in a professional environment beyond the classroom. Problem solving and sharing of information was vital, as were time-management and organizational skills. This fish exposure data was collected to be used to assess the amount of pesticides in streams in various locations around the country. Our goal was to provide a more expansive dataset to assess risk, study the human impact on the environment and protect human health.

All communication with my preceptor was either written or oral and to be understood, effective communication was necessary. While we discussed how to appropriately relay imputed data and our findings to lay-persons it did not end up being the focus of this project. Problem solving, idea sharing and discussion was vital to the success of this project. Neither my preceptor nor I specialize in fish data so we were both challenged to learn about fish and find a reasonable method that could logically estimate missing data. Data can be missing due to a variety of reasons such as limited fish/sample size, below limit of quantitation, analytical error, etc., but in this analysis left-censored (below limit of quantitation) data was imputed. This missing data can lead to massive data gaps and leave out information that could better assess

risk, and protect environmental and public health. Properly imputing below limit of quantitation data has been a subject of debate, do you exclude it, set it to zero, set it to all the same value or is there a better way? During this practicum, data was imputed by ranking data in the area below the quantifiable limit and assigning logically ranked values. Finding a way to logically rank the missing data required much discussion and sharing of ideas to develop an appropriate ranking method. In the limited hours in which this project was completed, organizational and time management skills were required. Organization, especially proper data management, was very important so that my work was clearly documented and could be used by others, and that my preceptor could figure out what I was trying to send him or explain.

This project was centered around assessing environmental hazards that pose risks to human health and safety. The pesticides and their residuals, measured in the exposed fish, pose known risks to human health and safety. It was important that our assessment method use an accurate representation of estimated levels in fish so as not to skew data. Applying this method of imputing left-censored data, has the potential to improve upon other methods that assign all missing values with the same amount. This method provides a more accurate distribution of data below the limit of quantitation. These pesticides and their residuals imputed in this study can be used to better assess the impact of pesticides on human health and the environment.

References

(CDC), C. for D. C. and P. (2010). Healthy weigh, overweight, and obesity among U.S. adults. *National Health and Nutrition*, 2.

(CDC), C. for D. C. and P. (2016). Leading Causes of Death. Retrieved March 11, 2017, from https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

American Heart Association, I. (2014). What is Cardiovascular Disease? Retrieved March 11, 2017, from http://www.heart.org/HEARTORG/Support/What-is-Cardiovascular-Disease_UCM_301852_Article.jsp#.WMRNJ_nyvX4

Bate, S. L., Dollard, S. C., & Cannon, M. J. (2010). Cytomegalovirus seroprevalence in the United States: the national health and nutrition examination surveys, 1988-2004. *Clinical Infectious Diseases*, *50*(11), 1439–47. https://doi.org/10.1086/652438

Baumann, R., Gube, M., Markert, A., Davatgarbenam, S., Kossack, V., Gerhards, B., … Brand, P. (2017). Systemic serum amyloid A as a biomarker for exposure to zinc and / or copper-containing metal fumes. *Nature Publishing Group*, (December 2016), 1–8. https://doi.org/10.1038/jes.2016.86

BD. (2016). BD Vacutainer Blood and Urine Collection- FAQs. Retrieved April 17, 2017, from http://www.bd.com/vacutainer/faqs/

Betjes, M. G. H., Litjens, N. H. R., & Zietse, R. (2007). Seropositivity for cytomegalovirus in patients with end-stage renal disease is strongly associated with atherosclerotic disease. *Nephrology, Dialysis, Transplantation : Official Publication of the European Dialysis and Transplant Association - European Renal Association*, *22*(June), 3298–3303. https://doi.org/10.1093/ndt/gfm348

Bialas, K. M., Swamy, G. K., Permar, S. R., & Permar Kristybialas, S. R. (2015). Perinatal cytomegalovirus and varicella zoster virus infections: epidemiology, prevention, and treatment HHS Public Access. *Clin Perinatol*, *42*(1), 1–19. https://doi.org/10.1016/j.clp.2014.10.006

Bind, M.-A., Baccarelli, A., Zanobetti, A., Tarantini, L., Suh, H., Vokonas, P., & Schwartz, J. (2013). Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort. *Epidemiology*, *23*(2), 332–340. https://doi.org/10.1097/EDE.0b013e31824523f0.Air

