Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Xinyi Zhao Date

The Associations between PFOA, PFOS and Cholesterol Levels in US Pregnant Women: NHANES 2003-2006

By

Xinyi Zhao

Master of Public Health
Epidemiology

_____

Lyndsey Darrow
Faculty Thesis Advisor

The Associations between PFOA, PFOS and Cholesterol Levels in US Pregnant Women: NHANES 2003-2006

By

Xinyi Zhao

B.Med.Sc., B.Sc.
Southern Medical University
2014

Faculty Thesis Advisor: Lyndsey Darrow, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Abstract

The Associations between PFOA, PFOS and Cholesterol Levels
in US Pregnant Women: NHANES 2003-2006

By Xinyi Zhao

Background: Perfluorooctanoic acid (PFOA) and perfluorooctane sulfate (PFOS) are two of the most studied types of perfluoroalkyl compounds which have been widely used in assorted consumer and industrial products. Several epidemiologic studies have shown positive statistically significant associations between PFOA and cholesterols as well as between PFOS and cholesterols, which may result in various adverse health effects. However, such associations during pregnancy may be different.
Methods: A cross-sectional study was performed on 174 pregnant women participating in the 2003-2006 National Health and Nutrition Examination Survey (NHANES). PFOA, PFOS and cholesterols (total cholesterol, high-density lipoprotein [HDL] and low-density lipoprotein [LDL]) were measured in serum. Linear regression was performed to determine the associations between PFOA, PFOS and cholesterol levels.
Results: A significant positive association between PFOS and LDL was found ($p < 0.05$), while the associations between PFOA and cholesterols as well as between PFOS and non-LDL cholesterols were not significant ($p > 0.05$). One ln-ng/mL increase in PFOS was associated with 21.0 (95% CI: 0.6, 41.4) increase in LDL, adjusting for age, race, education level, annual family income, BMI, total fat intake, serum cotinine, month of pregnancy and number of former pregnancies as confounders.
Conclusion: We observed little evidence that PFOA and PFOS affect cholesterol levels during pregnancy, with the possible exception of the effect of PFOS on LDL. However, since PFCs have been associated with other adverse health effects including pregnancy-induced hypertension and low birth weight, controlling the exposure to PFOA and PFOS is still highly necessary for pregnant women.

The Associations between PFOA, PFOS and Cholesterol Levels in US Pregnant Women: NHANES 2003-2006

By

Xinyi Zhao

B.Med.Sc., B.Sc.
Southern Medical University
2014

Faculty Thesis Advisor: Lyndsey Darrow, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Acknowledgement

I would first like to express my sincere gratitude to my faculty thesis advisor Dr. Lyndsey Darrow for her excellent guidance, encouragement, caring, and providing me with an excellent atmosphere for doing the thesis. Her patient, detailed guidance helped me in all the time of working on this thesis, including developing the research topic, collecting the data, designing the study, performing the statistical analysis and writing the thesis. I enjoyed working with her very much and have acquired a lot of valuable suggestions and skills from her on the real-life research. Also, I have learned a lot about the areas of environmental epidemiology and chemical exposures.

I would also like to thank Dr. David Kleinbaum and Dr. Eli Rosenberg, who were the instructors of the Epidemiologic Methods class and the Epidemiologic Modeling class of our department. Their classes gave me a systematic, detailed guidance of the methods for epidemiologic analysis. Most methods used for my thesis analysis were learned from their classes.

Finally, I must express my very profound gratitude to my parents, my boyfriend and my friends for providing me with unfailing support and continuous encouragement throughout my years of study at Emory and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you very much.

CONTENTS

BACKGROUND

Perfluoroalkyl compounds (PFCs) are a class of perfluorinated chemicals that have been widely used in assorted consumer and industrial products [1]. The particular chemical structure of PFCs is a fluorinated carbon backbone of different length that is terminated by a carboxylate or sulfonate functional group [2]. This amphipathic structure is the foundation for the properties of water and oil repellency and stain resistance [3], and thus to possess a lot of uses. The chemical structure also leads to persistence in the environment. Since the 1950s, PFCs have been used extensively in commercial products, such as manufacture of lubricants, surfactants, textile coatings and paper, food packaging, polishes, fire-retardant foams, etc. [4-6]. Since the use of PFCs is very extensive, exposure of the general U.S. population to PFCs is highly widespread, although variation in demographics, geographic location, and temporal factors may be present [7-9]. National biomonitoring surveys have reported that PFCs are detectable in the blood of over 98% of the US population; some of the most widely-known PFCs, particularly perfluorooctanoic acid (PFOA) and perfluorooctane sulfate (PFOS), have been detected worldwide in wildlife and persist in human bodies [10]. Exposure to PFCs is mostly owing to the ingestion from food or water that are contaminated by PFCs, as well as other paths such as house dust, air, breast milk for infants, and fumes from treated fabrics at home [11-13]. The average daily intake of PFCs was estimated to be up to 130 ng for PFOA, and up to 220 ng for PFOS per kilo bodyweight in the Western hemisphere [11]. Although production of PFOA and PFOS has been phased out in the United States currently, they are still in production in other countries.

PFOA and PFOS are two of the most studied types of perfluoroalkyl substances (PFASs), which belong to the perfluorinated carboxylates and the perfluorinated sulfonates, respectively [2]. Animal studies suggest that PFOS and PFOA are preferentially found in the blood, liver and kidney [2, 14]. Due to the stability of the special chemical structure, C-F bond, these compounds are highly resistant to metabolism and biodegradation, and are thus very persistent in human bodies [15]. It was estimated that the average serum half-life in human bodies is at 5.4 years for PFOS and 3.8 years for PFOA [16]. The long half-lives of PFOA and PFOS in human bodies indicate that these chemicals are bioaccumulative, which could lead to higher body burdens and allow for a multitude of potential adverse health outcomes. So far there have been a multitude of adverse health outcomes observed in animal studies of PFOA and PFOS, including developmental toxicity, hepatotoxicity, immunotoxicity, and tumors in certain organs [2, 17-19]. Several studies in humans suggested that exposure to PFOA and PFOS are associated with increased serum lipid levels such as cholesterol and triglycerides in humans. [20-24]. There is evidence that some adverse outcomes are mediated through the binding between PFASs and peroxisome proliferator-activated receptor alpha (PPARα), which involves in the regulation of metabolism of glucose and lipid in rodents and humans [25]. In laboratory studies, these chemicals have been found to possess strong abilities to disrupt the binding of serum protein ligand [26], and to interact with the nuclear receptor peroxisome PPARα, which plays a central role in adipogenesis and lipid metabolism [27]. Additionally, some toxicological studies of PFOA and PFOS have shown modulation of enlargement of the liver, sex hormone homeostasis, reduced body weight and hypolipidemia in rodent and primate models [2]. These findings have raised

interests in the potential role of PFOA and PFOS in lipid metabolism, including metabolism of cholesterol and triglycerides, in human bodies.

Cholesterol, including high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C), is a principal sterol or modified steroid that widely exists in human bodies [28]. According to the Friedewald equation, total cholesterol can be calculated using LDL, HDL and triglycerides as: [Total cholesterol] = [LDL] + [HDL] + [Triglycerides/5]. All animal cells are able to biosynthesize cholesterol, since it is an indispensable structural component to enable and maintain structural integrity and fluidity of cell membrane. Generally, the hepatic cells synthesize larger amounts of cholesterol than other types of cells. Based on this membrane structure, animal cells are able to change shape and move, which is different from bacteria and plant cells [28]. In addition to this critical roles within cells, cholesterol is also a precursor of the biosynthesis for many important endogenous substances in humans, including bile acids, vitamin D, and steroid hormones [29]. Cholesterol is not important, even completely absent in prokaryotes; but for some exceptions such as Mycoplasma, cholesterol is still essential for its growth [30]. Although cholesterol is broadly existing in animal cells, and plays an important role not only within cells but also in synthesis of other compounds that are essential for humans, high blood cholesterol, called hypercholesterolemia or hyperlipidemia, can lead to serious effects to human health. It has been shown that high serum cholesterol is a strong and independent risk factor for cardiovascular diseases in adults [31]. The presence of hyperlipidemia is directly associated with higher risk of coronary heart disease and potential cardiovascular events in the future [32]. However, less than half of people with increased levels of LDL-C levels have received treatment or

have been adequately treated, which make these patients remain at risk for new cardiovascular events [33].

Scientists have judged that there is a probable link between exposure to PFOA and high cholesterol [34]. Positive statistically significant associations between PFOA and cholesterol levels, as well as between PFOS and cholesterol levels, have been shown in several cross-sectional studies and a few longitudinal studies. There is evidence of both an increased risk of high diagnosed cholesterol and a shift in average cholesterol levels in relation to PFOA and PFOS. For example, a positive association between PFOA and cholesterol has been shown in two community studies [24, 35] and six occupational studies [20, 21, 23, 36-38]. A cross-sectional study has observed a positive association between PFOS and cholesterol levels among workers at one plant, while not at another plant [39]. In addition, several epidemiologic studies have found a significant positive association between total cholesterol and PFOA as well as PFOS in the general U.S. population [22]. Although these studies were conducted on different study populations or sample sizes, the general consensus across them is that there is an association between PFASs and cholesterol levels.

Based on the different situations for the associations between PFASs and lipid levels including cholesterol and triglycerides among different target populations, a variety of studies have been conducted to determine the associations in diverse populations. However, there have been few studies that determined such association in pregnant women, among whom the situation may be highly different from that in other populations. In a large-scale study evaluating the impact of pregnancy on the extent of exposure to PFCs in the U.S., pregnant women are observed to possess lower serum concentrations of

perfluorohexane sulfonic acid (PFHxS), perfluorononanoic acid (PFNA), PFOA, and PFOS than nonpregnant women, while only the differences for PFOS were significant [40]. Another nationwide cross-sectional study based on NHANES (National Health and Nutrition Examination Survey) data reported that many environmental chemicals, including organochlorine pesticides, PFCs, PBDEs, certain polychlorinated biphenyls, phenols, polycyclic aromatic hydrocarbons, perchlorate and phthalates, were detected in 99% to 100% U.S. pregnant women [41]. And generally, the levels of these environmental chemicals in pregnant women were lower than or similar to the levels in nonpregnant women, while the levels tended to be higher than those nonpregnant women if adjusting for physiological factors that could affect the levels of chemicals in pregnant women, namely the large blood volume expansion that occurs during pregnancy [41]. Apart from the difference of PFCs levels between in pregnant women and in nonpregnant women, cholesterol levels are dramatically altered during pregnancy. It has been shown that compared to nonpregnant women, pregnant women tend to have different lipid concentrations, both relative and absolute [42], and this may imply different vulnerabilities to the effects of PFOA, PFOS on cholesterol levels in pregnant women, compared to nonpregnant women [43]. In addition to the associated adverse health effects in the general population, high levels of plasma lipids including cholesterols may raise concerns for outcomes that are specific to pregnant women. Some studies have shown that altered plasma lipids during pregnancy, especially increased levels of triglycerides, are associated with pregnancy-induced hypertension [44] and preeclampsia [45]. Particularly, increases in the levels of some types of LDL cholesterol particles and triglycerides are considered “pro-atherogenic” during pregnancy, which may aggregate

oxidative stress and endothelial damage, and finally lead to preeclampsia in pregnant women [46].

According to the cross-sectional study conducted by Starling et al. (2014) on pregnant Norwegian women, five of the seven PFASs in the study were positively associated with HDL, and PFOS was positively associated with total cholesterol in addition to HDL. One ln-ng/mL increase in PFOS was associated with 8.96 (95% CI: 1.70, 16.22) mg/dL increase in total cholesterol and 4.39 (95% CI: 2.37, 6.42) mg/dL increase in HDL, while was not significantly associated with LDL ($p > 0.05$). The associations between PFOA and all three cholesterols were not significant ($p > 0.05$) [43]. These results suggested potential difference in vulnerabilities to the effects of PFOA, PFOS on cholesterol levels in pregnant women.

Due to the extensive use of PFCs, and their long-lasting persistence in both the environment and human bodies, the various adverse health effects of PFCs are of high concern. Together with the essential roles of cholesterol in human cells and the serious outcomes resulting from high cholesterol levels, any association between PFCs and cholesterol levels have important public health implications. As the levels of both PFCs and cholesterol are different between pregnant women and nonpregnant women, and more importantly, high cholesterol levels may lead to a variety of additional adverse health effects which are specific to pregnant women, we are interested in whether PFCs, especially PFOA and PFOS, are able to affect cholesterol levels during pregnancy. Therefore, we conducted a cross-sectional study using 4 years of NHANES (National Health and Nutrition Examination Survey) data from 2003 to 2006 to examine the association between PFOA and PFOS and cholesterol levels among pregnant women.

METHODS

Data Source and Study Population

The data for this study were extracted and merged from the public database of NHANES. The NHANES is a program of studies designed to measure the nutritional and health conditions of adults and children in the United States, conducted by the Centers for Disease Control and Prevention (CDC) [47]. The survey consists of interviews for various aspects including demographic, socioeconomic, dietary, and health-related questions, as well as physical and laboratory examinations including medical, dental, and physiological measurements [47]. All examinations and tests were administered by highly-trained medical staff. This survey examines a nationally representative sample of approximately 5,000 people each year, who are located in different counties across the country. The sampling of NHANES is representative of the noninstitutionalized, civilian U.S. population, since it is based on a plan of complex, multistage, stratified, probability cluster sampling [47]. Results from the survey can be used to determine the prevalence of diseases and related risk factors. Additionally, information from the survey can be used to evaluate people's nutritional status and its association with disease prevention and health promotion. Pregnant women were oversampled during 2003-2004 and 2005-2006, so we restricted the study period to 2003-2006 to allow for a large enough sample size that was distributed evenly across years of the study period.

Result of pregnancy tests were obtained from NHANES laboratory data. Pregnancy tests of NHANES were conducted on female participants aged 12-59 years and menstruating females aged 8-11 years who were representative of the pregnant women population in U.S. A urine test was completed first, and then a positive result of urine test

was confirmed using a serum test [48]. Only female participants with positive results of both pregnancy tests were included in this study. After filtering, there were in total 250 pregnant women during the 2003-2004 and 2005-2006 NHANES cycles. However, due to the missing data for the main exposures— PFOA and PFOS— and outcomes— cholesterol levels, we ultimately had 174 pregnant women with complete records of PFOA, PFOS, total cholesterol and HDL. Due to the insufficient number of pregnant women who had records for LDL (n = 92), we did not exclude pregnant women with missing data of LDL.

Measurement of Exposure and Outcome

PFOA, PFOS, and cholesterol levels (total cholesterol, HDL and LDL, respectively) were obtained from NHANES laboratory data. Participants of NHANES who were ≥ 12 years old and met the subsample requirements had their serum samples measured for PFCs. The quantitative detection of PFCs was conducted by solid phase extraction-high performance liquid chromatography-turboionspray ionization-tandem mass spectrometry (SPE-HPLC-TCI-MS/MS) [49]. Participants who were ≥ 3 years old and did not meet any of the exclusion criteria had total cholesterol and HDL measured using a plasma sample or serum sample. Total cholesterol was measured enzymatically in plasma or serum by a series of coupled chemical reactions that hydrolyze cholesteryl esters and oxidize the 3-OH group of cholesterol [50]. HDL was measured directly in serum [51]. LDL was calculated from measured values of total cholesterol, HDL and triglycerides, based on the Friedewald calculation: [LDL] = [total cholesterol] – [HDL] – [triglycerides/5], where triglycerides were measured enzymatically in serum by a series

of coupled chemical reactions [52]. All measurements for cholesterols were performed by Johns Hopkins University Lipoprotein Analytical Laboratory.

