China Health and Retirement Longitudinal Study

2011-2012 National Baseline Blood Data Users' Guide

Yaohui Zhao

Eileen Crimmins

Peifeng (Perry) Hu

Yisong Hu

Tao Ge

Jung Ki Kim

John Strauss

Gonghuan Yang

Xiangjun Yin

Yafeng Wang

March 2014

China Center for Economic Research

Peking University

Preface

This document describes the overall process, including the design, implementation and data release, of the blood samples for the China Health and Retirement Longitudinal Study national baseline survey in 2011-2012. This manual aims to enhance the users' understanding and application of the blood data.

The China Health and Retirement Longitudinal Study (CHARLS) is a survey of the mid-aged and elderly in China, run by the National School for Development (China Center for Economic Research) at Peking University (PKU). It is based on a sample of households with members aged 45 years or above. It attempts to set up a high quality public micro-database, which can provide a wide range of information from socio-economic status to health conditions, to serve the needs of scientific research on the mid-aged and elderly.

CHARLS is harmonized to the Health and Retirement Study (HRS) and related aging surveys such as the English Longitudinal Study of Ageing (ELSA) and the Survey of Health, Aging and Retirement in Europe (SHARE). Considering the enormous complexity involved in a national survey, we began with a pilot survey in just two provinces in 2008: Gansu, a poor inland province, and Zhejiang, a rich coastal province. The pilot survey collected data from 95 communities/villages in 32 counties/districts, covering 2,685 individuals living in 1,570 households. The pilot survey produced a set of high quality survey data, demonstrated that fielding an HRS-type survey in China is feasible. Based on pilot survey experiences, CHARLS conducted its national baseline survey in 2011-2012. To ensure sample representativeness, the CHARLS baseline survey covered 150 counties/districts, 450 villages/urban communities, across the country. We successfully interviewed 17,708 individuals in 10,257 households, reflecting the Chinese mid-aged and elderly population collectively.

Acknowledgements

The China Health and Retirement Longitudinal Study (CHARLS) is an enormous project that required the efforts of many people. We want to express our gratitude to the CHARLS research team, the field team, and every respondent. Thank you all for the time, energy, and passion you've devoted to the project, and also for your understanding and support for the CHARLS project.

CHARLS project is a collaborative effort of many scholars at home and abroad. The Principal Investigator is Professor Zhao Yaohui, National School of Development (China Center for Economic Research) at Peking University. We have additional Principal Investigators: Professor John Strauss from the University of Southern California, and Professor Gonghuan Yang now at the Peking Union Medical University. Dr. Peifeng (Perry) Hu of the University of California, Los Angeles and Professor Eileen Crimmins of the University of Southern California are co-Principal Investigators on the supplement grant having to do with analysis of blood data described in this User Guide.

The CHARLS fieldwork for blood was administered by a staff led by Yisong Hu, together with Dr. Xiangjun Yin of the National Center for Chronic and Non-communicable Disease Control and Prevention of the Chinese Center for Disease Control and Prevention (NCNCD, China CDC). The lab work done at Capital Medical University was overseen by Dr. Tao Ge.

Yun Wu, trained and assisted by Albert (Bas) Weerman from RAND Corporation, led the programmers who programmed the questionnaire into CAPI. Yafeng Wang led the effort in creating sampling weights. Haiyu Jin managed the CHARLS accounts, payments to interviewers and subcontractors.

The China Health and Retirement Longitudinal Study (CHARLS) has received critical support from both home and abroad. Behavioral and Social Research division of the National Institute on Aging (NIA) of the National Institutes of Health in the United States (grant numbers 1-R21-AG031372-01, 1-R21-AG033675-01A1, 1-R01-AG037031-01 and 1-R01-AG037031-03S1), the Management Science Division of the National Natural Science Foundation of China (grant number 70773002, 70910107022, 71130002), Beijing Representative Office of the World Bank (contract number 7145915) and Knowledge for Change Program of the World Bank Group (contract number 7159234) all provided critical financial support for our project. Here, we want to extend our deepest gratitude to all the

sponsors.

CHARLS received valuable support from many individuals and organizations home and abroad. We thank Dr. Richard Suzman, Dr. John Phillips and the staff at the Division of Behavioral Science of the NIA, Deputy Directors of management Science Division of the NSFC, Dr. Yijun Li and Dr. Ziyou Gao, and the Scientific Committee led by Professor Haizhen Li. We thank Peking University Vice President Wei Liu, former Peking University Vice President Jianhua Lin, Director of "985 program" office and Assistant President of Peking University Xiaoming Li, Director of the Office of Social Sciences and ISSS Qiang Li, Honorary Dean of the NSD Justin Yifu Lin, former Dean of the NSD Qiren Zhou, current Dean of the NSD Yang Yao, and officials from the National Bureau of Statistics, Ministry of Health, Ministry of Labor and Social Security, Ministry of Civil Affairs, Ministry of Education, and Ministry of Construction.

CHARLS benefitted tremendously from intellectual support from members of our international and domestic advisory board members led by Dr. James Smith of Rand Corporation, including James Banks, Lisa Berkman, David Bloom, Axel Borsch-Supan, Arie Kapteyn, Jinkook Lee, David Weir, Robert Willis, David Wise, Qiren Zhou, Fang Cai, Scott Rozelle, Cangping Wu, Yang Yao and Xuejin Zuo.

We thank our colleagues Yan Sun, Yue Zhang and Minyan Chen and staff from the Institute of Social Science Survey of Peking University, Dr. Linhong Wang and Dr. Zhixin Li from NCNCD, China CDC, Xiaoming Shi, and Zhaoxue Yin from the China CDC, 28 Provincial Centers of CDC, 150 County or District Centers of CDC and the health systems of the counties and cities, and all leaders from 450 communities involved in the survey for their assistance in ensuring the smoothness of the fieldwork. We are greatly indebted to them.