Biron, K. K. (2006). Antiviral drugs for cytomegalovirus diseases. *Antiviral Research*, *71*(2–3 SPEC. ISS.), 154–163. https://doi.org/10.1016/j.antiviral.2006.05.002

Blankenberg, S., Barbaux, S., & Tiret, L. (2003). Adhesion molecules and atherosclerosis. *Atherosclerosis*, *170*(2), 191–203. https://doi.org/10.1016/S0021-9150(03)00097-2

Brook, R. D. (2008). Cardiovascular effects of air pollution. *Clinical Science (London, England : 1979)*, *115*(6), 175–87. https://doi.org/10.1042/CS20070444

Cannon, M. J., Schmid, D. S., & Hyde, T. B. (2010). Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Reviews in Medical Virology*, *20*(1), 202–213. https://doi.org/10.1002/rmv

CDC/National Center for Health Statistics. (2015). NHANES - About the National Health and Nutrition Examination Survey. Retrieved July 21, 2017, from https://www.cdc.gov/nchs/nhanes/about_nhanes.htm

Centers for Disease Control and Prevention. (2016). CMV | Overview | Cytomegalovirus and Congenital CMV Infection | CDC. Retrieved April 20, 2017, from https://www.cdc.gov/cmv/overview.html

Crough, T., & Khanna, R. (2009). Immunobiology of Human Cytomegalovirus : from Bench to Bedside. *Clinical Microbiology Reviews*, *22*(1), 76–98. https://doi.org/10.1128/CMR.00034-08

Fabricant, C., & Fabricant, J. (1999). Atherosclerosis induced by infection with Marek's disease herpesvirus in chickens. *American Heart Journal*, *138*(5 Pt 2), S465-8. https://doi.org/a101710 [pii]

Farber, S., & Wolbach, S. B. (1932). Intranuclear and Cytoplasmic Inclusions ("Protozoan-like Bodies") In the Salivary Glands and Other Organs of Infants. *The American Journal of Pathology*, *8*(2), 123–135.

Fateh-Moghadam, S., Bocksch, W., Wessely, R., Jäger, G., Hetzer, R., & Gawaz, M. (2003). Cytomegalovirus infection status predicts progression of heart-transplant vasculopathy. *Transplantation*, *76*(10), 1470–1474. https://doi.org/10.1097/01.TP.0000090163.48433.48

Hiscock, R., Bauld, L., Amos, A., Fidler, J. A., & Munafò, M. (2012). Socioeconomic status and smoking: A review. *Annals of the New York Academy of Sciences*, *1248*(1), 107–123. https://doi.org/10.1111/j.1749-6632.2011.06202.x

Ho, M. (2008). The history of cytomegalovirus and its diseases. *Medical Microbiology and Immunology*, *197*(2), 65–73. https://doi.org/10.1007/s00430-007-0066-x

Hyde, T. B., Schmid, D. S., & Cannon, M. J. (2010). Cytomegalovirus seroconversion rates and risk factors: implications for congenital CMV. *Reviews in Medical Virology*, *20*(1), 311–326. https://doi.org/10.1002/rmv

Kaptoge, S., Angelantonio, E. D., & Pennells, L. (2012). C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction. *The New England Journal of Medicine*, 1310–1320. https://doi.org/10.1056/NEJMoa1107477

Knox, G. E. (1983). Cytomegalovirus: Import of Sexual Transmission. *Clinical Obstetrics and Gynecology*, *26*(1).

Lanzieri, T. M., Kruszon-Moran, D., Amin, M. M., Bialek, S. R., Cannon, M. J., Carroll, M. D., & Dollard, S. C. (2015). Seroprevalence of cytomegalovirus among children 1 to 5 years of age in the United States from the National Health and Nutrition Examination Survey of 2011 to 2012. *Clinical and Vaccine Immunology : CVI*, *22*(2), 245–7. https://doi.org/10.1128/CVI.00697-14

Mehta, S. K., Laudenslager, M. L., Stowe, R. P., Crucian, B. E., Sams, C. F., & Pierson, D. L. (2014). Multiple latent viruses reactivate in astronauts during Space Shuttle missions. *Brain, Behavior, and Immunity*, *41*(1), 210–217. https://doi.org/10.1016/j.bbi.2014.05.014