Covariates

Covariates for the study were chosen based on existing studies, which were reported to be potential confounders of the association between PFCs and serum lipids among U. S. adolescents, the U.S. general population and the middle-aged Danish population [22, 53, 54]. These covariates including age (years), race/ethnicity, education level, annual family income, body mass index (BMI), physical activity, daily fat intake (g/day), and serum cotinine levels (ng/mL) were obtained from demographics data, dietary data, examination data, laboratory data and questionnaire data of NHANES. Cotinine was used as a biomarker for exposure to tobacco smoke, as it is the predominant metabolite of nicotine [55]. Additionally, since the study population is pregnant women, we also included unique characteristics relevant to the study population—history of breastfeeding, month of pregnancy and number of former pregnancies— based on the study for the associations between PFCs and plasma lipids among pregnant Norwegian women [43]. Characteristics of pregnant women were extracted from NHANES questionnaire data of reproductive health. Due to the large number of missing observations for breastfeeding (n = 77), we restricted the pregnant characteristics to month of pregnancy and number of former pregnancies to ensure a sufficient sample size (N = 174).

Among all the covariates, race, education, annual family income and physical activity were in categorical scale, while other variables were continuous. Due to the large number of levels ($k \geq 5$) for these categorical variables relative to the small sample size (N = 174), we re-categorized the levels of the categorical variables instead of using the original

categorization of NHANES to avoid problems of overparameterization. Re-categorization was based on the distribution of the variable, categorization in existing studies, as well as the univariate association between outcome and each level of the variable. Possible problems of sparse data were examined using contingency tables. Bar plot and frequency tables were used to examine the distribution of participants in each level of the variable. After re-categorization, participant's race was classified as four levels: (1) Mexican American, (2) Non-Hispanic black, (3) Non-Hispanic white and (4) other races (including other Hispanic and multi-racial); education was classified as three levels: (1) less than high school, (2) high school diploma or GED and (3) more than high school; annual family income was classified as three levels: (1) under $20,000, (2) $20,000 to $44,999 and (3) $45,000 and over; physical activity was dichotomized as having moderate or vigorous activity over past 30 days or not.

Statistical Analysis

We performed linear regression to quantify the associations between cholesterol levels and PFCs. Assumptions for normality and homoscedasticity were primarily examined using histogram, normal probability plot and scattering plot. F tests were performed to examine the differences in cholesterols among levels of covariates, including race, education level, and annual family income. Linear models were fitted for each independent variable separately associated with total cholesterol, HDL and LDL, respectively. Through the simple linear models, we estimated unadjusted associations and 95% confidence intervals for each pair of independent variable and dependent variable. PFOA and PFOS were analyzed in log-transformed scales and quartiles due to skewness. Each continuous independent variable which showed potential non-linear association

with cholesterol was performed test of linear trend using deviance. Deviance is calculated as $D = -2[\log L_c - \log L_s]$, where L_c is the maximum likelihood for the current model, and L_s is the maximum likelihood for the saturated model. We fitted a model with the continuous variable and obtained the deviance of the model, and then categorized the variable with m levels and fitted a model with the categorical variable. Categorization for these variables were based on existing studies, distribution of variables, and trend of relationship suggested in scattering plots. We obtained the deviance for the second model and calculated the difference in deviance between these two model as $\Delta D = -2[\log L_{cont} - \log L_{cat}]$, where L_{cont} was the maximum likelihood for the model with the continuous variable, and L_{cat} was the maximum likelihood for the model with the categorical variable. Under null hypothesis, ΔD had a chi-squared distribution with (m-2) degrees of freedom; if $\Delta D > \chi^2_{\alpha,(m-2)}$, we would conclude that the relationship between the outcome and exposure is significantly non-linear. According to test of linear trend, serum cotinine, BMI and number of former pregnancies were defined categorically due to significant non-linearity ($p < 0.05$). Serum cotinine was dichotomized as > 0.1 ng/mL and ≤ 0.1 ng/mL; BMI was classified as normal (< 25 kg/m^2), overweight (25-29.99 kg/m^2), and obese and above (≥ 30 kg/m^2); number of former pregnancies was classified as 0, 1, 2 and ≥ 3.

Four multiple linear models were fitted to determine the adjusted associations between PFOA, PFOS and total cholesterol and HDL, respectively. Since the number of covariates (n = 10) is relatively large for the sample size of 174 participants, we used forward method to assess confounding while avoiding model overfitting. Detailed steps are: 1) Fit a univariate model for the main exposure and outcome, and add each covariate

separately into the model. Calculate the change in the coefficient estimate for the main exposure from the univariate model. 2) If the change is greater than 10%, identify the first confounder that maximizes the change in estimate for the main exposure. Add this confounder to the univariate model as the Model 1 for the next step. If the change is less than 10%, stop model selection and fit the final model as the univariate model. 3) Among the rest covariates, add each covariate one at a time to Model 1. Calculate the change in the coefficient estimate for the main exposure from the Model 1, and determine if the change is greater than 10%. 4) Similarly, identify the rest confounders one at a time until the change in estimate for main exposure is less than 10%, or the model is unstable due to potential problems of collinearity (suggested by the variance inflation of some coefficient estimates).

After the four models were fitted, we compared different models for each exposure. If the models for the same exposure did not control for same covariates, we added additional covariates tothe models to make the covariates that were controlled consistent across models. This was to simplify the interpretations for the model of each exposure, as well as gaining better precision for the coefficient estimates. After the final models were fitted, we assessed assumptions of linear regression, influential observations, outliers and collinearity for each model. Assumptions for linearity, normality and homoscedasticity were examined using residual plots and partial plots. Influential observations and outliers were detected using leverage and Cook's distance. Problem of collinearity was detected using Variance Inflation Factor (VIF). Variables with collinearity would be removed from the model. Due to the small sample size for participants with LDL data (N = 92), adjusted model for LDL was not fitted independently. Instead, we applied the final

models for total cholesterol and HDL to LDL and then performed diagnostics to the models.

In this study, all the data cleaning, descriptives and analyses were performed using SAS 9.4 and SAS University Edition. Significance level was set as 5% for all hypothesis tests.

RESULTS

Among 174 pregnant women in the study, average age was 26.3 (SD: 5.4) years. 45 (25.9%) participants were Mexican American, and 90 (51.7%) participants were non-Hispanic white. 90 (51.7%) participants had education of more than high school, and 63 (37.1%) participants had an annual family income of $45,000 and above. Average total fat intake was 83.7 (SD: 34.3) grams per day. BMI categorization of 64 (37.0%) participants were obese and above. 77 (44.3%) of the study sample reported no moderate or vigorous activity in the past 30 days. Serum cotinine levels of 52 (29.9%) participants were higher than 0.1 ng/mL, which was reported to be the average serum cotinine level among adult non-smokers in the United States [56]. 55 (36.7%) of the participants had been pregnant for 4 to 6 months, and 68 (45.3%) participants had been pregnant for more than 7 months. 52 (31.7%) women were pregnant for the first time [Table 1].

PFOA and PFOS were measured in the serum samples of 174 pregnant women in the study. The distributions of PFOA and PFOS were right-skewed. Median concentration of PFOA was 2.3 ng/mL, which was lower than median concentration of PFOS (10.1 ng/mL). Cholesterol levels, including total cholesterol, HDL and LDL, were normally distributed. Total cholesterol and HDL cholesterol were measured in the serum samples of all 174 pregnant women in the study, while LDL values were available for only 92 pregnant women in the study. Mean concentrations for total cholesterol, HDL and LDL were 223.0 (SD: 48.9) mg/dL, 69.9 (SD: 15.9) mg/dL and 115.9 (SD: 38.0) mg/dL, respectively [Table 1].

PFOA and PFOS were analyzed as continuous log-transformed variables and quartiles in linear regression due to skewness [Table 2, Table 3]. Unadjusted associations between

PFOA, PFOS and cholesterol levels were not significant ($p > 0.05$). Total cholesterol was significantly associated with annual family income ($p = 0.01$), race ($p = 0.02$) and month of pregnancy ($p < 0.01$). HDL was significantly associated with age ($p = 0.02$), education level ($p = 0.04$), serum cotinine ($p = 0.04$) and month of pregnancy ($p = 0.01$). LDL was significantly associated with month of pregnancy ($p < 0.01$). Compared to participants whose annual family income was under $20,000, mean total cholesterol of those having annual family income of $20,000 to $44,999 was significantly higher with a difference of 25.5 (95% CI: 7.1, 43.9) mg/dL, and mean total cholesterol of those having annual family income of $45,000 and above was significantly higher with a difference of 21.4 (95% CI: 3.7, 39.1) mg/dL. Compared to participants who did not obtain high school education, mean total cholesterol of those having education of more than high school was significantly higher with a difference of 18.5 (95% CI: 1.1, 35.8) mg/dL. A concentration of serum cotinine higher than 0.1 ng/mL was significantly associated with 5.5 (95% CI: 0.3, 10.6) mg/dL decrease in mean HDL, compared to a concentration of serum cotinine not exceeding 0.1 ng/mL. A 1-year increase in age was significantly associated with 0.5 (95% CI: 0.1, 1.0) mg/dL increase in HDL. A 1-month increase in gestation length was significantly associated with an 11.8 (95% CI: 8.9, 14.7) mg/dL increase in mean total cholesterol, 1.5 (95% CI: 0.4, 2.6) mg/dL increase in mean HDL, and 5.5 (95% CI: 2.0, 9.1) mg/dL increase in mean LDL [Table 2].

According to simple linear regression, PFOA was significantly associated with annual family income ($p = 0.02$), race ($p = 0.03$), and number of former pregnancies ($p < 0.01$). PFOS was significantly associated with annual family income ($p < 0.01$), race ($p < 0.01$), serum cotinine ($p = 0.03$), and number of former pregnancies ($p < 0.01$). Compared to

participants whose annual family income was under $20,000, those having annual family income of $45,000 and above have significantly higher mean log-PFOA with a difference of 0.3 (95% CI: 0.0, 0.6) ln-ng/mL, and similar significantly higher mean log-PFOS. Compared to non-Hispanic whites, Mexican Americans have significantly lower mean log-PFOA with a difference of 0.4 (95% CI: 0.1, 0.7) ln-ng/mL, and similar significantly lower mean log-PFOS. Compared to participants pregnant for the first time, those pregnant for the second time have significantly lower mean log-PFOA with a difference of 0.6 (95% CI: 0.3, 0.9) ln-ng/mL, and similar significantly lower mean log-PFOS; those pregnant for the third time have significantly lower mean log-PFOA with a difference of 0.6 (95% CI: 0.2, 0.9) ln-ng/mL, and significantly lower mean log-PFOS with a difference of 0.3 (95% CI: 0.1, 0.6) ln-ng/mL; those pregnant for more than three times have significantly lower mean log-PFOA with a difference of 0.4 (95% CI: 0.1, 0.7) ln-ng/mL, and significantly lower mean log-PFOS with a difference of 0.3 (95% CI: 0.0, 0.6) mg/dL. A concentration of serum cotinine higher than 0.1 ng/mL was significantly associated with 0.2 (95% CI: 0.0, 0.5) ln-ng/mL decrease in mean log-PFOS, compared to a concentration of serum cotinine not exceeding 0.1 ng/mL.

PFOA and cholesterol levels

Adjusting for age (years), race, education level, annual family income, BMI, total fat intake (g/day), serum cotinine, month of pregnancy and number of former pregnancies as confounders, one ln-ng/mL increase in PFOA was associated with 2.0 (95% CI: -8.9, 13.0) mg/dL decrease in total cholesterol, 1.3 (95% CI: -2.9, 5.5) decrease in HDL, and 1.4 (95% CI: -15.6, 18.4) increase in LDL. Associations between log-PFOA and the three types of cholesterols were not significant ($p > 0.05$) and confidence intervals were wide. Based on

the adjusted model, a 1-month increase in gestational length was significantly associated with 11.7 (95% CI: 8.2, 15.2) mg/dL increase in mean total cholesterol, and 6.4 (95% CI: 1.4, 11.4) mg/dL increase in mean LDL. Compared to non-Hispanic whites, mean total cholesterol of non-Hispanic blacks was significantly lower with a difference of 27.6 (95% CI: 2.9, 52.3) mg/dL [Table 3-1]. According to residual plots and partial plots, assumptions for linearity, normality, or homoscedasticity were not violated. There was no collinearity between independent variables (VIF < 5). Cook's distance and leverage detected one influential observation for each model. But since none of the values were implausible, we kept all observations in the model.

PFOS and cholesterol levels

Adjusting for age (years), race, education level, annual family income, BMI, total fat intake (g/day), serum cotinine, month of pregnancy and number of former pregnancies as confounders, one ln-ng/mL increase in PFOS was associated with 1.3 (95% CI: -11.4, 13.9) mg/dL increase in total cholesterol, 0.4 (95% CI: -4.5, 5.3) mg/dL increase in HDL, and 21.0 (95% CI: 0.6, 41.4) mg/dL increase in LDL. Only the association between PFOS and LDL was significant ($p < 0.05$). Compared to participants pregnant for the first time, mean total cholesterol of those pregnant for the second time was significantly higher with a difference of 20.9 (1.3, 40.5) mg/dL, mean LDL of those pregnant for the second time was significantly higher with a difference of 29.1 (5.1, 53.0) mg/dL, and mean LDL of those pregnant for the third time was significantly higher with a difference of 30.3 (0.9, 59.7) mg/dL [Table 3-2]. Based on residual plots and partial plots, there was no violation of assumptions for linearity, normality, or homoscedasticity. No collinearity was detected between independent variables (VIF < 5). There was one influential

observation detected for the model of total cholesterol model and HDL, using Cook's distance and leverage. But we kept all observations in the model since none of the values were implausible.

In addition to log-transformed scales of PFOA and PFOS, we also fitted multiple linear models to determine the adjusted associations between PFOA quartiles, PFOS quartiles and cholesterol levels, adjusting for the same set of confounders [Table 3-3]. Quartiles of PFOA and PFOS were not significantly associated with total cholesterol, HDL or LDL [Table 3-3], which was similar to the results of the log-PFCs models, except for the associations between log-PFOS and LDL which was significantly positively associated in the log-PFOS model [Table 3-2]. The point estimates for PFOS quartiles indicated higher LDL with higher PFOS, but all confidence intervals included the null value.

DISCUSSION

In this study, we identified the associations between serum concentrations of PFOA, PFOS and cholesterol levels in a representative sample of U.S. pregnant females. A significant positive association between log-PFOS and LDL was found, while the associations between PFOA and cholesterols as well as between PFOS and non-LDL cholesterols were not significant. Associations between PFOA, PFOS and cholesterol levels were found to be confounded by age (years), race, education level, annual family income, BMI, total fat intake (g/day), serum cotinine, month of pregnancy and number of former pregnancies. Among these confounders, month of pregnancy was the most significant independent variable across all adjusted models, which suggested a strong association between gestational length and change in cholesterols.

There have been few studies that explored the associations between PFCs and cholesterol levels in pregnant women, among which the cross-sectional study conducted by Starling et al. (2014) on pregnant Norwegian women was the most relevant to our study [43]. The results for PFOA of this study were consistent with those found by Starling et al., in which the associations between PFOA and all three cholesterols were not significant, using either log-PFOA or PFOA quartiles [43]. However, the results for log-PFOS of this study was not consistent with those found by Starling et al., in which the association between log-PFOS and non-LDL cholesterols was significant positive while the association between log-PFOS and LDL was not significant [43]. Associations between PFOS quartiles and cholesterols were basically consistent with those found by Starling et al., except for the association between PFOS quartiles and HDL which was significant positive [43]. Another related study by Skuladottir et al. (2015) examined the

confounding by diet in the association between concentrations of PFCs and total cholesterol in pregnant Danish women [57]. The study of Skuladottir et al. showed positive and similar associations between PFOA quartiles, PFOS quartiles and serum total cholesterols [57], which was not consistent with the results by either Starling et al. or our study.