The survey could not have taken place without the understanding and support of all households participated in CHARLS project. The data provided not only lays the foundation for academic studies on the Chinese aging problem, but also throws light on the future development of social welfare system for our government. We extend our sincerest thanks.

Contents

1 Introduction

The aim of collecting whole blood in CHARLS is to produce a fuller profile of biological risk and physiological functioning in China, both for scientists studying health in China and to make international comparisons. China is undergoing dramatic economic and social change which is linked to changes in physiological status and risks for both infectious and chronic disease. Examination of differences across China and changes over time in this risk will increase our understanding of how economic development is linked to a transition in health and physiological functions. Furthermore, examining biomarkers within the large representative sample of CHARLS with the wealth of individual social, economic, behavioral, psychological, and other physical performance and health data will allow us to understand the mechanisms or pathways through which economic development and social change affect health.

With the economic and demographic changes in China, we expect that the prevalence of recognized risk factors for cardiovascular disease, including hypertension and adverse lipid levels (indicated by cholesterols and triglycerides), will increase. We expect the increase of food availability and the decrease in physical activity will be linked to increasing risk for obesity and diabetes (indicated by fasting glucose and glycosylated hemoglobin). On the other hand, malnutrition should be reduced (indicated by improved hemoglobin levels). Decreases in infectious exposure and treatment of infectious conditions should be linked to lower probability of extreme inflammatory burden among individuals. However, increased risk of atherosclerosis and metabolic syndrome may lead to higher prevalence of chronic and low-grade inflammation in the population (which may be detected by high sensitivity C-reactive protein-hsCRP). Kidney functioning may be affected differently by two trends. Improving treatment for kidney and urinary tract infections and severe urinary retention may improve kidney functioning (indicated by levels of creatinine and cystatin C). On the other hand, the increased prevalence of hypertension, diabetes, and atherosclerosis may increase the risk of kidney impairment. These changes linked to economic development will lead to marked differentials in health risk in the population now as well as changes in the process of aging.

CHARLS has the following non-blood biomarkers, which were collected as part of the main survey: anthropometric measurements: height, weight, waist circumference, right lower leg length and upper arm length, blood pressure, pulse rate, lung peak flow rate, grip strength, timed sit-to-stand, timed walk, and balance measures (see the National

Baseline User Guide). In addition, we have collected venous blood samples and assayed them for hsCRP, HbA1c, total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, glucose, blood urea nitrogen (BUN), creatinine, uric acid, and cystatin C. In addition, a complete blood count (CBC) analysis was been done at local CDC laboratories (this included hemoglobin, hematocrit, white blood cell count, platelet counts and mean corpuscular volume).

The rationale for adding biomarkers to population surveys is that they validate and add nuance to self-reported health information, they allow richer modeling of pathways of influence between the socioeconomic and the physical, and they may capture aspects of health unknown to survey participants (Weir 2008). For instance, aging is accompanied by an increasing likelihood of dysregulation within physiological systems that may precede the diagnosis of disease and disability. Biomarkers such as blood pressure and cholesterol levels may indicate preclinical or pre-morbid dysregulation which may be unknown to survey participants, especially those without regular preventive health care. Because research has increasingly clarified that many of the important health outcomes associated with aging have similar risk factors (e.g. mortality, loss of both physical and cognitive functioning, and cardiovascular disease), the value of collecting biomarkers related to these health outcomes along with socioeconomic and psychological information in population surveys has become widely recognized (Crimmins and Seeman 2001, 2004). The research community has become both convinced of the value of biomarkers and their collection has been included in a growing number of current population studies including HRS, the English Longitudinal Study of Ageing (ELSA), the Japanese Longitudinal Study of Aging (JSTAR),[1] the Indonesia Family Life Survey (IFLS), the Mexican Family Life Survey (MxFLS), the World Health Organization sponsored Study on Global Aging and Adult Health (SAGE) surveys, data in Germany in the Survey of Health, Ageing and Retirement in Europe (SHARE), and others that are planning on their incorporation (e.g. LASI). While many of these studies use dried blood spots (DBS) for their blood assays, ELSA uses venous blood. For CHARLS, we also collected venous blood.

[1] JSTAR links their data to other data with venous blood samples on the same respondents.

2 Blood Sample Collection, Processing, Transportation, and Storage

For the blood collection in the national survey we used staff of the Chinese Center for Disease Control and Prevention (China CDC), while for the main survey, including the collection of non-blood biomarkers, we used CHARLS enumerators. We have worked in tandem with China CDC staff in developing the main questionnaire, field protocols, and other aspects of the survey. Working with China CDC had several advantages in this very large country, especially as CHARLS became fully national in scope. Most importantly for this project, working with the China CDC allowed us to collect venous blood instead of relying on dried blood spots. China CDC has a nationwide network and trained staff in all counties. Each county CDC has basic laboratory facilities, even in the most remote regions. Moreover they have experience in large-scale surveys of collecting blood, processing samples de-centrally and shipping them to Beijing for analyses at a central lab. The collection of venous blood, blood processing, initial local analysis of complete blood cell counts, and specimen shipment to Beijing for the first national wave of CHARLS was supported by the Chinese National Natural Science Foundation funds under a project entitled, "Health Needs of the Mature and Older Adults in China". The central laboratory assays of the samples were supported by a competing revision R01 from the National Institute on Aging.