Pak, C., Mcarthur, R., Eun, H.-M., & Yoon, J.-W. (1988). ASSOCIATION OF CYTOMEGALOVIRUS INFECTION WITH AUTOIMMUNE TYPE 1 DIABETES. *The Lancet*, *332*(8601), 1–4. https://doi.org/10.1016/S0140-6736(88)92941-8

Popović, M., Smiljanić, K., Dobutović, B., Syrovets, T., Simmet, T., & Isenović, E. R. (2012). Human cytomegalovirus infection and atherothrombosis. *Journal of Thrombosis and Thrombolysis*, *33*(2), 160–172. https://doi.org/10.1007/s11239-011-0662-x

Prince, H. E., & Lapé-Nixon, M. (2014). Role of cytomegalovirus (CMV) IgG avidity testing in diagnosing primary CMV infection during pregnancy. *Clinical and Vaccine Immunology*, *21*(10), 1377–1384. https://doi.org/10.1128/CVI.00487-14

Riley, H. D. J. (1997). History of Cytomegalovirus.pdf. *Southern Medical Journal*.

Schleiss, M. R. (2005). Antiviral therapy of congenital cytomegalovirus infection. *Seminars in Pediatric Infectious Diseases*, *16*(1 SPEC.ISS.), 50–59. https://doi.org/10.1053/j.spid.2004.09.012

Schleiss, M. R. (2008). Cytomegalovirus vaccine development. *Current Topics in Microbiology and Immunology*, *325*, 361–382. https://doi.org/10.1007/978-3-540-77349-8_20

Schleiss, M. R. (2016). Preventing Congenital Cytomegalovirus Infection: Protection to a "T." *Trends in Microbiology*, *24*(3), 170–172. https://doi.org/10.1016/j.tim.2016.01.007

Simanek, A. M., Dowd, J. B., Pawelec, G., Melzer, D., Dutta, A., & Aiello, A. E. (2011). Seropositivity to Cytomegalovirus , Inflammation , All- Cause and Cardiovascular Disease-Related Mortality in the United States, *6*(2). https://doi.org/10.1371/journal.pone.0016103

Söderberg-Nauclér, C. (2006). Does cytomegalovirus play a causative role in the development of various inflammatory diseases and cancer? *Journal of Internal Medicine*, *259*(3), 219–246. https://doi.org/10.1111/j.1365-2796.2006.01618.x

Span, A. H., Grauls, G., Bosman, F., van Boven, C. P., & Bruggeman, C. A. (1992). Cytomegalovirus infection induces vascular injury in the rat. *Atheroscl*, *93*(1–2), 41–52.

Terrazzini, N., Bajwa, M., Thomas, D., Smith, H., & Kern, F. (2014). Gender differences and age-specific associations between body mass index and other cardiovascular risk factors in CMV infected and uninfected people. *Immunology Letters*, *162*(1), 316–322. https://doi.org/10.1016/j.imlet.2014.09.011

Walker, A., & Braglia, L. (2017). Package " openxlsx " Read, Write and Edit XLSX Files.

Wang, C., Dollard, S. C., Amin, M. M., & Bialek, S. R. (2016). Cytomegalovirus IgM seroprevalence among women of reproductive age in the United States. *PLoS ONE*, *11*(3), 8–15. https://doi.org/10.1371/journal.pone.0151996

Weller, T. H., & Hanshaw, J. B. (1962). Virologic and Clinical Observations on Cytomegalic Inclusion Disease. *The New England Journal of Medicine*, *266*(24), 1233–1244.

WHO | Infertility definitions and terminology. (2016). Retrieved April 25, 2017, from http://www.who.int/reproductivehealth/topics/infertility/definitions/en/

Wickham, H. (2016). Package " plyr " Tools for Splitting, Applying and Combining Data.

Wickham, H., Bryan, J., RStudio, Kalicinski, M., Valery, K., Leitienne, C., … Hoerl, D. (2017). Package " readxl " Read Excel Files.

Wickham, H., & RStudio. (2016). Package " scales " Scale Functions for Visualization.

Zakynthinos, E., & Pappa, N. (2009). Inflammatory biomarkers in coronary artery disease. *Journal of Cardiology*, *53*, 317–333. https://doi.org/10.1016/j.jjcc.2008.12.007