In contrast to the studies in nonpregnant populations that generally found consistent results of significant positive associations between PFOA, PFOS and cholesterol levels [20-24, 35-39], the associations between PFOA, PFOS and cholesterols appears inconsistent across studies regarding pregnant women. This inconsistency may be due to: (1) the variation in pregnant stage of the study sample: the strong association between gestational length and cholesterol levels may imply different vulnerabilities to the effects of PFCs on cholesterols in pregnant women during different stages of pregnancy – consequently, different distributions of pregnant stages may lead to different results of the PFCs-cholesterol associations; (2) different study population –for example, the studies of Starling et al. and Skuladottir et al. was based on pregnant Norwegian women and pregnant Danish women, respectively [43, 57], while our study was based on pregnant U.S. women; the genetic factors of different races and the environmental factors of different countries may also lead to different vulnerabilities to the effects of PFCs on cholesterols, and thus lead to different results; (3) different covariates controlled in the analyses –for example, the study of Starling et al. controlled for demographics, smoking, dietary, pre-pregnant BMI, and various reproductive health characteristics including nulliparous or inter-pregnancy interval, gestational weeks, and duration of breastfeeding previous child [43]; the study of Skuladottir et al. controlled for demographics, pre-

pregnancy BMI, smoking, and detailed dietary factors including vegetables, meat and meat products, and total caloric intake [57]; our study controlled for demographics, BMI, dietary, serum cotinine, and two reproductive health characteristics including month of pregnancy and number of former pregnancies. Controlling for different set of confounders or insufficient controlling for strong confounders may also lead to different results of the associations; (4) insufficient statistical power: it is possible that PFCs truly affect cholesterol levels among pregnant women, but the number of pregnant women available for our study from NHANES was too small to detect these associations.

We observed little evidence that PFOA and PFOS affect cholesterol levels during pregnancy, with the possible exception of the effect of PFOS on LDL. However, since PFCs could allow for various potential adverse health effects including developmental toxicity, hepatotoxicity, immunotoxicity and tumors [2, 17-19] as well as pregnancy-induced hypertension [44] and low birth weight [58], controlling the exposure to PFOA and PFOS is still highly necessary for pregnant women.

Strengths and Weaknesses

The data for this study were extracted from NHANES, the sampling of which is based on a complex, multistage, stratified, probability cluster sampling plan [47]. Since the exposures and outcomes were measured using strict laboratory approaches in professional and authoritative institutions such as CDC and Johns Hopkins University [48], values of exposures and outcomes were guaranteed to be valid and accurate, and this decreased the possibilities of information bias. In the analyses, we used statistical tests to examine the linearity of the relationship between cholesterols and each independent variable, and ultimately modeled PFCs in both log-transformed scales and quartiles as well as

categorized several continuous covariates in the model to optimize the fitness of the non-linear relationships. Additionally, we controlled for different kinds of covariates including demographics, dietary, behaviors and reproductive health factors in the multiple regression models, which allowed for a comprehensive adjustment for the associations. Furthermore, since there have been few studies that explored the associations between PFCs and cholesterol levels in pregnant women, this study could provide important results to fill in the gap.

There were also some limitations for this study. The initial sample size was small (N = 174), and the actual sample size (n = 134) used in the final models –except for the LDL models which used only 66 observations due to large number of missing data of LDL – was smaller because of missing observations of covariates. The small sample size may make the results more susceptible to potential outliers and influential observations. Also, to avoid the problems of overparameterization for the small sample size as well as to simplify the interpretations, the study did not consider interactions, which may be significant if included in the models. And since the confounding effects in the missing observations were unclear, there could be potential bias in the study results. Moreover, the association between PFCs and cholesterols could be confounded by additional important characteristics such as pre-pregnancy BMI and history of breastfeeding. However, we did not obtain data for these variables due to the limited data sources, which might make our models not fully adjusted.

Future Directions

Give the widespread exposure to PFCs and the importance of cholesterols on pregnant women's health, more large-scale studies need to be conducted to draw the attentions of

scientists and public health professionals to this area. Future studies are expected to be conducted in larger and more general study populations of pregnant women, but not limited to a particular area or country. Also, more possible confounders (especially reproductive health characteristics) are expected to be considered in future studies to allow for more accurate estimates of the associations between cholesterols and PFCs.

REFERENCES

1. Olsen, G.W., et al., *Historical Comparison of Perfluorooctanesulfonate, Perfluorooctanoate, and Other Fluorochemicals in Human Blood.* Environmental Health Perspectives, 2005. **113**(5): p. 539-545.
2. Lau, C., et al., *Perfluoroalkyl acids: a review of monitoring and toxicological findings.* Toxicol Sci, 2007. **99**(2): p. 366-94.
3. Conder, J.M., et al., *Are PFCAs bioaccumulative- A critical review and comparison with regulatory criteria and persistent lipophilic compounds.* Environ Sci Technol, 2008. **42**(4): p. 995–1003.
4. Olsen, G.W., et al., *Decline in Perfluorooctanesulfonate and Other Polyfluoroalkyl Chemicals in American Red Cross Adult Blood Donors, 2000-2006.* Environ. Sci. Technol., 2008. **42**: p. 4989–4995.
5. Steenland, K., T. Fletcher, and D.A. Savitz, *Epidemiologic evidence on the health effects of perfluorooctanoic acid (PFOA).* Environ Health Perspect, 2010. **118**(8): p. 1100-8.
6. Raymer, J.H., et al., *Concentrations of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) and their associations with human semen quality measurements.* Reprod Toxicol, 2012. **33**(4): p. 419-27.
7. Fromme, H., et al., *Occurrence of perfluorinated substances in an adult German population in southern Bavaria.* Int Arch Occup Environ Health, 2007. **80**: p. 313–319.
8. Harada, K., et al., *Historical and geographical aspects of the increasing perfluorooctanoate and perfluorooctane sulfonate contamination in human serum in Japan.* Chemosphere, 2007. **66**(2): p. 293-301.
9. Karrman, A., et al., *Levels of 12 perfluorinated chemicals in pooled Australian serum, collected 2002–2003, in relation to age, gender, and region.* Environ Sci Technol, 2006. **40**: p. 3742–3748.

10. Calafat, A.M., et al., *Polyfluoroalkyl chemicals in the U.S. population: data from the National Health and Nutrition Examination Survey (NHANES) 2003-2004 and comparisons with NHANES 1999-2000.* Environ Health Perspect, 2007. **115**(11): p. 1596-602.

11. Trudel, D., et al., *Estimating consumer exposure to PFOS and PFOA.* Risk Anal, 2008. **28**(2): p. 251-69.

12. Fromme, H., et al., *Pre- and postnatal exposure to perfluorinated compounds (PFCs).* Environ Sci Technol, 2010. **44**: p. 7123–9.

13. Haug, L.S., et al., *Characterisation of human exposure pathways to perfluorinated compounds--comparing exposure estimates with biomarkers of exposure.* Environ Int, 2011. **37**(4): p. 687-93.

14. Kennedy, G.L., et al., *The Toxicology of Perfluorooctanoate.* Critical Reviews in Toxicology, 2004. **34**(4): p. 351-384.

15. Lau, C., *Perfluoroalkyl acids: Recent activities and research progress.* Reprod Toxicol, 2009. **27**: p. 209-211.

16. Olsen, G.W., et al., *Half-life of serum elimination of perfluorooctanesulfonate,perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers.* Environ Health Perspect, 2007. **115**(9): p. 1298-305.

17. Biegel, L.B., et al., *Mechanisms of extrahepatic tumor induction by peroxisome proliferators in male CD rats.* Toxicol Sci, 2001. **60**: p. 44-55.

18. White, S.S., et al., *Gestational PFOA exposure of mice is associated with altered mammary gland development in dams and female offspring.* Toxicol Sci, 2007. **96**(1): p. 133-44.

19. DeWitt, J.C., et al., *Immunotoxicity of perfluorooctanoic acid and perfluorooctane sulfonate and the role of peroxisome proliferator-activated receptor alpha.* Crit Rev Toxicol, 2009. **39**(1): p. 76-94.

20. Costa, G., S. Sartori, and D. Consonni, *Thirty years of medical surveillance in perfluorooctanoic acid production workers.* J Occup Environ Med, 2009. **51**(3): p. 364-72.

21. Sakr, C.J., et al., *Cross-sectional study of lipids and liver enzymes related to a serum biomarker of exposure (ammonium perfluorooctanoate or APFO) as part of a general health survey in a cohort of occupationally exposed workers.* J Occup Environ Med, 2007. **49**(10): p. 1086-96.

22. Nelson, J.W., E.E. Hatch, and T.F. Webster, *Exposure to polyfluoroalkyl chemicals and cholesterol, body weight, and insulin resistance in the general U.S. population.* Environ Health Perspect, 2010. **118**(2): p. 197-202.

23. Olsen, G.W., et al., *Epidemiologic Assessment of Worker Serum Perfluorooctanesulfonate (PFOS) and Perfluorooctanoate (PFOA) Concentrations and Medical Surveillance Examinations.* Journal of Occupational and Environmental Medicine, 2003. **45**(3): p. 260-270.

24. Steenland, K., et al., *Association of perfluorooctanoic acid and perfluorooctane sulfonate with serum lipids among adults living near a chemical plant.* Am J Epidemiol, 2009. **170**(10): p. 1268-78.

25. Abbott, B.D., et al., *Perfluorooctanoic acid induced developmental toxicity in the mouse is dependent on expression of peroxisome proliferator activated receptor-alpha.* Toxicol Sci, 2007. **98**(2): p. 571-81.

26. Luebker, D.J., et al., *Interactions of fluorochemicals with rat liver fatty acid-binding protein.* Toxicology, 2002. **176**(3): p. 175–85.

27. Rigamonti, E., G. Chinetti-Gbaguidi, and B. Staels, *Regulation of macrophage functions by PPAR-alpha, PPAR-gamma, and LXRs in mice and men.* Arterioscler Thromb Vasc Biol, 2008. **28**(6): p. 1050-9.

28. Haberland, E.M. and A.J. Reynolds, *Self-association of Cholesterol in Aqueous Solution.* Proc Nat Acad Sci, 1973. **70**(8): p. 2313-2316.

29. Hanukoglu, I., *Steroidogenic enzymes: structure, function, and role in regulation of steroid hormone biosynthesis.* J Steroid Biochem Mol Biol, 1992. **43**(8): p. 779–804.

30. Razin, S. and J.G. Tully, *Cholesterol Requirement of Mycoplasmas.* Journal of Bacteriology, 1970. **102**(2): p. 306–310.

31. Gordon, D.J., et al., *High-Density Lipoprotein Cholesterol and Cardiovascular Disease. Four Prospective American Studies.* Circulation, 1989. **79**(1): p. 8-15.

32. Willerson, J.T. and P.M. Ridker, *Inflammation as a cardiovascular risk factor.* Circulation, 2004. **109**(21 Suppl 1): p. II2-10.

33. Go, A.S., et al., *Heart disease and stroke statistics--2014 update: a report from the American Heart Association.* Circulation, 2014. **129**(3): p. e28-e292.

34. C8 Science Panel. *Probable Link Evaluation for heart disease (including high blood pressure, high cholesterol, coronary artery disease)*. Oct 29, 2012 Mar 2, 2016]; Available from: http://www.c8sciencepanel.org/pdfs/Probable_Link_C8_Heart_Disease_29Oct2012.pdf.

35. Emmett, E.A., et al., *Community exposure to perfluorooctanoate: relationships between serum levels and certain health parameters.* J Occup Environ Med, 2006. **48**(8): p. 771-9.

36. Olsen, G.W. and L.R. Zobel, *Assessment of lipid, hepatic, and thyroid parameters with serum perfluorooctanoate (PFOA) concentrations in fluorochemical production workers.* Int Arch Occup Environ Health, 2007. **81**(2): p. 231-46.

37. Olsen, G.W., et al., *Plasma cholecystokinin and hepatic enzymes, cholesterol and lipoproteins in ammonium perfluorooctanoate production workers.* Drug Chem Toxicol, 2000. **23**(4): p. 603-20.

38. Sakr, C.J., et al., *Longitudinal study of serum lipids and liver enzymes in workers with occupational exposure to ammonium perfluorooctanoate.* J Occup Environ Med, 2007. **49**(8): p. 872-9.

39. Olsen, G.W., et al., *Serum perfluorooctane sulfonate and hepatic and lipid clinical chemistry tests in fluorochemical production employees.* J Occup Environ Med, 1999. **41**(9): p. 799–806.

40. Jain, R.B., *Effect of pregnancy on the levels of selected perfluoroalkyl compounds for females aged 17-39 years: data from National Health and Nutrition Examination Survey 2003-2008.* J Toxicol Environ Health A, 2013. **76**(7): p. 409-21.

41. Woodruff, T.J., A.R. Zota, and J.M. Schwartz, *Environmental chemicals in pregnant women in the United States: NHANES 2003-2004.* Environ Health Perspect, 2011. **119**(6): p. 878-85.

42. Piechota, W. and A. Staszewski, *Reference ranges of lipids and apolipoproteins in pregnancy.* Eur J Obstet Gynecol Reprod Biol, 1992. **45**: p. 27–35.

43. Starling, A.P., et al., *Perfluoroalkyl substances and lipid concentrations in plasma during pregnancy among women in the Norwegian Mother and Child Cohort Study.* Environ Int, 2014. **62**: p. 104-12.

44. Vrijkotte, T.G., et al., *Maternal lipid profile during early pregnancy and pregnancy complications and outcomes: the ABCD study*. J Clin Endocrinol Metab, 2012. **97**(11): p. 3917-25.

45. Sattar, N., et al., *Lipoprotein subfraction concentrations in preeclampsia: pathogenic parallels to atherosclerosis.* Obstet Gynecol, 1997. **89**: p. 403–8.

46. Llurba, E., et al., *Atherogenic lipoprotein subfraction profile in preeclamptic women with and without high triglycerides: different pathophysiologic subsets in preeclampsia.* Metabolism, 2005. **54**(11): p. 1504-9.

47. National Center for Health Statistics. *Introduction to the National Health and Nutrition Examination Survey*. Feb 3, 2014; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/2013-14_overview_brochure.pdf.

48. National Health and Nutrition Examination Survey. *Laboratory Procedure Manual*. Jan 2005; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/lab.pdf.

49. CDC Division of Laboratory Sciences. *Laboratory Protocol (Polyfluorochemicals Compounds)*. Aug 2009; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/pfc_d_met.pdf.

50. Johns Hopkins University Lipoprotein Analytical Laboratory. *Laboratory Procedure Manual (Total Cholesterol)*. Nov 2007; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/tchol_d_met_h912.pdf.

51. Johns Hopkins University Lipoprotein Analytical Laboratory. *Laboratory Procedure Manual (HDL-Cholesterol)*. Dec 2007; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/hdl_d_met_cholesterol_hdl_h912.pdf.

52. Johns Hopkins University Lipoprotein Analytical Laboratory. *Laboratory Procedure Manual (Triglycerides)*. Mar 2008; Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/trigly_d_met_triglyceride_h717.pdf.

53. Geiger, S.D., et al., *The association between PFOA, PFOS and serum lipid levels in adolescents.* Chemosphere, 2014. **98**: p. 78-83.

54. Eriksen, K.T., et al., *Association between plasma PFOA and PFOS levels and total cholesterol in a middle-aged Danish population.* PLoS One, 2013. **8**(2): p. e56969.

55. P.;, D.L., et al., *(S)-(-)-Cotinine, the major brain metabolite of nicotine, stimulates nicotinic receptors to evoke [3H]dopamine release from rat striatal slices in a calcium-dependent manner*. J Pharmacol Exp Ther, 1999. **288**(3): p. 905-11.

56. Benowitz, N.L., et al., *Optimal serum cotinine levels for distinguishing cigarette smokers and nonsmokers within different racial/ethnic groups in the United States between 1999 and 2004.* Am J Epidemiol, 2009. **169**(2): p. 236-48.

57. Skuladottir, M., et al., *Examining confounding by diet in the association between perfluoroalkyl acids and serum cholesterol in pregnancy.* Environ Res, 2015. **143**(Pt A): p. 33-8.

58. Darrow, L.A., C.R. Stein, and K. Steenland, *Serum perfluorooctanoic acid and perfluorooctane sulfonate concentrations in relation to birth outcomes in the Mid-Ohio Valley, 2005-2010.* Environ Health Perspect, 2013. **121**(10): p. 1207-13.