Three tubes of venous blood were collected from each respondent by medically-trained staff from the China CDC, based on a standard protocol. We asked respondents to have fasted overnight, but we took blood even if they were not fasting and we noted their status, which is a variable in the data. Over 92% of respondents who gave blood reported that they were fasting. For most respondents the blood collection was done at centralized locations. In urban areas the district CDC was the central location. In rural areas, the county CDC stations or the town/village health centers were used, whichever was closer. If the location was remote, the village health center was used as the collection center. If none of these were possible, the exam and blood collection was done at the respondent's home, using the same standards as at the central places (these procedures are covered by a standard CDC manual, "Blood Collection and Handling").

Identification of the appropriate person was done by matching the name, sex, date of birth, barcode (given by CHARLS enumerators at the main interview) and national identity card number, by county CDC staff. Respondents were told to bring their barcodes and identity

cards to the site of the blood collection. The county CDC staff in the field checked the barcodes. The staff had the respondent's identity card numbers but minus the last four digits. They then used the actual card to fill in the last 4 digits so that we could later verify that the person was the correct one.

The first tube of blood, a 2 mL tube, was used for a complete blood count (CBC) test, including white blood cells, hemoglobin, hematocrit, platelet counts, and mean corpuscular volume. These were measured on automated analyzers available at county CDC stations or town/village health centers. Over 97% of these laboratories participate in a regular assay quality control program.

After collection, these fresh venous blood samples were transported, at 4oC temperature, to either local CDC laboratories or township level hospitals near the study sites. For the overwhelming majority of the respondents (75%) who had fasting blood specimens collected at centralized locations, the CBC was measured within 141 minutes of collection. The median time from collection to CBC assay was 97 minutes.

Second, a 4 mL tube of whole blood was collected to obtain plasma (yield a little under 2 mL of plasma) and buffy coat, which contains predominantly white blood cells. The venous blood was processed and divided into these two components within the same timeframe as the CBC measurement. Transport to the local lab (if transport is required) was at $4\,^{\circ}$C. After the venous blood was separated into plasma and buffy coat, the plasma was then stored in three 0.5 mL cryovials and the buffy coat in a separate cryovial. These cryovials were then immediately stored frozen at $-20\,^{\circ}$C and transported to the Chinese CDC in Beijing within 2 weeks where they were placed in a deep freezer and stored at $-80\,^{\circ}$C until assay at CMU laboratory.

Finally, one 2 mL tube was collected for the HbA1c assay. This 2 mL tube of whole blood was stored immediately and during shipment at $4\,^{\circ}$C as described above, and transported to the China CDC in Beijing within 2 weeks, where it was placed at $-80\,^{\circ}$C in a deep freezer for the HbA1c assay.

3 Handling of Blood Samples and Development of Data Base

Two China CDC staff worked full time on maintaining storage of the blood samples; they logged in blood tubes and cryovials as they were received. All tubes and cryovials have barcodes attached to them. For a given respondent each barcode has the same number. When the blood samples were first collected in the field, the barcode was scanned into the CAPI file for the respondent, so that we had a clear link between the barcode number and the CHARLS ID. When the county CDC labs conducted the CBC and when the CMU laboratory conducted its assays, the barcodes were scanned again at the labs in order to make sure we had the correct identification. Files were maintained so that the whereabouts of all samples could be determined at any time. As samples were removed from freezers and sent for assay, files were sent with barcodes and not survey ID numbers. China CDC staff personally carried the samples to the CMU laboratory. Each assay result was merged with the data file by two China CDC staff to ensure accuracy.

4 Methods for Blood-Based Bioassays

In addition to the CBC, which was performed in the laboratories at the county level, we measured high-sensitivity CRP, glycosylated hemoglobin (HbA1c), a lipid panel (total, HDL, LDL cholesterol, and triglycerides), glucose, BUN, creatinine, uric acid, and cystatin C from frozen plasma or whole blood samples. These assays were performed at the Youanmen Center for Clinical Laboratory of Capital Medical University. The assay methods used in this laboratory, coefficients of variation (CVs), and detection limits are summarized below.

Table 1: Methods, Coefficient of Variation, and Detection Limits of the Bioassays

Biomarkers	Method	Coefficient of variation		Detection limits
		Within-assay	Between-assay	
hsCRP	Immunoturbidimetric assay	$< 1.3\%$	$< 5.7\%$	0.1 - 20 mg/L
Glycosylated hemoglobin (HbA1c)	Boronate affinity HPLC	1.90%	2.10%	0 - 40%
Total-C	Enzymatic colormetric test	0.80%	1.70%	3 - 800 mg/dL
HDL-C	Enzymatic colormetric test	1.00%	1.30%	3 - 120 mg/dL
LDL-C	Enzymatic colormetric test	0.70%	1.20%	3 - 400 mg/dL
Triglycerides	Enzymatic colormetric test	1.50%	1.80%	4-1000 mg/dL
Blood urea nitrogen (BUN)	Enzymatic UV method with urease	<4.4%	<4.1%	5- 100 mg/dL
Creatinine	Rate-blanked and compensated Jaffe creatinine method	<1.6%	<2.1%	0.1 - 25 mg/dL
Glucose	Enzymatic colormetric test	0.90%	1.80%	2 - 450 mg/dL
Uric acid	UA Plus method	1.10%	1.90%	Up to 20 mg/dL
Cystatin C	Particle-enhanced turbimetric assay	$< 5\%$	$< 5\%$	0.5 - 8 mg/L

This CMU laboratory has regular external quality assessment organized by the Chinese Ministry of Health and conducts assay quality control samples on a daily basis. The Center

has been accredited by the Beijing Health Bureau and offers approximately 200 different clinical assays in immunology, chemistry, hematology, virology, and molecular biology. This laboratory has excellent performance during annual evaluation by External Quality Assurance (EQA) Program organized by the National Center for Clinical Laboratories, China Ministry of Health.