Table 1. Characteristics[a] of a cohort of U.S. pregnant women, based on NHANES data from 2003 to 2006 (N=174).

Characteristics	N (%) or Mean ± SD
Age (years)	26.3 ± 5.4
Race	
Mexican American	45 (25.9)
Non-Hispanic White	90 (51.7)
Non-Hispanic Black	27 (15.5)
Other	12 (6.9)
Education	
Less than high school	46 (26.4)
High school diploma or GED	38 (21.8)
More than high school	90 (51.7)
Annual family income	
Under $20,000	54 (31.8)
$20,000 to $44,999	53 (31.2)
$45,000 and over	63 (37.1)
Body Mass Index (kg/m^2)[b]	
Normal	49 (28.3)
Overweight	60 (34.7)
Obese and above	64 (37.0)
Total fat intake (g/day)	83.7 ± 34.3
No moderate or vigorous activity	77 (44.3)
Serum cotinine > 0.1 ng/mL	52 (29.9)
Month of pregnancy	
1-3	27 (18.0)
4-6	55 (36.7)
≥ 7	68 (45.3)
Previous pregnancy	
0	52 (31.7)
1	47 (28.7)
2	31 (18.9)
≥ 3	34 (20.7)
Perfluorooctanoic acid (ng/mL)[c]	2.3 (1.2, 3.2)
Perfluorooctane sulfonic acid (ng/mL)[c]	10.1 (6.1, 14.7)
Serum Cholesterol levels	
Total cholesterol (mg/dL)	223.0 ± 48.9
HDL-cholesterol (mg/dL)	69.9 ± 15.9
LDL-cholesterol (mg/dL)	115.9 ± 38.0

Abbreviations: N, number; SD, standard deviation.

[a] Number of missing observations: 4 for annual family income, 1 for BMI (kg/m^2), 17 for total fat intake (g/day), 24 for month of pregnancy, 10 for previous pregnancy, and 82 for LDL.

[b] Normal: < 25 kg/m^2; overweight: 25-29.99 kg/m^2; obese and above: ≥ 30 kg/m^2.

[c] Report median (lower quartile, upper quartile) due to skewed distribution.

Table 2. Unadjusted associations between PFOA[a], PFOS[b], participants characteristics[c] and serum cholesterol levels (mg/dL) of a cohort of U.S. pregnant women, based on NHANES data from 2003 to 2006 (N=174).

Characteristics	Total cholesterol (mg/dL)		HDL cholesterol (mg/dL)		LDL cholesterol (mg/dL)	
	β estimate	95% CI	β estimate	95% CI	β estimate	95% CI
PFOA (ng/mL)						
Quartile 1		*Referent*		*Referent*		*Referent*
Quartile 2	-14.1	(-34.1, 6.0)	-6.1	(-12.6, 0.4)	-6.3	(-27.3, 14.7)
Quartile 3	-3.8	(-24.4, 16.8)	-5.1	(-11.8, 1.6)	13.5	(-8.9, 36.0)
Quartile 4	-21.7	(-42.0, -1.3)	-6.6	(-13.2, 0.1)	4.6	(-17.2, 26.4)
Per ln-ng/mL	-8.1	(-17.5, 1.3)	-2.7	(-5.8, 0.3)	1.3	(-8.5, 11.2)
PFOS (ng/mL)						
Quartile 1		*Referent*		*Referent*		*Referent*
Quartile 2	2.1	(-18.7, 22.8)	1.7	(-5.1, 8.5)	-3.5	(-26.3, 19.3)
Quartile 3	-6.8	(-27.3, 13.7)	-1.4	(-8.1, 5.3)	1.5	(-19.6, 22.6)
Quartile 4	-14.3	(-35.1, 6.5)	-0.9	(-7.7, 5.9)	1.4	(-21.1, 23.9)
Per IQR	-5.1	(-11.7, 1.4)	-0.6	(-2.7, 1.6)	0.8	(-6.2, 7.8)
Per ln-ng/mL	-8.7	(-19.9, 2.4)	-1.7	(-5.4, 1.9)	4.3	(-8.0, 16.6)
Age (years)	0.8	(-0.5, 2.2)	0.5	(0.1, 1.0)	-0.2	(-1.7, 1.3)
Race						
Non-Hispanic White		*Referent*		*Referent*		*Referent*
Mexican American	-2.9	(-20.2, 14.4)	2.4	(-3.4, 8.2)	-10.3	(-29.8, 9.1)
Non-Hispanic Black	-33.1	(-53.8, 12.3)	0.0	(-6.9, 6.9)	-26.7	(-48.7, -4.7)
Other	-9.3	(-38.4, 19.7)	2.0	(-7.7, 11.7)	-11.9	(-40.2, 16.5)
Education						
Less than high school		*Referent*		*Referent*		*Referent*
High school diploma / GED	6.5	(-14.5, 27.5)	-2.2	(-8.9, 4.7)	6.8	(-17.5, 31.1)
More than high school	18.5	(1.1, 35.8)	4.9	(-0.7, 10.5)	8.8	(-9.2, 26.7)
Annual family income						
Under $20,000		*Referent*		*Referent*		*Referent*
$20,000 to $44,999	25.5	(7.1, 43.9)	-2.3	(-8.4, 3.8)	10.2	(-10.9, 31.3)
$45,000 and over	21.4	(3.7, 39.1)	0.5	(-5.4, 6.3)	13.6	(-5.8, 33.0)
Body Mass Index (kg/m^2)[d]						
Normal		*Referent*		*Referent*		*Referent*
Overweight	20.1	(1.7, 38.6)	4.1	(-1.9, 10.2)	-2.3	(-22.8, 18.2)
Obese and above	12.9	(-5.3, 31.1)	0.3	(-5.7, 6.3)	-5.3	(-25.6, 15.1)
Total fat intake (g/day)	0.1	(-0.1, 0.3)	0.1	(0.0, 0.1)	0.0	(-0.3, 0.2)
No moderate / vigorous activity	4.9	(-9.9, 19.7)	1.1	(-3.7, 5.9)	5.9	(-9.9, 21.7)
Serum cotinine > 0.1 ng/mL	-14.7	(-30.6, 1.2)	-5.5	(-10.6, -0.3)	-2.2	(-18.9, 14.6)
Month of pregnancy	11.8	(8.9, 14.7)	1.5	(0.4, 2.6)	5.5	(2.0, 9.1)
Previous pregnancy						
0		*Referent*		*Referent*		*Referent*
1	18.6	(-0.7, 38.0)	2.0	(-4.4, 8.3)	20.9	(-0.2, 42.0)
2	-5.1	(-26.9, 16.7)	-5.2	(-12.3, 1.9)	3.8	(-19.2, 26.7)
≥ 3	11.5	(-9.7, 32.7)	-2.1	(-9.0, 4.9)	-0.7	(-26.2, 24.8)

Abbreviations: CI, confidence interval; IQR, interquartile range; N, number.

[a] PFOA Quartiles: Quartile 1 (≤ 1.2 ng/mL), Quartile 2 (1.3-2.3 ng/mL), Quartile 3 (2.4-3.2 ng/mL), Quartile 4 (≥ 3.3 ng/mL).

[b] PFOS Quartiles: Quartile 1 (≤ 6.1 ng/mL), Quartile 2 (6.2-10.0 ng/mL), Quartile 3 (10.1-14.7 ng/mL), Quartile 4 (≥ 14.8 ng/mL).

[c] Number of missing observations: 4 for annual family income, 1 for BMI (kg/m^2), 17 for total fat intake (g/day), 24 for month of pregnancy, 10 for previous pregnancy, and 82 for LDL.

[d] Normal: < 25 kg/m^2; overweight: 25-29.99 kg/m^2; obese and above: ≥ 30 kg/m^2.

Table 3-1. Adjusted associations between PFOA (ln-ng/mL) and serum cholesterol levels (mg/dL) of a cohort of U.S. pregnant women, based on NHANES data from 2003 to 2006 (N=134)[a].

Characteristics	Total cholesterol (mg/dL)		HDL cholesterol (mg/dL)		LDL cholesterol (mg/dL)	
	β estimate	95% CI	β estimate	95% CI	β estimate	95% CI
Intercept	121.6	(62.4, 180.8)	56.6	(33.9, 79.4)	56.0	(-38.5, 150.5)
PFOA (ln-ng/mL)	-2.0	(-13.0, 8.9)	-1.3	(-5.5, 2.9)	1.4	(-15.6, 18.4)
Age (years)	0.2	(-1.6, 1.9)	0.3	(-0.4, 1.0)	0.1	(-2.7, 2.9)
Race						
Non-Hispanic White		*Referent*		*Referent*		*Referent*
Mexican American	-6.6	(-26.3, 13.1)	1.2	(-6.3, 8.8)	-10.8	(-43.4, 21.8)
Non-Hispanic Black	-27.6	(-52.3, -2.9)	4.1	(-5.4, 13.5)	-11.9	(-42.3, 18.6)
Other	-10.0	(-37.9, 17.9)	1.3	(-9.4, 12.0)	-7.3	(-42.2, 27.5)
Education						
Less than high school		*Referent*		*Referent*		*Referent*
High school diploma or GED	17.5	(-4.6, 39.5)	-2.4	(-10.8, 6.1)	17.9	(-13.4, 49.1)
More than high school	7.2	(-13.9, 28.4)	1.1	(-7.0, 9.2)	4.6	(-22.6, 31.8)
Annual family income						
Under $20,000		*Referent*		*Referent*		*Referent*
$20,000 to $44,999	16.1	(-3.5, 35.7)	-1.4	(-8.9, 6.2)	7.9	(-20.5, 36.3)
$45,000 and over	12.1	(-10.3, 34.4)	-3.9	(-12.5, 4.7)	13.9	(-21.5, 49.3)
Body Mass Index (kg/m^2)[b]						
Normal		*Referent*		*Referent*		*Referent*
Overweight	-1.7	(-21.8, 18.4)	-0.3	(-8.0, 7.4)	-12.8	(-41.6, 16.1)
Obese and above	-8.6	(-29.6, 12.4)	-3.4	(-11.5, 4.6)	-13.1	(-42.3, 16.1)
Total fat intake (g/day)	0.2	(-0.1, 0.4)	0.1	(0.0, 0.1)	0.0	(-0.3, 0.3)
Serum cotinine > 0.1 ng/mL	11.0	(-9.5, 31.4)	-1.4	(-9.3, 6.4)	15.0	(-15.5, 45.6)
Month of pregnancy	11.7	(8.2, 15.2)	1.1	(-0.3, 2.4)	6.4	(1.4, 11.4)
Previous pregnancy						
0		*Referent*		*Referent*		*Referent*
1	18.9	(-0.4, 38.3)	1.2	(-6.3, 8.6)	19.9	(-3.5, 43.3)
2	9.7	(-13.0, 32.4)	-6.2	(-14.9, 2.6)	22.8	(-7.3, 52.9)
≥ 3	14.6	(-8.5, 37.7)	-2.8	(-11.7, 6.1)	5.6	(-27.4, 38.6)

Abbreviations: CI, confidence interval; N, number.

[a] Smaller than the sample size due to missing values of covariates. N = 66 for LDL model due to additional missing values for LDL.

[b] Normal: < 25 kg/m^2; overweight: 25-29.99 kg/m^2; obese and above: ≥ 30 kg/m^2.

Table 3-2. Adjusted associations between PFOS (ln-ng/mL) and serum cholesterol levels (mg/dL) of a cohort of U.S. pregnant women, based on NHANES data from 2003 to 2006 (N=134)[a].

Characteristics	Total cholesterol (mg/dL)		HDL cholesterol (mg/dL)		LDL cholesterol (mg/dL)	
	β estimate	95% CI	β estimate	95% CI	β estimate	95% CI
Intercept	114.6	(49.5, 179.7)	53.2	(28.1, 78.2)	-9.9	(-113.8, 93.9)
PFOS (ln-ng/mL)	1.3	(-11.4, 13.9)	0.4	(-4.5, 5.3)	21.0	(0.6, 41.4)
Age (years)	0.2	(-1.6, 2.0)	0.4	(-0.3, 1.1)	-0.1	(-2.6, 2.4)
Race						
Non-Hispanic White		*Referent*		*Referent*		*Referent*
Mexican American	-6.4	(-26.1, 13.3)	1.3	(-6.2, 8.9)	-0.1	(-32.8, 32.7)
Non-Hispanic Black	-27.1	(-51.6, -2.6)	4.5	(-4.9, 13.9)	-13.8	(-41.6, 14.1)
Other	-9.3	(-37.2, 18.5)	1.6	(-9.1, 12.4)	-0.8	(-34.4, 32.8)
Education						
Less than high school		*Referent*		*Referent*		*Referent*
High school diploma or GED	17.3	(-4.7, 39.3)	-2.5	(-11.0, 6.0)	19.6	(-10.3, 49.4)
More than high school	7.6	(-13.7, 28.8)	1.3	(-6.9, 9.4)	6.3	(-19.9, 32.4)
Annual family income						
Under $20,000		*Referent*		*Referent*		*Referent*
$20,000 to $44,999	16.1	(-3.6, 35.7)	-1.4	(-9.0, 6.1)	9.6	(-17.3, 36.5)
$45,000 and over	10.6	(-11.7, 32.8)	-4.7	(-13.2, 3.9)	13.1	(-19.6, 45.8)
Body Mass Index (kg/m^2)[b]						
Normal		*Referent*		*Referent*		*Referent*
Overweight	-1.3	(-21.6, 19.1)	-0.1	(-7.9, 7.8)	-4.6	(-33.4, 24.2)
Obese and above	-8.9	(-29.8, 12.0)	-3.7	(-11.7, 4.4)	-10.5	(-38.2, 17.2)
Total fat intake (g/day)	0.2	(-0.1, 0.4)	0.1	(0.0, 0.1)	0.1	(-0.2, 0.4)
Serum cotinine > 0.1 ng/mL	10.0	(-10.3, 30.3)	-2.0	(-9.8, 5.8)	16.1	(-13.0, 45.3)
Month of pregnancy	11.8	(8.4, 15.3)	1.2	(-0.2, 2.5)	7.6	(2.8, 12.3)
Previous pregnancy						
0		*Referent*		*Referent*		*Referent*
1	20.9	(1.3, 40.5)	2.2	(-5.3, 9.7)	29.1	(5.1, 53.0)
2	10.5	(-12.0, 32.9)	-5.7	(-14.3, 2.9)	30.3	(0.9, 59.7)
≥ 3	15.4	(-7.6, 38.4)	-2.3	(-11.2, 6.5)	6.2	(-25.4, 37.9)

Abbreviations: CI, confidence interval; N, number.

[a] Smaller than the sample size due to missing values of covariates. N = 66 for LDL model due to additional missing values for LDL.

[b] Normal: < 25 kg/m^2; overweight: 25-29.99 kg/m^2; obese and above: ≥ 30 kg/m^2.

Table 3-3. Adjusted associations between PFOA quartiles[a], PFOS quartiles[b] and serum cholesterol levels (mg/dL) of a cohort of U.S. pregnant women, based on NHANES data from 2003 to 2006 (N=134)[c].

PFCs[d]	Total cholesterol (mg/dL)		HDL cholesterol (mg/dL)		LDL cholesterol (mg/dL)	
	β estimate	95% CI	β estimate	95% CI	β estimate	95% CI
PFOA (ng/mL)						
Quartile 1		*Referent*		*Referent*		*Referent*
Quartile 2	-16.3	(-36.7, 4.2)	-4.1	(-12.0, 3.8)	-20.5	(-50.0, 9.1)
Quartile 3	-1.6	(-24.3, 21.1)	-2.5	(-11.3, 6.3)	-3.7	(-42.6, 35.1)
Quartile 4	-4.0	(-26.5, 18.4)	-4.2	(-12.9, 4.5)	0.5	(-33.7, 34.7)
PFOS (ng/mL)						
Quartile 1		*Referent*		*Referent*		*Referent*
Quartile 2	0.7	(-19.7, 21.1)	3.6	(-4.3, 11.4)	-3.2	(-30.9, 24.5)
Quartile 3	-0.4	(-21.0, 20.2)	1.3	(-6.6, 9.2)	11.2	(-16.9, 39.3)
Quartile 4	10.3	(-13.6, 34.2)	3.7	(-5.5, 12.8)	25.8	(-11.7, 63.2)

Abbreviations: CI, confidence interval; N, number.

[a] PFOA Quartiles: Quartile 1 (≤ 1.2 ng/mL), Quartile 2 (1.3-2.3 ng/mL), Quartile 3 (2.4-3.2 ng/mL), Quartile 4 (≥ 3.3 ng/mL).

[b] PFOS Quartiles: Quartile 1 (≤ 6.1 ng/mL), Quartile 2 (6.2-10.0 ng/mL), Quartile 3 (10.1-14.7 ng/mL), Quartile 4 (≥ 14.8 ng/mL).

[c] Smaller than the sample size due to missing values for covariates. N = 66 for LDL model due to additional missing values for LDL.

[d] PFOA and PFOS were analyzed separately with cholesterols.

APPENDICES

Appendix I: Additional Tables for Model Selection

Table I. Model selection (confounding assessment) for association between PFOA and total cholesterol, using forward selection method (N=174).

Model	Covariate	Estimate[a]	Change in Estimate[b]
1	-	-8.12	-
2	Age	-7.41	8.7%
3	PregMonth	-1.61	80.2%
4	Fat	-8.22	1.3%
5	Cotinine	-7.15	12.0%
6	Edu	-7.76	4.4%
7	Income	-10.06	24.0%
8	Race	-9.74	20.0%
9	Npreg	-5.12	36.9%
10	BMI	-6.80	16.2%
11	Activity	-7.93	2.3%
12	PregMonth + Age	-0.69	57.1%
13	PregMonth + Fat	-2.27	41.2%
14	PregMonth + Cotinine	-1.47	8.4%
15	PregMonth + Edu	-1.72	6.8%
16	PregMonth + Income	-3.10	92.8%
17	PregMonth + Race	-3.68	128.8%
18	PregMonth + Npreg	2.53	257.3%
19	PregMonth + BMI	-1.19	25.8%
20	PregMonth + Activity	-1.58	2.0%
21	PregMonth + Npreg + Age	2.91	15.2%
22	PregMonth + Npreg + Fat	2.24	11.5%
23	PregMonth + Npreg + Cotinine	2.72	7.5%
24	PregMonth + Npreg + Edu	2.36	6.6%
25	PregMonth + Npreg + Income	1.11	55.9%
26	PregMonth + Npreg + Race	0.26	89.9%
27	PregMonth + Npreg + BMI	3.08	21.7%
28	PregMonth + Npreg + Activity	2.50	1.3%
29	PregMonth + Npreg + Race + Age	0.60	132.8%
30	PregMonth + Npreg + Race + Fat	-0.13	150.4%
31	PregMonth + Npreg + Race + Cotinine	0.06	75.7%
32	PregMonth + Npreg + Race + Edu	0.47	81.3%
33	PregMonth + Npreg + Race + Income	-0.54	310.0%
34	PregMonth + Npreg + Race + BMI	0.79	208.0%
35	PregMonth + Npreg + Race + Activity	0.09	64.5%
36	PregMonth + Npreg + Race + Income + Age	-0.37	32.1%
37	PregMonth + Npreg + Race + Income + Fat	-1.05	94.7%
38	PregMonth + Npreg + Race + Income + Cotinine	-1.07	99.3%
39	PregMonth + Npreg + Race + Income + Edu	-0.45	16.7%
40	PregMonth + Npreg + Race + Income + BMI	0.10	119.0%
41	PregMonth + Npreg + Race + Income + Activity	-0.78	44.7%
42	PregMonth + Npreg + Race + Income + BMI + Age	0.29	179.6%

43	PregMonth + Npreg + Race + Income + BMI + Fat	-0.17	263.4%
44	PregMonth + Npreg + Race + Income + BMI + Cotinine	-0.43	523.1%
45	PregMonth + Npreg + Race + Income + BMI + Edu	0.13	22.9%
46	PregMonth + Npreg + Race + Income + BMI + Activity	-0.13	231.8%
47	PregMonth + Npreg + Race + Income + BMI + Cotinine + Age	-0.21	51.7%
48	PregMonth + Npreg + Race + Income + BMI + Cotinine + Fat	-1.46	237.0%
49	PregMonth + Npreg + Race + Income + BMI + Cotinine + Edu	-0.39	10.3%
50	PregMonth + Npreg + Race + Income + BMI + Cotinine + Activity	-0.60	38.4%
51	PregMonth + Npreg + Race + Income + BMI + Cotinine + Fat + Age	-1.58	8.2%
52	PregMonth + Npreg + Race + Income + BMI + Cotinine + Fat + Edu	-2.14	46.7%
53	PregMonth + Npreg + Race + Income + BMI + Cotinine + Fat + Act	-1.50	2.9%

[a] Coefficient estimate for ln-PFOA: the estimated change in total cholesterol per one ln-PFOA increase in serum.
[b] Change in the coefficient estimate for ln-PFOA, compared to the former selected model that includes less 1 covariate.

Table II. Model selection (confounding assessment) for association between PFOS and total cholesterol, using forward selection method (N=174).

Model	Covariate	Estimate[a]	Change in Estimate[b]
1	-	-8.74	-
2	Age	-8.49	2.9%
3	PregMonth	-4.67	46.5%
4	Fat	-7.91	9.5%
5	Cotinine	-7.22	17.3%
6	Edu	-8.95	2.4%
7	Income	-12.99	48.6%
8	Race	-6.64	24.0%
9	Npreg	-7.11	18.7%
10	BMI	-6.95	20.5%
11	Activity	-8.41	3.8%
12	Income + Age	-12.77	1.7%
13	Income + PregMonth	-6.32	51.4%
14	Income + Fat	-11.12	14.4%
15	Income + Cotinine	-11.49	11.5%
16	Income + Edu	-12.08	7.0%
17	Income + Race	-9.83	24.3%
18	Income + Npreg	-8.30	36.1%
19	Income + BMI	-10.98	15.4%
20	Income + Activity	-12.73	2.0%
21	Income + PregMonth + Age	-5.65	10.6%
22	Income + PregMonth + Fat	-5.45	13.8%
23	Income + PregMonth + Cotinine	-6.53	3.3%
24	Income + PregMonth + Edu	-6.10	3.4%
25	Income + PregMonth + Race	-5.21	17.6%
26	Income + PregMonth + Npreg	-0.49	92.2%
27	Income + PregMonth + BMI	-5.95	5.8%
28	Income + PregMonth + Activity	-6.15	2.6%
29	Income + PregMonth + Npreg + Age	-0.36	26.5%
30	Income + PregMonth + Npreg + Fat	-0.18	63.7%
31	Income + PregMonth + Npreg + Cotinine	-0.61	23.6%
32	Income + PregMonth + Npreg + Edu	-0.44	11.8%
33	Income + PregMonth + Npreg + Race	0.59	219.6%
34	Income + PregMonth + Npreg + BMI	-0.01	97.0%
35	Income + PregMonth + Npreg + Activity	-0.48	2.1%
36	Income + PregMonth + Npreg + Race + Age	0.70	18.1%
37	Income + PregMonth + Npreg + Race + Fat	1.67	181.5%
38	Income + PregMonth + Npreg + Race + Cotinine	0.22	63.4%
39	Income + PregMonth + Npreg + Race + Edu	0.81	36.8%
40	Income + PregMonth + Npreg + Race + BMI	0.83	39.9%
41	Income + PregMonth + Npreg + Race + Activity	0.55	7.8%
42	Income + PregMonth + Npreg + Race + Fat + Age	1.54	7.7%
43	Income + PregMonth + Npreg + Race + Fat + Cotinine	0.89	46.5%
44	Income + PregMonth + Npreg + Race + Fat + Edu	1.57	5.5%
45	Income + PregMonth + Npreg + Race + Fat + BMI	2.02	21.3%
46	Income + PregMonth + Npreg + Race + Fat + Activity	1.63	2.3%
47	Income + PregMonth + Npreg + Race + Fat + Cotinine + Age	0.88	1.3%

48	Income + PregMonth + Npreg + Race + Fat + Cotinine + Edu	0.87	2.5%
49	Income + PregMonth + Npreg + Race + Fat + Cotinine + BMI	1.03	15.1%
50	Income + PregMonth + Npreg + Race + Fat + Cotinine + Activity	0.89	0.2%

[a] Coefficient estimate for ln-PFOS: the estimated change in total cholesterol per one ln-PFOS increase in serum.
[b] Change in the coefficient estimate for ln-PFOS, compared to the former selected model that includes less 1 covariate.

Table III. Model Selection (confounding assessment) for association between PFOA and HDL, using forward selection method (N=174).

Model	Covariate	Estimate[a]	Change in Estimate[b]
1	-	-2.73	-
2	Age	-2.24	17.8%
3	PregMonth	-2.45	10.2%
4	Fat	-2.62	3.8%
5	Cotinine	-2.36	13.4%
6	Edu	-2.47	9.3%
7	Income	-3.37	23.6%
8	Race	-2.60	4.8%
9	Npreg	-3.12	14.4%
10	BMI	-2.41	11.7%
11	Activity	-2.69	1.4%
12	Income + Age	-2.71	19.5%
13	Income + PregMonth	-3.17	6.0%
14	Income + Fat	-3.09	8.3%
15	Income + Cotinine	-2.88	14.5%
16	Income + Edu	-2.79	17.2%
17	Income + Race	-3.16	6.2%
18	Income + Npreg	-3.71	10.0%
19	Income + BMI	-2.99	11.3%
20	Income + Activity	-3.36	0.3%
21	Income + Age + PregMonth	-2.47	9.0%
22	Income + Age + Npreg	-2.68	1.2%
23	Income + Age + Cotinine	-2.39	12.0%
24	Income + Age + Edu	-2.36	13.1%
25	Income + Age + Race	-2.54	6.4%
26	Income + Age + Npreg	-2.96	9.2%
27	Income + Age + BMI	-2.42	10.7%
28	Income + Age + Activity	-2.72	0.3%
29	Income + Age + Edu + PregMonth	-2.18	7.8%
30	Income + Age + Edu + Fat	-2.40	1.6%
31	Income + Age + Edu + Cotinine	-2.11	10.6%
32	Income + Age + Edu + Race	-2.03	13.8%
33	Income + Age + Edu + Npreg	-2.52	6.8%
34	Income + Age + Edu + BMI	-2.08	11.8%
35	Income + Age + Edu + Activity	-2.33	1.4%
36	Income + Age + Edu + Race + PregMonth	-1.89	6.9%
37	Income + Age + Edu + Race + Fat	-2.10	3.3%
38	Income + Age + Edu + Race + Cotinine	-1.92	5.5%
39	Income + Age + Edu + Race + Npreg	-2.35	15.4%
40	Income + Age + Edu + Race + BMI	-1.82	10.7%
41	Income + Age + Edu + Race + Activity	-2.04	0.2%
42	Income + Age + Edu + Race + Npreg + PregMonth	-1.90	18.8%
43	Income + Age + Edu + Race + Npreg + Fat	-2.19	6.9%
44	Income + Age + Edu + Race + Npreg + Cotinine	-2.20	6.1%
45	Income + Age + Edu + Race + Npreg + BMI	-2.08	11.4%
46	Income + Age + Edu + Race + Npreg + Activity	-2.41	2.9%
47	Income + Age + Edu + Race + Npreg + PregMonth + Fat	-1.87	1.9%

48	Income + Age + Edu + Race + Npreg + PregMonth + Cotinine	-1.81	4.9%
49	Income + Age + Edu + Race + Npreg + PregMonth + BMI	-1.74	8.8%
50	Income + Age + Edu + Race + Npreg + PregMonth + Activity	-1.99	4.3%

[a] Coefficient estimate for ln-PFOA: the estimated change in HDL per one ln-PFOA increase in serum.
[b] Change in the coefficient estimate for ln-PFOA, compared to the former selected model that includes less 1 covariate.

Table IV. Model Selection (confounding assessment) for association between PFOS and HDL, using forward selection method (N=174).

Model	Covariate	Estimate[a]	Change in Estimate[b]
1	-	-1.71	-
2	Age	-1.56	9.3%
3	PregMonth	-0.65	62.3%
4	Fat	-1.85	8.2%
5	Cotinine	-1.11	35.5%
6	Edu	-1.74	1.2%
7	Income	-2.98	74.0%
8	Race	-1.35	21.0%
9	Npreg	-1.59	7.0%
10	BMI	-1.28	25.4%
11	Activity	-1.63	4.6%
12	Income + Age	-2.39	19.8%
13	Income + PregMonth	-1.56	47.8%
14	Income + Fat	-2.49	16.6%
15	Income + Cotinine	-2.12	28.9%
16	Income + Edu	-2.42	18.9%
17	Income + Race	-2.51	15.9%
18	Income + Npreg	-2.47	17.2%
19	Income + BMI	-2.47	17.1%
20	Income + Activity	-2.92	2.0%
21	Income + PregMonth + Age	-0.95	39.2%
22	Income + PregMonth + Fat	-1.27	18.6%
23	Income + PregMonth + Cotinine	-1.03	33.7%
24	Income + PregMonth + Edu	-1.17	24.6%
25	Income + PregMonth + Race	-1.68	7.6%
26	Income + PregMonth + Npreg	-0.90	42.4%
27	Income + PregMonth + BMI	-1.32	15.2%
28	Income + PregMonth + Activity	-1.54	1.2%
29	Income + PregMonth + Npreg + Age	-0.38	57.3%
30	Income + PregMonth + Npreg + Fat	-0.47	48.1%
31	Income + PregMonth + Npreg + Cotinine	-0.34	62.3%
32	Income + PregMonth + Npreg + Edu	-0.47	48.1%
33	Income + PregMonth + Npreg + Race	-1.02	14.3%
34	Income + PregMonth + Npreg + BMI	-0.56	37.9%
35	Income + PregMonth + Npreg + Activity	-0.89	0.3%
36	Income + PregMonth + Npreg + Continine + Age	-0.17	49.2%
37	Income + PregMonth + Npreg + Continine + Fat	0.05	114.4%
38	Income + PregMonth + Npreg + Continine + Edu	-0.06	81.3%
39	Income + PregMonth + Npreg + Continine + Race	-0.57	68.4%
40	Income + PregMonth + Npreg + Continine + BMI	0.02	105.7%
41	Income + PregMonth + Npreg + Continine + Activity	-0.32	4.1%
42	Income + PregMonth + Npreg + Continine + Fat + Age	0.17	246.4%
43	Income + PregMonth + Npreg + Continine + Fat + Edu	0.29	496.3%
44	Income + PregMonth + Npreg + Continine + Fat + Race	-0.18	476.6%
45	Income + PregMonth + Npreg + Continine + Fat + BMI	0.45	833.2%
46	Income + PregMonth + Npreg + Continine + Fat + Activity	0.08	68.2%
47	Income + PregMonth + Npreg + Continine + Fat + BMI + Age	0.53	17.6%

48	Income + PregMonth + Npreg + Continine + Fat + BMI + Edu	0.58	26.9%
49	Income + PregMonth + Npreg + Continine + Fat + BMI + Race	0.17	63.0%
50	Income + PregMonth + Npreg + Continine + Fat + BMI + Activity	0.44	3.2%
51	Income + PregMonth + Npreg + Continine + Fat + BMI + Race + Age	0.29	73.6%
52	Income + PregMonth + Npreg + Continine + Fat + BMI + Race + Edu	0.35	105.5%
53	Income + PregMonth + Npreg + Continine + Fat + BMI + Race + Activity	0.13	24.6%
54	Income + PregMonth + Npreg + Continine + Fat + BMI + Race + Edu + Age	0.38	9.7%
55	Income + PregMonth + Npreg + Continine + Fat + BMI + Race + Edu + Activity	0.33	3.4%

[a] Coefficient estimate for ln-PFOS: the estimated change in HDL per one ln-PFOS increase in serum.
[b] Change in the coefficient estimate for ln-PFOS, compared to the former selected model that includes less 1 covariate.