The laboratory used quality control (QC) samples daily during the testing of the CHARLS study samples (from February of 2013 to June of 2013). The total numbers of QC samples tested were 117 for blood chemistry tests and lipid panels, 59 for HbA1c assay, and 42 for hsCRP assay. All test results from these QC samples were within the target range (within two standard deviations of mean QC control concentrations).

Perry Hu and Tao Ge checked initial results on a weekly basis to make sure that assays appeared to be within range, that values seem appropriate, and that the quality-control (QC) samples indicate reliability in the process.

5 Sampling and Sample Weights

The CHARLS national baseline survey was conducted in 28 provinces, 150 countries / districts, 450 villages/urban communities, across the country. The CHARLS sample is representative of people aged 45 and over, living in households; institutionalized mid-aged and elderly are not sampled, but Wave 1 respondents who later enter into an institution will be followed. We successfully interviewed 17,708 individuals in 10,257 households, reflecting the Chinese mid-aged and elderly population collectively. Details of the sampling can be found in Zhao et al. (2013).

The target sample for taking blood samples was the entire group of 17,708 main respondents and spouses from the main CHARLS national baseline. Out of this, we collected blood samples for 11,847 individuals, a response rate of 67% (see Table 2). Women had a slightly higher response rate than men (69% versus 65%), and rural respondents had a higher rate: 71% versus 60% for urban respondents.

Table 2: Response Rate of Blood Sample

	Total			Urban			Rural		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
Target Sample	17,708	8,476	9,232	7,171	3,351	3,820	10,537	5,125	5,412
Complete Sample	11,847	5,504	6,343	4,317	1,962	2,355	7,530	3,542	3,988
Response Rate	0.67	0.65	0.69	0.60	0.59	0.62	0.71	0.69	0.74

Because of non-response to the blood collection, we calculate sample weights with a correction for household and individual non-response as well as non-participation in the blood collection, which we call bloodweight in the data. To construct this weight we start with the individual-level sample weights we constructed for the main data. These weights are constructed from our weights for households. We use weights with corrections for both household and individual non-response (households that were sampled but did not respond to the baseline survey; individual non-response for the main questionnaire, given that the household did respond), because most researchers will want to use the blood data together with individual data on age, sex and other attributes found in the individual data.

For the non-response correction, we use an inverse probability weighting that we multiply with the uncorrected household weight.[2]The inverse probability weighting factor is

[2]See Jeffrey Wooldridge, 2002, *Econometric Analysis for Cross Section and Panel Data*, Cambridge: MIT Press, for details.

constructed by first estimating a logit regression of whether the individual participates in the blood collection (using main respondents and their spouses as the observations) as a function of village-level dummies, dummies for the respondent's age, sex and their interactions, and schooling level dummies. Once the logit regressions are run, we take the inverse of predicted probability for each household and cap them at the 99th percentile, so that no observation is unduly weighted. This is our inverse probability weight that we use to multiply by the (inverted) individual weight (with household and individual non-response), taken from the main data. These non-response corrections do require a selection on observables assumption (that is, there is no selection on unobservables, Wooldridge, 2002), which is strong.

The results are presented in Table 3. Women are more likely to get their blood samples taken. There is a strong age-gradient for men, with younger men being less likely to have their blood taken (perhaps because they are more likely to be working and did not take the time). The age gradient for women is much flatter. Education levels do not significantly affect the odds of being measured.

Table 3: Logit Regression for Inverse Probability Weight

Dependent Variable: Participation in Blood Test		
Independent Variables	Coefficient	Std.
Female	0.433***	0.075
Age Groups		
50- (reference)		
50-54	0.235***	0.086
55-59	0.353***	0.079
60-64	0.540***	0.086
65-69	0.632***	0.099
70+	0.496***	0.087
Female cross Age Groups		
Female & 50-54	0.005	0.119
Female & 55-59	-0.071	0.109
Female & 60-64	-0.296**	0.116
Female & 65-69	-0.329**	0.137
Female & 70+	-0.618***	0.114
Education Groups		
Illiteracy (reference)		
Literate	0.076	0.059
Primary	0.044	0.059
Junior and Above	-0.023	0.060
Constant	1.970***	0.609
Community dummies	Yes	
Observations	17708	

** Significant at p=.05; *** significant at p=.01.

6 Values from Blood-Based Assays

Descriptive statistics from the CBC assays are shown in Table 4; values from the other blood-based assays are shown in Table 5. Non-parametric kernel density plots in Figures 1-9 are constructed after omitting a small number of extreme values. These densities appear to be reasonable. The mean and median levels are comparable in the measures of total cholesterol, HDL, LDL, glucose and uric acid, showing somewhat normal distribution in the density plots, while CRP, HbA1c, triglycerides, BUN, creatinine, and cystatin C have density plots skewed to the right. Table 6 provides estimates of the population frequencies at high risk levels for those assays that have agreed upon, international threshold levels. Twenty-eight percent of the mid-aged and elderly Chinese population have low levels of HDL cholesterol, men at more risk than women for this measure. High levels of triglycerides is a problem in China, with 15.2% of the mid-aged and elderly population suffering from this. However only 10% of the mid-aged and elderly population have high total cholesterol, and about the same for high LDL (measured directly). In both cases, the levels for women are higher than for men. Some 18.7% have elevated CRP levels, a little higher for men. Some 13.7% have low hemoglobin levels, women being more at risk for this measure, while 13.5% have elevated blood glucose levels, about the same for men and women. A smaller percentage have elevated HbA1c (4.8%) than elevated glucose levels. The correlation between HbA1c and plasma glucose is 0.64 ($p<.0001$), and the correlation between high risk HbA1c and high risk glucose is 0.41 ($p<.0001$). This discrepancy between Hba1c and glucose levels is reduced somewhat because of the presence of samples from non-fasting persons. About 8% of the respondents provided non-fasting blood samples and glucose levels from these specimens may be higher than if they had been fasted. The correlation between high HbA1c and high plasma glucose is 0.31 among those who were not fasted; while it is 0.43 among the fasters. Second, it is possible that the values of HbA1c obtained from the samples that were frozen upon arrival in Beijing are lower than true values. While a number of studies report that frozen samples can be used, it is not standard clinical laboratory practice. Previous studies have shown that the results from HbA1c assays that use the affinity high performance liquid chromatography (HPLC) method (which we used) are stable on whole blood specimens that have been stored at $-70\,^{\circ}\mathrm{C}$ for over 2 decades (Steffes et al. 2005). However, the ambient conditions and the length of time whole blood samples are stored during transport from the field to the lab have been shown to affect results (Little et al. 2007). Storage at freezing temperature has