Appendix II: SAS Code

```
*************************************************************************;
* Program 1: Main program for data cleaning, descriptives and analysis
* Programmer: Xinyi Zhao    Date: Apr. 2016;
*****************************************;

proc contents data=sasuser.ts;
run;

*-------------------;
*  Creating Format  ;
*-------------------;

proc format;
value eduf
  1 = "Less than high school"
  2 = "High school diploma or GED"
  3 = "More than high school";

value incomef
  1 = "Under $20,000"
  2 = "$20,000 to $44,999"
  3 = "$45,000 and over";

value racef
  1 = "Mexican American"
  2 = "Non-hispanic White"
  3 = "Non-hispanic Black"
  4 = "Other";

value npregf
  0 = "0"
  1 = "1"
  2 = "2"
  3 = "3+";

value pregmf
  1 = "1-3"
  2 = "4-6"
  3 = "7+";

value bmif
  1 = "Normal (<25)"
  2 = "Overweight (25-29.99)"
  3 = "Obese and above (30+)";

value cotf
  0 = "<= 0.1 ng/mL"
  1 = ">0.1 ng/mL";

value yesnof
  0 = "No"
  1 = "Yes";

value quartilef
```

```
  1 = "Q1"
  2 = "Q2"
  3 = "Q3"
  4 = "Q4";
run;

*-----------------;
*  Data Cleaning  ;
*-----------------;

data ts;
set sasuser.ts;
/*re-categorize annual family income*/
if income gt 0 and income le 4 then incomecat=1;
if income eq 13 then incomecat=1;
if income ge 5 and income le 7 then incomecat=2;
if income gt 7 and income lt 13 then incomecat=3;
/*1 = under $20,000
  2 = $20,000 to $44,999
  3 = $45,000 and over */

/* re-categorize race*/
if race=1 then racecat=1;
if race=2 or race=5 then racecat=4;
if race=3 then racecat=2;
if race=4 then racecat=3;
/* 1 = Mexican American
   2 = Non-hispanic White
   3 = Non hispanic Black
   4 = Other */

/* re-categorize former pregnancy times */
npr=npreg-1;
if npr=0 then np=0;
if npr=1 then np=1;
if npr=2 then np=2;
if npr ge 3 then np=3;
/* 0 = 0
   1 = 1
   2 = 2
   3 = 3+ */

/* dichotomize smoking */
if smk=1 then smkcat=1;
if smk=2 then smkcat=1;
if smk=3 then smkcat=0;

/* re-code alcohol use*/
if alc=1 then alq=1;
if alc=2 then alq=0;

/*pregnancy month category*/
if pregm ge 1 and pregm le 3 then tri=1;
if pregm ge 4 and pregm le 6 then tri=2;
if pregm ge 7 then tri=3;

/*BMI category*/
```

```
if bmi ge 0 and bmi lt 25 then bmicat=1;
if bmi ge 25 and bmi lt 30 then bmicat=2;
if bmi ge 30 then bmicat=3;
/* 1 = normal
   2 = overweight
   3 = obese and above*/

/* cotinine category*/
if cot ge 0 and cot le 0.1 then cotcat=0;
if cot gt 0.1 then cotcat=1;

/*log-transformed PFOA and PFOS*/
log_PFOA=log(PFOA);
log_PFOS=log(PFOS);

/*PFOA and PFOS category (by quartiles)*/
/*PFOA: Q1=1.2, Q2=2.3, Q3=3.2, Q4=13.6
  PFOS: Q1=6.1, Q2=10.05, Q3=14.70, Q4=39.0*/
if pfoa ge 0 and pfoa le 1.2 then pfoaq=1;
if pfoa gt 1.2 and pfoa le 2.3 then pfoaq=2;
if pfoa gt 2.3 and pfoa le 3.2 then pfoaq=3;
if pfoa gt 3.2 then pfoaq=4;

if pfos ge 0 and pfos le 6.1 then pfosq=1;
if pfos gt 6.1 and pfos le 10.05 then pfosq=2;
if pfos gt 10.05 and pfos le 14.70 then pfosq=3;
if pfos gt 14.70 then pfosq=4;

label incomecat="Annual Family Income Category"
      racecat="Race Category"
      alq="Had at least 12 alcohol drinks in any one year?"
      smkcat="Current Smoker"
      npr="Number of Previous Pregnancy"
      np="# Previous Pregnancy"
      tri="Pregnancy Month Group"
      bmicat="BMI Classification"
      cotcat="Serum Cotinine Level"
      pfoaq="PFOA Quartile"
      pfosq="PFOS Quartile";

format edu eduf.
       incomecat incomef.
       racecat racef.
       np npregf.
       tri pregmf.
       bmicat bmif.
       cotcat cotf.
       act yesnof.
       pfoaq quartilef.
       pfosq quartilef.;
run;

*check re-categorized variables;
proc freq data=ts;
tables income*incomecat/list;
run;
```

```
proc freq data=ts;
tables race*racecat/list;
run;

proc freq data=ts;
tables alc*alq/list;
run;

proc freq data=ts;
tables smk*smkcat/list;
run;

proc freq data=ts;
tables npreg*npr*np/list;
run;

proc freq data=ts;
tables pregm*tri/list;
run;

proc freq data=ts;
tables bmi*bmicat/list;
run;

proc freq data=ts;
tables cot*cotcat/list;
run;

proc freq data=ts;
tables pfoa*pfoaq/list;
run;

proc freq data=ts;
tables pfos*pfosq/list;
run;

proc contents data=ts;
run;

*--------------------------;
*  Descriptive Statistics  ;
*--------------------------;

proc means data=ts mean std median q1 q3 nmiss;
var age bmi fat cot pfoa pfos tcho hdl ldl;
run;

proc freq data=ts;
tables pfoaq pfosq racecat edu incomecat act pregm np tri bmicat cotcat;
run;

proc univariate data=ts plot;
var tcho hdl ldl pfoa pfos;
histogram tcho hdl ldl pfoa pfos/normal;
probplot tcho hdl ldl pfoa pfos;
run;
```

```
*----------------------------;
*  Univariate Associations  ;
*----------------------------;

* (1) continuous/dichotomous variable;
%macro uni(x);
proc univariate data=ts;
var &x;
histogram &x/normal;
run;

proc sgplot data=ts;
scatter Y=tcho X=&x;
reg Y=tcho X=&x;
label;
run;

proc sgplot data=ts;
scatter Y=hdl X=&x;
reg Y=hdl X=&x;
label;
run;

proc sgplot data=ts;
scatter Y=ldl X=&x;
reg Y=ldl X=&x;
label;
run;

proc reg data=ts;
model tcho=&x/clb;
run;

proc reg data=ts;
model hdl=&x/clb;
run;

proc reg data=ts;
model ldl=&x/clb;
run;

proc reg data=ts;
model log_pfoa=&x/clb;
run;

proc reg data=ts;
model log_pfos=&x/clb;
run;
%mend;

options mprint symbolgen;

%uni(pfoa);
%uni(pfos);
%uni(pfoaq);
%uni(pfosq);
%uni(log_pfoa);
```

```
%uni(log_pfos);
%uni(age);
%uni(bmi);
%uni(fat);
%uni(cot);
%uni(pregm);
%uni(npr);
%uni(act);
%uni(cotcat);

* (2) categorical variable (>2 levels);
%macro unic(x, ref);
proc sgplot data=ts;
vbar &x;
run;

proc glm data=ts;
class &x(ref=&ref);
model tcho=&x/solution clparm;
run;

proc glm data=ts;
class &x(ref=&ref);
model hdl=&x/solution clparm;
run;

proc glm data=ts;
class &x(ref=&ref);
model ldl=&x/solution clparm;
run;

proc glm data=ts;
class &x(ref=&ref);
model log_pfoa=&x/solution clparm;
run;

proc glm data=ts;
class &x(ref=&ref);
model log_pfos=&x/solution clparm;
run;
%mend;

options mprint symbolgen;

%unic(pfoaq, "Q1");
%unic(pfosq, "Q1");
%unic(edu, "Less than high school");
%unic(incomecat, "Under $20,000");
%unic(racecat, "Non-hispanic White");
%unic(tri, "1-3");
%unic(np, "0");
%unic(bmicat, "Normal (<25)");
%unic(act, "Yes");

* Pregnancy month and exposures;
proc sgplot data=ts;
scatter Y=pfoa X=pregm;
```

```
reg Y=pfoa X=pregm;
label;
run;

proc sgplot data=ts;
scatter Y=pfos X=pregm;
reg Y=pfos X=pregm;
label;
run;

proc reg data=ts;
model pfoa=pregm;
run;

proc reg data=ts;
model pfos=pregm;
run;

proc glm data=ts;
class tri(ref="1-3");
model pfoa=tri/solution;
run;

proc glm data=ts;
class tri(ref="1-3");
model pfos=tri/solution;
run;

**** Compare different scales of predictors;
* continuous scale;
%macro con(x);
proc mixed method=ml data=ts;
model tcho=&x/solution;
run;

proc mixed method=ml data=ts;
model hdl=&x/solution;
run;

proc mixed method=ml data=ts;
model ldl=&x/solution;
run;
%mend;

%con(pfoa);
%con(log_pfoa);
%con(pfos);
%con(log_pfos);
%con(bmi);
%con(cot);
%con(cotcat);
%con(pregm);
%con(npr);

* categorical scale;
%macro cat(x, ref);
proc mixed method=ml data=ts;
```

```
class &x(ref=&ref);
model tcho=&x/solution;
run;

proc mixed method=ml data=ts;
class &x(ref=&ref);
model hdl=&x/solution;
run;

proc mixed method=ml data=ts;
class &x(ref=&ref);
model ldl=&x/solution;
run;
%mend;

%cat(pfoaq, "Q1");
%cat(pfosq, "Q1");
%cat(bmicat, "Normal (<25)");
%cat(tri, "1-3");
%cat(np, "0");

*-----------------------------;
*  Multivariate Associations  ;
*-----------------------------;

****** Model 1: PFOA and total cholesterol;

* (1) only exposure;
proc glm data=ts;
model tcho=log_pfoa/solution;
run;

* (2) 2-predictor model;
%macro twopredcon(x);
proc glm data=ts;
model tcho=log_pfoa &x/solution;
run;
%mend;

%macro twopredcat(x,ref);
proc glm data=ts;
class &x(ref=&ref);
model tcho=log_pfoa &x/solution;
run;
%mend;

%twopredcon(age);
%twopredcon(pregm);
%twopredcon(fat);
%twopredcon(cotcat);
%twopredcat(edu, "Less than high school");
%twopredcat(incomecat, "Under $20,000");
%twopredcat(racecat, "Non-hispanic White");
%twopredcat(np, "0");
%twopredcat(bmicat, "Normal (<25)");
%twopredcat(act, "Yes");
```

```
* (3) 3-predictor model;

%macro threepredcon(x);
proc glm data=ts;
model tcho=log_pfoa pregm &x/solution;
run;
%mend;

%macro threepredcat(x,ref);
proc glm data=ts;
class &x(ref=&ref);
model tcho=log_pfoa pregm &x/solution;
run;
%mend;

%threepredcon(age);
%threepredcon(fat);
%threepredcon(cotcat);
%threepredcat(edu, "Less than high school");
%threepredcat(incomecat, "Under $20,000");
%threepredcat(racecat, "Non-hispanic White");
%threepredcat(np, "0");
%threepredcat(bmicat, "Normal (<25)");
%threepredcat(act, "Yes");

* (4) 4-predictor model;

%macro fourpredcon(x);
proc glm data=ts;
class np(ref="0");
model tcho=log_pfoa pregm np &x/solution;
run;
%mend;

%macro fourpredcat(x,ref);
proc glm data=ts;
class np(ref="0") &x(ref=&ref);
model tcho=log_pfoa pregm np &x/solution;
run;
%mend;

%fourpredcon(age);
%fourpredcon(fat);
%fourpredcon(cotcat);
%fourpredcat(edu, "Less than high school");
%fourpredcat(incomecat, "Under $20,000");
%fourpredcat(racecat, "Non-hispanic White");
%fourpredcat(bmicat, "Normal (<25)");
%fourpredcat(act, "Yes");

* (5) 5-predictor model;

%macro fivepredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White");
model tcho=log_pfoa pregm np racecat &x/solution;
run;
```

```
%mend;

%macro fivepredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") &x(ref=&ref);
model tcho=log_pfoa pregm np racecat &x/solution;
run;
%mend;

%fivepredcon(age);
%fivepredcon(fat);
%fivepredcon(cotcat);
%fivepredcat(edu, "Less than high school");
%fivepredcat(incomecat, "Under $20,000");
%fivepredcat(bmicat, "Normal (<25)");
%fivepredcat(act, "Yes");

* (6) 6-predictor model;

%macro sixpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000");
model tcho=log_pfoa pregm np racecat incomecat &x/solution;
run;
%mend;

%macro sixpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") &x(ref=&ref);
model tcho=log_pfoa pregm np racecat incomecat &x/solution;
run;
%mend;

%sixpredcon(age);
%sixpredcon(fat);
%sixpredcon(cotcat);
%sixpredcat(edu, "Less than high school");
%sixpredcat(bmicat, "Normal (<25)");
%sixpredcat(act, "Yes");

* (7) 7-predictor model;

%macro sevenpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)");
model tcho=log_pfoa pregm np racecat incomecat bmicat &x/solution;
run;
%mend;

%macro sevenpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") &x(ref=&ref);
model tcho=log_pfoa pregm np racecat incomecat bmicat &x/solution;
```

```
run;
%mend;

%sevenpredcon(age);
%sevenpredcon(fat);
%sevenpredcon(cotcat);
%sevenpredcat(edu, "Less than high school");
%sevenpredcat(act, "Yes");

* (8) 8-predictor model;

%macro eightpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)");
model tcho=log_pfoa pregm np racecat incomecat bmicat cotcat &x/solution;
run;
%mend;

%macro eightpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") &x(ref=&ref);
model tcho=log_pfoa pregm np racecat incomecat bmicat cotcat &x/solution;
run;
%mend;

%eightpredcon(age);
%eightpredcon(fat);
%eightpredcat(edu, "Less than high school");
%eightpredcat(act, "Yes");

* (9) 9-predictor model;

%macro ninepredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)");
model tcho=log_pfoa pregm np racecat incomecat bmicat cotcat fat &x/solution;
run;
%mend;

%macro ninepredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") &x(ref=&ref);
model tcho=log_pfoa pregm np racecat incomecat bmicat cotcat fat &x/solution;
run;
%mend;

%ninepredcon(age);
%ninepredcat(edu, "Less than high school");
%ninepredcat(act, "Yes");

* add additional covariates to be consistent with HDL model;

proc glm data=ts;
```

```
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") edu(ref="Less than high school");
model tcho=log_pfoa pregm np racecat incomecat bmicat cotcat fat edu
age/solution;
run;


****** Model 2: PFOS and total cholesterol;

* (1) only exposure;
proc glm data=ts;
model tcho=log_pfos/solution;
run;

* (2) 2-predictor model;
%macro twopredcon(x);
proc glm data=ts;
model tcho=log_pfos &x/solution;
run;
%mend;

%macro twopredcat(x,ref);
proc glm data=ts;
class &x(ref=&ref);
model tcho=log_pfos &x/solution;
run;
%mend;

%twopredcon(age);
%twopredcon(pregm);
%twopredcon(fat);
%twopredcon(cotcat);
%twopredcat(edu, "Less than high school");
%twopredcat(incomecat, "Under $20,000");
%twopredcat(racecat, "Non-hispanic White");
%twopredcat(np, "0");
%twopredcat(bmicat, "Normal (<25)");
%twopredcat(act, "Yes");

* (3) 3-predictor model;

%macro threepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model tcho=log_pfos incomecat &x/solution;
run;
%mend;

%macro threepredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model tcho=log_pfos incomecat &x/solution;
run;
%mend;

%threepredcon(age);
%threepredcon(pregm);
```

```
%threepredcon(fat);
%threepredcon(cotcat);
%threepredcat(edu, "Less than high school");
%threepredcat(racecat, "Non-hispanic White");
%threepredcat(np, "0");
%threepredcat(bmicat, "Normal (<25)");
%threepredcat(act, "Yes");

* (4) 4-predictor model;

%macro fourpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model tcho=log_pfos incomecat pregm &x/solution;
run;
%mend;

%macro fourpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model tcho=log_pfos incomecat pregm &x/solution;
run;
%mend;

%fourpredcon(Age);
%fourpredcon(fat);
%fourpredcon(cotcat);
%fourpredcat(edu, "Less than high school");
%fourpredcat(racecat, "Non-hispanic White");
%fourpredcat(np, "0");
%fourpredcat(bmicat, "Normal (<25)");
%fourpredcat(act, "Yes");

* (5) 5-predictor model;

%macro fivepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") np(ref="0");
model tcho=log_pfos incomecat pregm np &x/solution;
run;
%mend;

%macro fivepredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") np(ref="0") &x(ref=&ref);
model tcho=log_pfos incomecat pregm np &x/solution;
run;
%mend;

%fivepredcon(age);
%fivepredcon(fat);
%fivepredcon(cotcat);
%fivepredcat(edu, "Less than high school");
%fivepredcat(racecat, "Non-hispanic White");
%fivepredcat(bmicat, "Normal (<25)");
%fivepredcat(act, "Yes");
```

```
* (6) 6-predictor model;

%macro sixpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000");
model tcho=log_pfos pregm np racecat incomecat &x/solution;
run;
%mend;

%macro sixpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") &x(ref=&ref);
model tcho=log_pfos pregm np racecat incomecat &x/solution;
run;
%mend;

%sixpredcon(age);
%sixpredcon(fat);
%sixpredcon(cotcat);
%sixpredcat(edu, "Less than high school");
%sixpredcat(bmicat, "Normal (<25)");
%sixpredcat(act, "Yes");

* (7) 7-predictor model;

%macro sevenpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000");
model tcho=log_pfos pregm np racecat incomecat fat &x/solution;
run;
%mend;

%macro sevenpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") &x(ref=&ref);
model tcho=log_pfos pregm np racecat incomecat fat &x/solution;
run;
%mend;

%sevenpredcon(age);
%sevenpredcon(cotcat);
%sevenpredcat(edu, "Less than high school");
%sevenpredcat(bmicat, "Normal (<25)");
%sevenpredcat(act, "Yes");

* (8) 8-predictor model;

%macro eightpredcon(x);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000");
model tcho=log_pfos pregm np racecat incomecat fat cotcat &x/solution;
run;
```

```
%mend;

%macro eightpredcat(x,ref);
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") &x(ref=&ref);
model tcho=log_pfos pregm np racecat incomecat fat cotcat &x/solution;
run;
%mend;

%eightpredcon(age);
%eightpredcat(edu, "Less than high school");
%eightpredcat(bmicat, "Normal (<25)");
%eightpredcat(act, "Yes");

* add additional covariates to be consistent with HDL model;
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") edu(ref="Less than high school");
model tcho=log_pfos pregm np racecat incomecat bmicat cotcat fat edu
age/solution;
run;


****** Model 3: PFOA and HDL;

* (1) only exposure;
proc glm data=ts;
model HDL=log_pfoa/solution;
run;

* (2) 2-predictor model;
%macro twopredcon(x);
proc glm data=ts;
model HDL=log_pfoa &x/solution;
run;
%mend;

%macro twopredcat(x,ref);
proc glm data=ts;
class &x(ref=&ref);
model HDL=log_pfoa &x/solution;
run;
%mend;

%twopredcon(age);
%twopredcon(pregm);
%twopredcon(fat);
%twopredcon(cotcat);
%twopredcat(edu, "Less than high school");
%twopredcat(incomecat, "Under $20,000");
%twopredcat(racecat, "Non-hispanic White");
%twopredcat(np, "0");
%twopredcat(bmicat, "Normal (<25)");
%twopredcat(act, "Yes");

* (3) 3-predictor model;
```

```
%macro threepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model HDL=log_pfoa incomecat &x/solution;
run;
%mend;

%macro threepredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfoa incomecat &x/solution;
run;
%mend;

%threepredcon(age);
%threepredcon(pregm);
%threepredcon(fat);
%threepredcon(cotcat);
%threepredcat(edu, "Less than high school");
%threepredcat(racecat, "Non-hispanic White");
%threepredcat(np, "0");
%threepredcat(bmicat, "Normal (<25)");
%threepredcat(act, "Yes");

* (4) 4-predictor model;

%macro fourpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model HDL=log_pfoa incomecat age &x/solution;
run;
%mend;

%macro fourpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfoa incomecat age &x/solution;
run;
%mend;

%fourpredcon(pregm);
%fourpredcon(fat);
%fourpredcon(cotcat);
%fourpredcat(edu, "Less than high school");
%fourpredcat(racecat, "Non-hispanic White");
%fourpredcat(np, "0");
%fourpredcat(bmicat, "Normal (<25)");
%fourpredcat(act, "Yes");

* (5) 5-predictor model;

%macro fivepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high school");
model HDL=log_pfoa incomecat age edu &x/solution;
run;
```

```
%mend;

%macro fivepredcat(x,ref);
proc glm data=ts;
class  incomecat(ref="Under $20,000") edu(ref="Less than high school")
&x(ref=&ref);
model HDL=log_pfoa incomecat age edu &x/solution;
run;
%mend;

%fivepredcon(pregm);
%fivepredcon(fat);
%fivepredcon(cotcat);
%fivepredcat(racecat, "Non-hispanic White");
%fivepredcat(np, "0");
%fivepredcat(bmicat, "Normal (<25)");
%fivepredcat(act, "Yes");

* (6) 6-predictor model;

%macro sixpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high school")
racecat(ref="Non-hispanic White");
model HDL=log_pfoa incomecat age edu racecat &x/solution;
run;
%mend;

%macro sixpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high
school")racecat(ref="Non-hispanic White") &x(ref=&ref);
model HDL=log_pfoa incomecat age edu racecat &x/solution;
run;
%mend;

%sixpredcon(pregm);
%sixpredcon(fat);
%sixpredcon(cotcat);
%sixpredcat(np, "0");
%sixpredcat(bmicat, "Normal (<25)");
%sixpredcat(act, "Yes");

* (7) 7-predictor model;

%macro sevenpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high school")
racecat(ref="Non-hispanic White")np(ref="0");
model HDL=log_pfoa age incomecat edu racecat np &x/solution;
run;
%mend;

%macro sevenpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high
school")racecat(ref="Non-hispanic White")np(ref="0") &x(ref=&ref);
```

```
model HDL=log_pfoa age incomecat edu racecat np &x/solution;
run;
%mend;

%sevenpredcon(pregm);
%sevenpredcon(fat);
%sevenpredcon(cotcat);
%sevenpredcat(bmicat, "Normal (<25)");
%sevenpredcat(act, "Yes");

* (8) 8-predictor model;

%macro eightpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high school")
racecat(ref="Non-hispanic White")np(ref="0");
model HDL=log_pfoa age incomecat edu racecat np pregm &x/solution;
run;
%mend;

%macro eightpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") edu(ref="Less than high
school")racecat(ref="Non-hispanic White")np(ref="0") &x(ref=&ref);
model HDL=log_pfoa age incomecat edu racecat np pregm &x/solution;
run;
%mend;

%eightpredcon(fat);
%eightpredcon(cotcat);
%eightpredcat(bmicat, "Normal (<25)");
%eightpredcat(act, "Yes");

* add additional covariates to be consistent with TCHO model;
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") edu(ref="Less than high school");
model hdl=log_pfoa pregm np racecat incomecat bmicat cotcat fat edu
age/solution;
run;

****** Model 4: PFOS and HDL;

* (1) only exposure;
proc glm data=ts;
model HDL=log_pfos/solution;
run;

* (2) 2-predictor model;
%macro twopredcon(x);
proc glm data=ts;
model HDL=log_pfos &x/solution;
run;
%mend;

%macro twopredcat(x,ref);
proc glm data=ts;
```

```
class &x(ref=&ref);
model HDL=log_pfos &x/solution;
run;
%mend;

%twopredcon(age);
%twopredcon(pregm);
%twopredcon(fat);
%twopredcon(cotcat);
%twopredcat(edu, "Less than high school");
%twopredcat(incomecat, "Under $20,000");
%twopredcat(racecat, "Non-hispanic White");
%twopredcat(np, "0");
%twopredcat(bmicat, "Normal (<25)");
%twopredcat(act, "Yes");

* (3) 3-predictor model;

%macro threepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model HDL=log_pfos incomecat &x/solution;
run;
%mend;

%macro threepredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfos incomecat &x/solution;
run;
%mend;

%threepredcon(age);
%threepredcon(pregm);
%threepredcon(fat);
%threepredcon(cotcat);
%threepredcat(edu, "Less than high school");
%threepredcat(racecat, "Non-hispanic White");
%threepredcat(np, "0");
%threepredcat(bmicat, "Normal (<25)");
%threepredcat(act, "Yes");

* (4) 4-predictor model;

%macro fourpredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000");
model HDL=log_pfos incomecat pregm &x/solution;
run;
%mend;

%macro fourpredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfos incomecat pregm &x/solution;
run;
%mend;
```

```
%fourpredcon(Age);
%fourpredcon(fat);
%fourpredcon(cotcat);
%fourpredcat(edu, "Less than high school");
%fourpredcat(racecat, "Non-hispanic White");
%fourpredcat(np, "0");
%fourpredcat(bmicat, "Normal (<25)");
%fourpredcat(act, "Yes");

* (5) 5-predictor model;

%macro fivepredcon(x);
proc glm data=ts;
class incomecat(ref="Under $20,000") np(ref="0");
model HDL=log_pfos incomecat pregm np &x/solution;
run;
%mend;

%macro fivepredcat(x,ref);
proc glm data=ts;
class incomecat(ref="Under $20,000") np(ref="0") &x(ref=&ref);
model HDL=log_pfos incomecat pregm np &x/solution;
run;
%mend;

%fivepredcon(age);
%fivepredcon(fat);
%fivepredcon(cotcat);
%fivepredcat(edu, "Less than high school");
%fivepredcat(racecat, "Non-hispanic White");
%fivepredcat(bmicat, "Normal (<25)");
%fivepredcat(act, "Yes");

* (6) 6-predictor model;

%macro sixpredcon(x);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000");
model HDL=log_pfos pregm np cotcat incomecat &x/solution;
run;
%mend;

%macro sixpredcat(x,ref);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfos pregm np cotcat incomecat &x/solution;
run;
%mend;

%sixpredcon(age);
%sixpredcon(fat);
%sixpredcat(edu, "Less than high school");
%sixpredcat(racecat, "Non-hispanic White");
%sixpredcat(bmicat, "Normal (<25)");
%sixpredcat(act, "Yes");
```

```
* (7) 7-predictor model;

%macro sevenpredcon(x);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000");
model HDL=log_pfos pregm np cotcat incomecat fat &x/solution;
run;
%mend;

%macro sevenpredcat(x,ref);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") &x(ref=&ref);
model HDL=log_pfos pregm np cotcat incomecat fat &x/solution;
run;
%mend;

%sevenpredcon(age);
%sevenpredcat(edu, "Less than high school");
%sevenpredcat(racecat, "Non-hispanic White");
%sevenpredcat(bmicat, "Normal (<25)");
%sevenpredcat(act, "Yes");

* (8) 8-predictor model;

%macro eightpredcon(x);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)");
model HDL=log_pfos pregm np cotcat incomecat fat bmicat &x/solution;
run;
%mend;

%macro eightpredcat(x,ref);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
&x(ref=&ref);
model HDL=log_pfos pregm np cotcat incomecat fat bmicat &x/solution;
run;
%mend;

%eightpredcon(age);
%eightpredcat(edu, "Less than high school");
%eightpredcat(racecat, "Non-hispanic White");
%eightpredcat(act, "Yes");

* (9) 9-predictor model;

%macro ninepredcon(x);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
racecat(ref="Non-hispanic White");
model HDL=log_pfos pregm np cotcat incomecat fat bmicat racecat &x/solution;
run;
%mend;

%macro ninepredcat(x,ref);
proc glm data=ts;
```

```
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
racecat(ref="Non-hispanic White") &x(ref=&ref);
model HDL=log_pfos pregm np cotcat incomecat fat bmicat racecat &x/solution;
run;
%mend;

%ninepredcon(age);
%ninepredcat(edu, "Less than high school");
%ninepredcat(act, "Yes");

* (10) 10-predictor model;

%macro tenpredcon(x);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
racecat(ref="Non-hispanic White") edu(ref="Less than high school");
model HDL=log_pfos pregm np cotcat incomecat fat bmicat racecat edu
&x/solution;
run;
%mend;

%macro tenpredcat(x,ref);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
racecat(ref="Non-hispanic White") edu(ref="Less than high school")
&x(ref=&ref);
model HDL=log_pfos pregm np cotcat incomecat fat bmicat racecat edu
&x/solution;
run;
%mend;

%tenpredcon(age);
%tenpredcat(act, "Yes");

* add additional covariates to be consistent with TCHO model;
proc glm data=ts;
class np(ref="0") racecat(ref="Non-hispanic White") incomecat(ref="Under
$20,000") bmicat(ref="Normal (<25)") edu(ref="Less than high school");
model hdl=log_pfos pregm np racecat incomecat bmicat cotcat fat edu
age/solution;
run;

********** FINAL MODELS;

* use macro to simplify coding because all models control for the same
covariates;
%macro fm(x, y);
proc glm data=ts;
class np(ref="0") incomecat(ref="Under $20,000") bmicat(ref="Normal (<25)")
racecat(ref="Non-hispanic White") edu(ref="Less than high school");
model &y=&x pregm np cotcat incomecat fat bmicat racecat edu age/solution
clparm;
run;
%mend;

%fm(log_pfoa, tcho);
%fm(log_pfoa, hdl);
```

```
%fm(log_pfoa, ldl);
%fm(log_pfos, tcho);
%fm(log_pfos, hdl);
%fm(log_pfos, ldl);

%macro fm(x, y);
proc glm data=ts;
class &x(ref="Q1") np(ref="0") incomecat(ref="Under $20,000")