also been related to assay value; the average frozen HbA1c was lower than fresh samples after a year, and the relative error ratio between HbA1c from frozen samples and from fresh samples was significantly higher at high levels of fresh HbA1c (Liotta et al. 2013). So the use of a conventional high risk level for HbA1c in this data may not be appropriate.

Assays on cystating C, a measure of kidney function, were done on a subsample of those who provided blood samples (N= 8,878). This was done for cost reasons, the cystatin c assay proved to be more expensive than other assays we conducted. We analyzed all respondents with blood samples aged 70 and over and a random sub-sample of those under 70. We chose to assay everyone aged 70 and over because renal insufficiency,is generally more common in older adults, who tend to have more chronic medical conditions and are taking more medications that may affect kidney functions. We use 1.44 mg/L as a cutoff for high risk cystatin C as recommended for samples ages 50 and over by the manufacturer of the assay (DakoCytomation). Doing this, the percent at high risk is 7.1% and men have higher levels of risk than women. The conventional cutoff is greater than 1.55 mg/L (Norlund et al. 1997). Using this level of cutoff would reduce the % with high risk CRP to 4.3%. Using either cutoff level, the proportion with high risk cystatin C is somewhat high compared to that of high risk serum creatinine (1.31%). The correlation between levels of creatinine and cystatin C is 0.6637 ($p<.0001$), and the correlation between the percent at high risk of the measures is 0.4094 ($p<.0001$).

In summary, we believe that the CHARLS national baseline wave has successfully collected and analyzed venous blood specimens in a large population representative sample. The distributions of bioassay data look good and should be very useful for conducting scientific analyses. CHARLS, like other HRS-type aging surveys, collects blood samples every other wave. Wave 3, in 2015, will be the next wave with blood sample collection and bioassay data.

Table 4: Descriptive Statistics of Complete Blood Count (CBC) in CHARLS (Weighted)

	N (unweighted)	Mean (SD)	Median	5% pct	25%	75%	95%	Ranges
White Blood Cell (WBC) (in thousands)	11528	6.17 (2.06)	5.90	3.8	4.9	7.1	9.5	1.14-110.80
Hemoglobin (Hb) (g/dL)	11530	14.28 (2.16)	14.20	11.2	12.9	15.4	17.6	5.44-27.88
Hematocrit (HCT) (%)	11839	41.20 (6.23)	41.30	30.4	37.5	45.0	51.0	14.10-78.40
Mean Corpuscular Volume (MCV)	11532	90.40 (8.43)	91.20	73.7	86.9	95.4	101.9	20.70-132.90
Platelets $(10^9/L)$	11531	208.71 (72.06)	203.00	102.0	160.0	252.0	330.0	10.0-1990.0

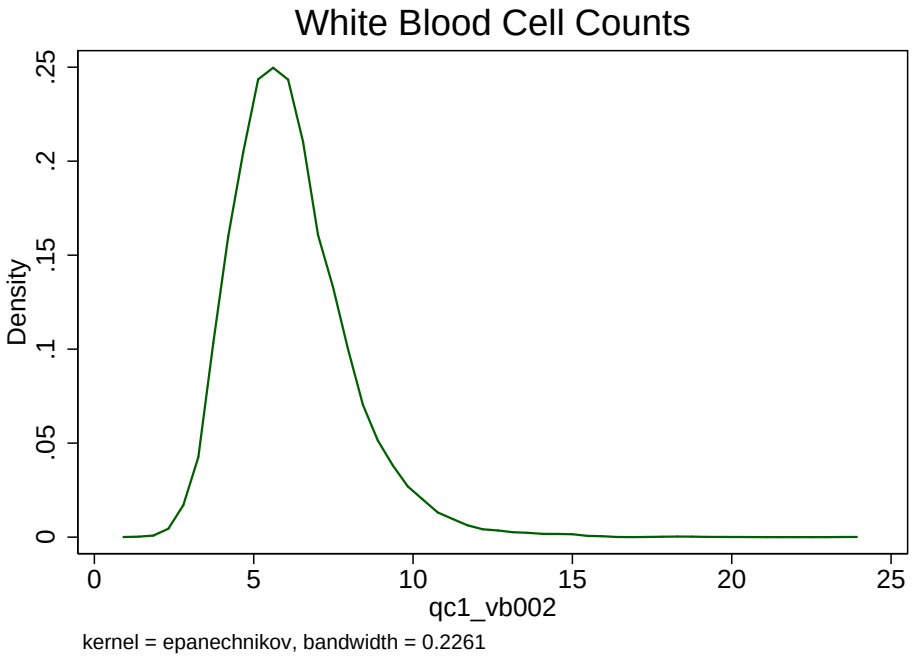
Table 5: Descriptive Statistics of Blood-based Biomarkers in CHARLS (Weighted)

	N (unweighted)	Mean (SD)	Median	5% pct	25%	75%	95%	Ranges
C-Reactive Protein (CRP) (mg/L)	11664	2.75 (7.28)	1.04	0.29	0.55	2.23	9.34	0.01-178.10
Glycated Hemoglobin (%)	11706	5.24 (0.82)	5.10	4.4	4.8	5.4	6.4	3.5-14.5
Total Cholesterol (mg/dL)	11655	190.71 (38.32)	187.89	134.54	164.69	213.40	259.02	23.97-627.07
HDL Cholesterol (mg/dL)	11663	49.57 (15.15)	47.94	28.22	39.05	58.38	76.16	3.09-158.89
LDL Cholesterol (mg/dL)	11642	115.02 (34.69)	112.89	63.40	92.01	137.24	173.58	0.39-385.83
Triglycerides (mg/dL)	11656	135.35 (108.86)	106.20	50.45	76.11	157.53	309.75	2.66-1904.52
Blood Urea Nitrogen (BUN) (mg/dL)	11662	15.71 (4.85)	15.04	9.47	12.49	18.12	23.86	3.75-89.32
Creatinine (mg/dL)	11634	0.80 (0.26)	0.77	0.54	0.66	0.89	1.13	0.18-11.55
Glucose (mg/dL)	11636	109.36 (35.26)	101.88	82.44	94.14	113.22	160.92	18.0-722.7
Uric Acid (mg/dL)	11664	4.60 (1.33)	4.44	2.76	3.66	5.32	6.98	0.26-13.31
Cystatin C	8878	1.03 (0.32)	0.98	0.68	0.85	1.15	1.52	0.36-8.75

Table 6: Percent with High Risk Levels by Gender in CHARLS (Weighted)

	Total	Men	Women
% C-Reactive Protein (CRP) >3 mg/L	18.66	20.35	17.13
% Glycated Hemoglobin >=6.5%	4.76	4.31	5.19
% Total Cholesterol >=240 mg/dL	9.97	7.65	12.12
% HDL Cholesterol <40 mg/dL	28.40	31.69	25.39
% LDL Cholesterol >160 mg/dL	9.46	7.35	11.42
% Triglycerides $\geqslant$ 200 mg/dL	15.24	14.24	16.16
% Blood Urea Nitrogen (BUN) >20 mg/dL	15.17	18.44	12.17
% Creatinine >1.4 mg/dL	1.32	2.43	0.29
% Glucose $\geqslant$ 126 mg/dL	13.45	13.99	12.95
% Low Hemoglobin (<13 g/dL for men; <12 g/dL for women)	13.65	11.62	15.52
% High Cystatin C (>1.44 mg/L)	7.09	9.18	5.21

Figure 1: Density Plots for White Blood Cell Counts and Hemoglobin (Weighted)



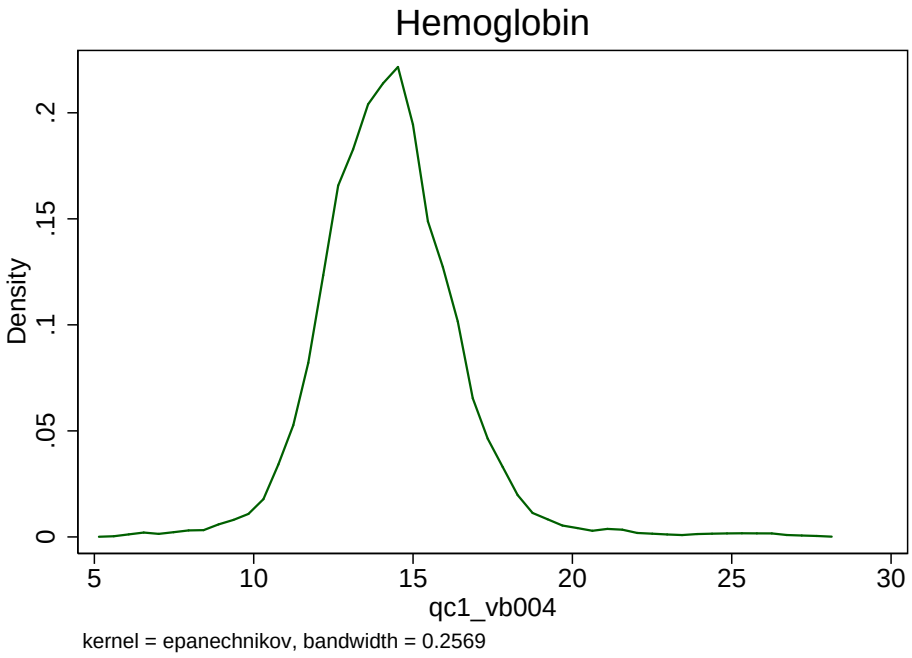


Figure 2: Density Plots for Hematocrit and MCV (Weighted)