bmicat(ref="Normal (<25)") racecat(ref="Non-hispanic White") edu(ref="Less
than high school");
model &y=&x pregm np cotcat incomecat fat bmicat racecat edu age/solution
clparm;
run;
%mend;

%fm(pfoaq, tcho);
%fm(pfoaq, hdl);
%fm(pfoaq, ldl);
%fm(pfosq, tcho);
%fm(pfosq, hdl);
%fm(pfosq, ldl);

*-------------------------;
*    Model Diagnostics    ;
*-------------------------;

* to get VIF, we need to use proc reg instead of proc glm;
* to perform proc reg, we need to create indicator variables first;

data diag;
set ts;

incomecat2=0; incomecat3=0;
if incomecat=2 then incomecat2=1;
if incomecat=3 then incomecat3=1;
if incomecat=. then do; incomecat2=.; incomecat3=.; end;

edu2=0; edu3=0;
if edu=2 then edu2=1;
if edu=3 then edu3=1;
if edu=. then do; edu2=.; edu3=.; end;

bmicat2=0; bmicat3=0;
if bmicat=2 then bmicat2=1;
if bmicat=3 then bmicat3=1;
if bmicat=. then do; bmicat2=.; bmicat3=.; end;

racecat1=0; racecat3=0; racecat4=0;
if racecat=1 then racecat1=1;
if racecat=3 then racecat3=1;
if racecat=4 then racecat4=1;
if racecat=. then do; racecat1=.; racecat3=.; racecat4=.; end;

np1=0; np2=0; np3=0;
if np=1 then np1=1;
if np=2 then np2=1;
if np=3 then np3=1;
```

```
if np=. then do; np1=.; np2=.; np3=.; end;

run;

proc freq data=diag;
tables incomecat*incomecat2*incomecat3/list;
run;

proc freq data=diag;
tables edu*edu2*edu3/list;
run;

proc freq data=diag;
tables bmicat*bmicat2*bmicat3/list;
run;

proc freq data=diag;
tables racecat*racecat1*racecat3*racecat4/list;
run;

proc freq data=diag;
tables np*np1*np2*np3/list;
run;

* Diagnostics;
%macro dgn(x, y, t, n);
proc reg data=diag;
model &y=&x incomecat2 incomecat3 pregm np1 np2 np3 cotcat fat bmicat2
bmicat3
      racecat1 racecat3 racecat4 edu2 edu3 age/clb partial pcorr2 influence R
vif;
id id;
  output
   out=resid
   R=residual
   rstudent=jackres
   cookd=cooks
   H =leverage;
run;

data resid;
  set resid;
  outlier_jk = (abs(jackres) > &t);
  outlier_cooks = (cooks > 18/&n);
  outlier_leverage = (leverage > 2*18/&n);
  if sum(outlier_jk, outlier_cooks, outlier_leverage) >=1;
  nm = nmiss(of &x &y incomecat pregm np cotcat fat bmicat racecat edu age);
  if nm > 0 then delete;
  keep id residual cooks leverage jackres outlier_jk outlier_cooks
outlier_leverage;
run;

proc print data=resid;
run;
%mend;

%dgn(log_pfoa, tcho, 1.98, 134);
```

```
%dgn(log_pfoa, hdl, 1.98, 134);
%dgn(log_pfoa, ldl, 2, 66);
%dgn(log_pfos, tcho, 1.98, 134);
%dgn(log_pfos, hdl, 1.98, 134);
%dgn(log_pfos, ldl, 2, 66);
```

```
*************************************************************************;
* Program 0: Program to merge, combine the raw data from NHANES;
* Programmer: Xinyi Zhao    Date: Jan. 2016;
*****************************************;

libname thesis "H:\Thesis\data";

*---------------------------;
* Transform raw data format ;
*---------------------------;
/*
libname inlib xport 'H:\Thesis\data\COT_D.xpt';
proc copy in=inlib out=thesis;
run;
*/
*-------------------------------;
* Combine, Merge and Clean Data ;
*-------------------------------;

**** Data for main variables (PFOA, PFOS and cholesterol levels);

* Macro for data from 2005-2012;
%macro yrdata(dsn, pf, cho, ldl, hdl, dt, yr);
data preg;
set &dsn;
if URXPREG=1; /*1=positive*/
run;

/*merge with Polyfluoroalkyl Chemicals data*/
proc sort data=preg; by seqn; run;
proc sort data=&pf; by seqn; run;

data temp1;
merge preg(in=a) &pf(in=b);
by seqn;
if a;
if LBXPFOA=. and LBXPFOS=. then delete; /*LBXPFOA=PFOA LBXPFOS=PFOS*/
run;

/*merge with Total cholestrol data*/
proc sort data=temp1; by seqn; run;
proc sort data=&cho; by seqn; run;

data temp2;
merge temp1(in=a) &cho(in=b);
by seqn;
if a;
run;

/*merge with LDL data*/
proc sort data=temp2; by seqn; run;
proc sort data=&ldl; by seqn; run;

data temp3;
merge temp2(in=a) &ldl(in=b);
by seqn;
if a;
```

```
run;

/*merge with HDL data*/
proc sort data=temp3; by seqn; run;
proc sort data=&hdl; by seqn; run;

data temp4;
merge temp3(in=a) &hdl(in=b);
by seqn;
if a;
run;

data &dt;
set temp4;
year="&yr";
if LBXTC=. and LBDLDL=. and LBDHDD=. then delete;
/*LBXTC=Total cholestrol (mg/dl) LBDLDL=LDL (mg/dl) LBDHDD=HDL (mg/dl)*/
/*missing data of HDL and LDL is included in the missing data of total
cholesterol (checked)*/
run;

proc freq data=&dt;
tables URXPREG;
title "Available data from &yr";
run;

%mend;
options mprint symbolgen;

/*2005-2006*/
%yrdata(thesis.preg_0506, thesis.pf_0506, thesis.cho_0506, thesis.ldl_0506,
thesis.hdl_0506, dt0506, 2005-2006);


* Macro for data from 1999-2004;
* -- use another macro because some codings are different from those for
2005-2012 data;
%macro yrdata(dsn, pf, chohdl, ldl, dt, yr);
data preg;
set &dsn;
if URXPREG=1; /*1=positive*/
run;

/*merge with Polyfluoroalkyl Chemicals data*/
proc sort data=preg; by seqn; run;
proc sort data=&pf; by seqn; run;

data temp1;
merge preg(in=a) &pf(in=b);
by seqn;
if a;
if LBXPFOA=. and LBXPFOS=. then delete; /*LBXPFOA=PFOA LBXPFOS=PFOS*/
run;

/*merge with Total cholestrol and HDL data*/
proc sort data=temp1; by seqn; run;
proc sort data=&chohdl; by seqn; run;
```

```
data temp2;
merge temp1(in=a) &chohdl(in=b);
by seqn;
if a;
rename LBXHDD=LBDHDD;
run;

/*merge with LDL data*/
proc sort data=temp2; by seqn; run;
proc sort data=&ldl; by seqn; run;

data temp3;
merge temp2(in=a) &ldl(in=b);
by seqn;
if a;
run;

data &dt;
set temp3;
year="&yr";
if LBXTC=. and LBDLDL=. and LBDHDD=. then delete;
/*LBXTC=Total cholestrol (mg/dl) LBDLDL=LDL (mg/dl) LBDHDD=HDL (mg/dl)*/
/*missing data of HDL and LDL is included in the missing data of total
cholesterol (checked)*/
run;

proc freq data=&dt;
tables URXPREG;
title "Available data from &yr";
run;

%mend;
options mprint symbolgen;

/*2003-2004*/
%yrdata(thesis.preg_0304, thesis.pf_0304, thesis.cho_hdl_0304,
thesis.ldl_0304, dt0304, 2003-2004);

data thesis.thesis;
set dt0304 dt0506;
rename LBXPFOA=PFOA
       LBXPFOS=PFOS
       LBXTC=TCHO
       LBDLDL=LDL
       LBDHDD=HDL;
keep SEQN URXPREG LBXPFOA LBXPFOS LBXTC LBDLDL LBDHDD year;
run;

proc freq data=thesis.thesis;
tables year;
run;

* check data availability;
data check;
set thesis;
if PFOA ne . then PFOAt=1;
```

```
if PFOS ne . then PFOSt=1;
if TCHO ne . then TCHOt=1;
if LDL ne . then LDLt=1;
if HDL ne . then HDLt=1;

proc freq data=check;
tables PFOAt PFOSt TCHOt LDLt HDLt;
run;

proc freq data=check;
tables PFOAt*TCHOt PFOAt*LDLt PFOAt*HDLt
       PFOSt*TCHOt PFOSt*LDLt PFOSt*HDLt;
run;

**** Data for covariates;

*** Pregnancy month;

data month0304;
set thesis.tri_0304;
if RHD152=. then delete;
if RHD152=77 then delete;
if RHD152=99 then delete;
rename RHD152=pregm;
keep SEQN RHD152;
run;

proc freq data=month0304;
tables pregm;
run;

data month0506;
set thesis.tri_0506;
if RHD152=. then delete;
if RHD152=77 then delete;
if RHD152=99 then delete;
rename RHD152=pregm;
keep SEQN RHD152;
run;

proc freq data=month0506;
tables pregm;
run;

data month;
set month0304 month0506;
run;

proc sort data=month; by seqn; run;
proc sort data=thesis.thesis; by seqn; run;

data thesis.thesis2;
merge thesis.thesis(in=a) month(in=b);
by seqn;
if a;
run;
```

```
proc freq data=thesis.thesis2;
tables pregm;
run;

**** other potential covariates;
/*
RIDAGEYR = age
RIDRETH1 = race
DMDEDUC = education
INDFMINC = annual family income
BMXBMI = BMI
SMQ040 = current smoker (Do you now smoke cigarettes)
ALQ101 = alcohol intake (Had at least 12 alcohol drinks/1 yr?)
DR1TTFAT = Total fat intake (g) in day 1
DR2TTFAT = Total fat intake (g) in day 2
PAD200 = Vigorous activity over past 30 days
PAD320 = Moderate activity over past 30 days
RHQ200 = Now breastfeeding a child?
RHQ210 = Breastfed any of your children?
*/

*** demographics (age, race, education and income);

data demo;
set thesis.demo0304 thesis.demo0506;
rename RIDAGEYR = age
       RIDRETH1 = race
       DMDEDUC2 = edua
       DMDEDUC3 = eduy
       INDFMINC = income;
if DMDEDUC3 > 66 then DMDEDUC3 = .;
if DMDEDUC2 > 5 then DMDEDUC2 = .;
if INDFMINC > 13 then INDFMINC = .;
/* re-categorize education*/
if DMDEDUC3 gt 0 and DMDEDUC3 lt 13 then edu_y = 1; /* 1 = less than high
school */
if DMDEDUC3 = 55 then edu_y = 1;
if DMDEDUC3 = 66 then edu_y = 1;
if DMDEDUC3 = 13 then edu_y = 2; /* 2 = high school diploma or GED*/
if DMDEDUC3 = 14 then edu_y = 2;
if DMDEDUC3 = 15 then edu_y = 3; /* 3 = more than high school*/
if DMDEDUC2 gt 0 and DMDEDUC2 lt 3 then edu_a = 1;
if DMDEDUC2 = 3 then edu_a = 2;
if DMDEDUC2 = 4 then edu_a = 3;
if DMDEDUC2 = 5 then edu_a = 3;
keep SEQN RIDAGEYR RIDRETH1 DMDEDUC2 DMDEDUC3 INDFMINC edu_a edu_y;
run;

proc freq data=demo;
tables edua*edu_a/list;
run;

proc freq data=demo;
tables eduy*edu_y/list;
run;

proc sort data=demo; by seqn; run;
```

```
proc sort data=thesis.thesis2; by seqn; run;

data t3;
merge thesis.thesis2(in=a) demo(in=b);
by seqn;
if a;
if  age gt 0 and age le 19 then edu = edu_y;
if age gt 19 then edu = edu_a;
label edu = "Education level";
run;

proc print data=t3;
var eduy edu_y edua edu_a edu;
where age > 19;
run;

proc print data=t3;
var eduy edu_y edua edu_a edu;
where age le 19;
run;

*** BMI;

data bmx;
set thesis.bmx0304 thesis.bmx0506;
rename BMXBMI = bmi;
keep SEQN BMXBMI;
run;

proc sort data=bmx; by seqn; run;
proc sort data=t3; by seqn; run;

data t4;
merge t3(in=a) bmx(in=b);
by seqn;
if a;
drop edua eduy edu_a edu_y;
run;

*** physical activity;

data pad;
set thesis.pad0304 thesis.pad0506;
rename PAD320 = actm
       PAD200 = actv;
if PAD320 = 1 then act_m = 1;
if PAD320 = 2 or PAD320 = 3 then act_m = 0;
if PAD320 > 3 then act_m = .;
if PAD200 = 1 then act_v = 1;
if PAD200 = 2 or PAD200 = 3 then act_v = 0;
if PAD200 > 3 then act_v = .;
if act_v = 1 or act_m = 1 then act = 1; /* 1 = did moderate activity or above
*/
if act_v = 0 and act_m = 0 then act = 0;
keep SEQN PAD320 PAD200 act_m act_v act;
run;
```

```
proc freq data=pad;
tables actm*act_m/list;
run;

proc freq data=pad;
tables actv*act_v/list;
run;

proc freq data=pad;
tables act_m*act_v*act/list;
run;

proc sort data=pad; by seqn; run;
proc sort data=t4; by seqn; run;

data t5;
merge t4(in=a) pad(in=b);
by seqn;
if a;
drop actm actv act_m act_v;
label act = "Moderate or vigorous activity over past 30 days";
run;

*** smoking (current smoker);

data smk;
set thesis.smk0304 thesis.smk0506;
rename SMQ040 = smk;
if SMQ040 gt 3 then SMQ040 = .;
keep SEQN SMQ040;
run;

proc sort data=smk; by seqn; run;
proc sort data=t5; by seqn; run;

data t6;
merge t5(in=a) smk(in=b);
by seqn;
if a;
run;

*** alcohol use;

data alc;
set thesis.alc0304 thesis.alc0506;
rename ALQ101 = alc
       ALQ130 = alc2;
if ALQ101 gt 2 then ALQ101 = .;
if ALQ130 gt 50 then ALQ130 = .;
keep SEQN ALQ101 ALQ130;
run;

proc sort data=alc; by seqn; run;
proc sort data=t6; by seqn; run;

data t7;
merge t6(in=a) alc(in=b);
```

```
by seqn;
if a;
run;

* total fat intake (daily);

/*calculate daily fat intake by data of two non-consecutive days*/
proc sort data=thesis.diet03041; by seqn; run;
proc sort data=thesis.diet03042; by seqn; run;

data diet0304;
merge thesis.diet03041 thesis.diet03042;
by seqn;
rename DR1TTFAT = fat1
       DR2TTFAT = fat2;
fat = (DR1TTFAT + DR2TTFAT)/2;
keep SEQN DR1TTFAT DR2TTFAT fat;
run;

proc sort data=thesis.diet05061; by seqn; run;
proc sort data=thesis.diet05062; by seqn; run;

data diet0506;
merge thesis.diet05061 thesis.diet05062;
by seqn;
rename DR1TTFAT = fat1
       DR2TTFAT = fat2;
fat = (DR1TTFAT + DR2TTFAT)/2;
keep SEQN DR1TTFAT DR2TTFAT fat;
run;

data diet;
set diet0304 diet0506;
drop fat1 fat2;
run;

proc sort data=diet; by seqn; run;
proc sort data=t7; by seqn; run;

data t8;
merge t7(in=a) diet(in=b);
by seqn;
if a;
label fat = "Total fat intake (g/day)";
run;

*** breast-feeding (ever or not);

data brst;
set thesis.brst0304 thesis.brst0506;
rename RHQ200 = bfc
       RHQ210 = bff
       RHQ160 = npreg;
if RHQ200 gt 2 then RHQ200 = .;
if RHQ210 gt 2 then RHQ210 = .;
if RHQ200 = 1 then bf_c = 1; /* 1 = current breast-feeding */
if RHQ200 = 2 then bf_c = 0;
```

```
if RHQ210 = 1 then bf_f = 1; /* 1 = former breast-feeding */
if RHQ210 = 2 then bf_f = 0;
if RHQ160 gt 50 then RHQ160 = .;
keep SEQN RHQ210 RHQ200 bf_c bf_f RHQ160;
run;

proc freq data=brst;
tables bfc*bf_c/list;
run;

proc freq data=brst;
tables bff*bf_f/list;
run;

proc sort data=brst; by seqn; run;
proc sort data=t8; by seqn; run;

data thesis.thesis3;
merge t8(in=a) brst(in=b);
by seqn;
if a;
if bf_c = 1 or bf_f = 1 then bf = 1; /* 1 = ever breast-feeding*/
if bf_c = 0 and bf_f = 0 then bf = 0;
drop bfc bff;
label bf = "Ever breast-feeding >= 1 child"
      bf_c = "Current breast-feeding"
      bf_f = "Former breast-feeding";
run;

*** Cotinine;

data cot;
set thesis.cot0304 thesis.cot0506;
rename LBXCOT = cot;
keep SEQN LBXCOT;
run;

proc sort data=cot; by seqn; run;
proc sort data=thesis.thesis3; by seqn; run;

data thesis.thesis4;
merge thesis.thesis3(in=a) cot(in=b);
by seqn;
if a;
run;

data thesis.ts;
set thesis.thesis4;
drop  URXPREG;
rename SEQN=id;
run;
```