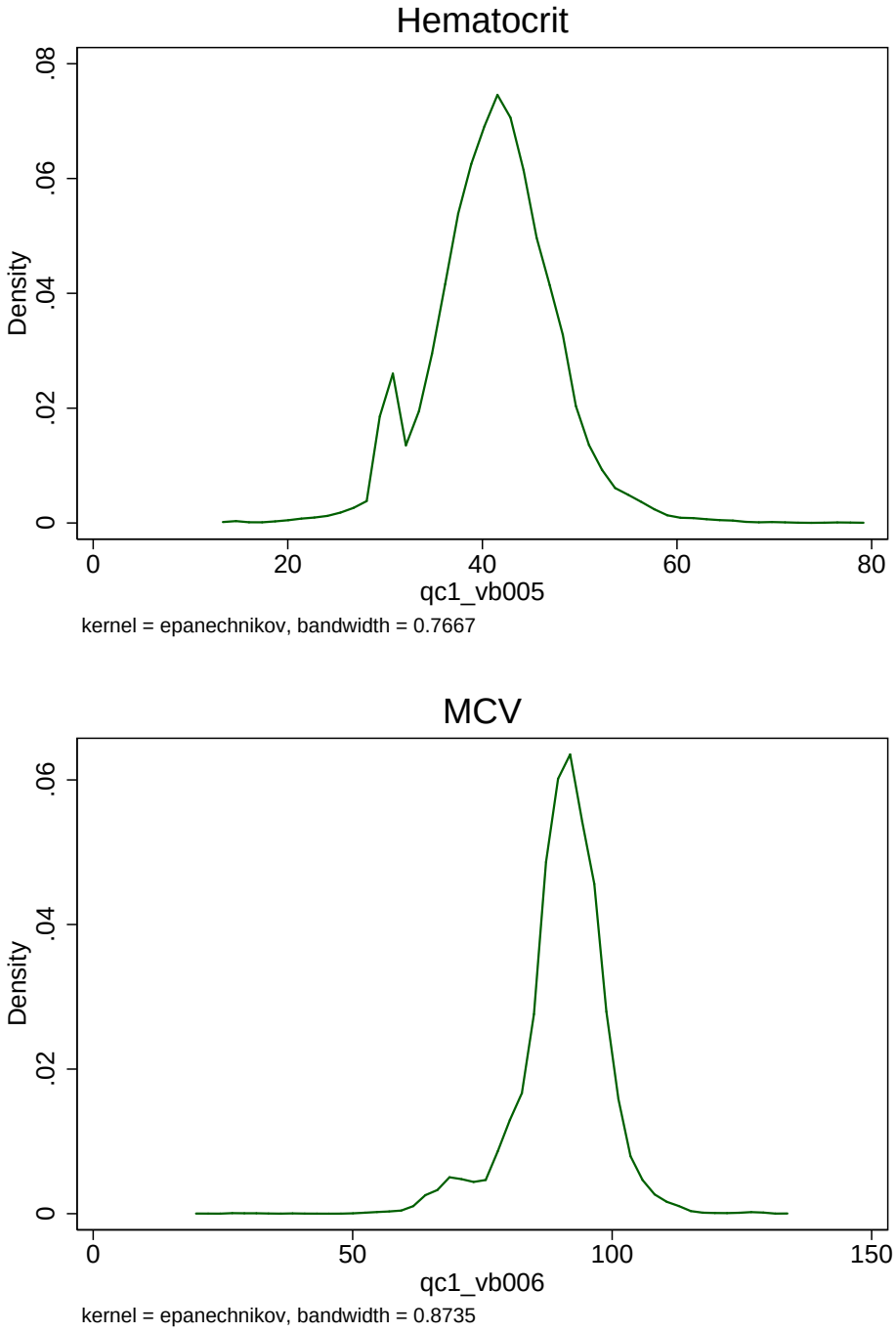


Figure 3: Density Plots for Platelets and CRP

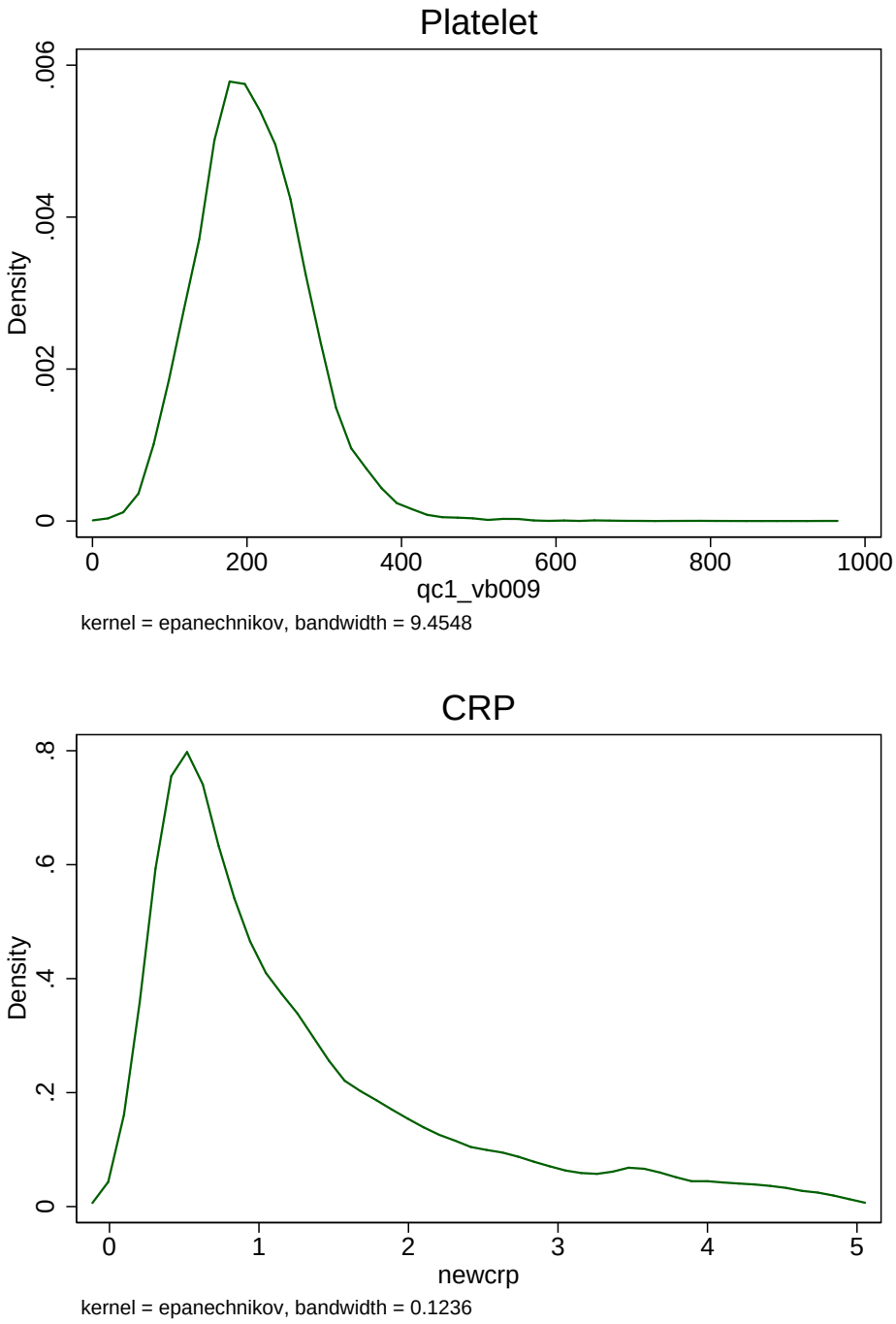
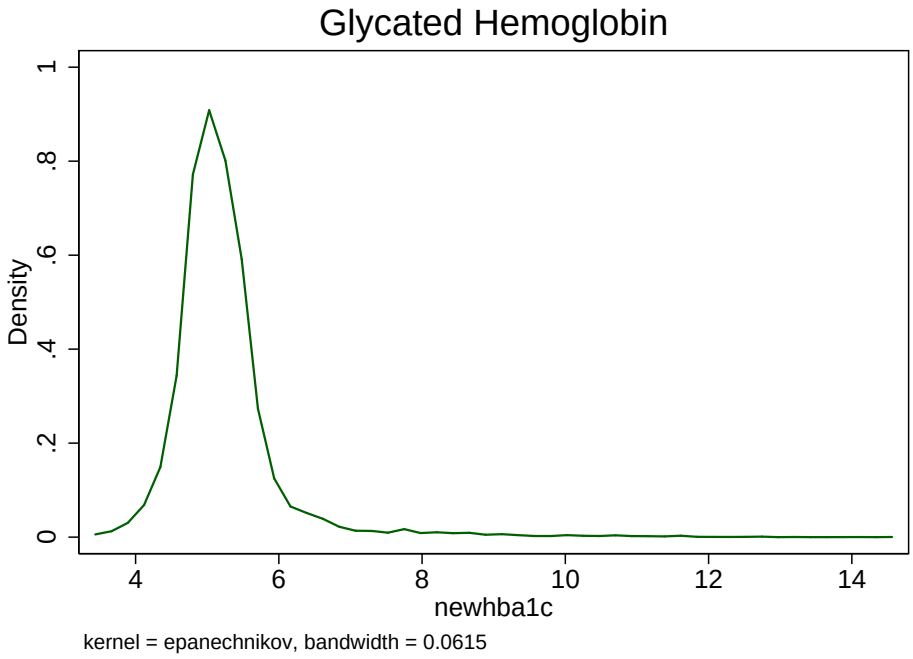


Figure 4: Density Plots for Glycated Hemoglobin and Total Cholesterol (Weighted)



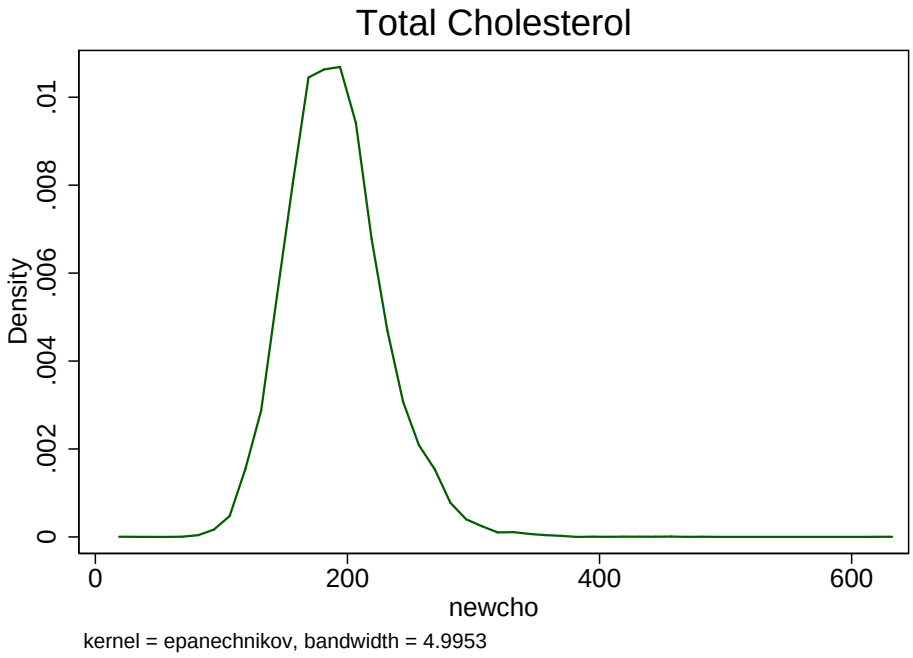


Figure 5: Density Plots for HDL and LDL Cholesterol (Weighted)

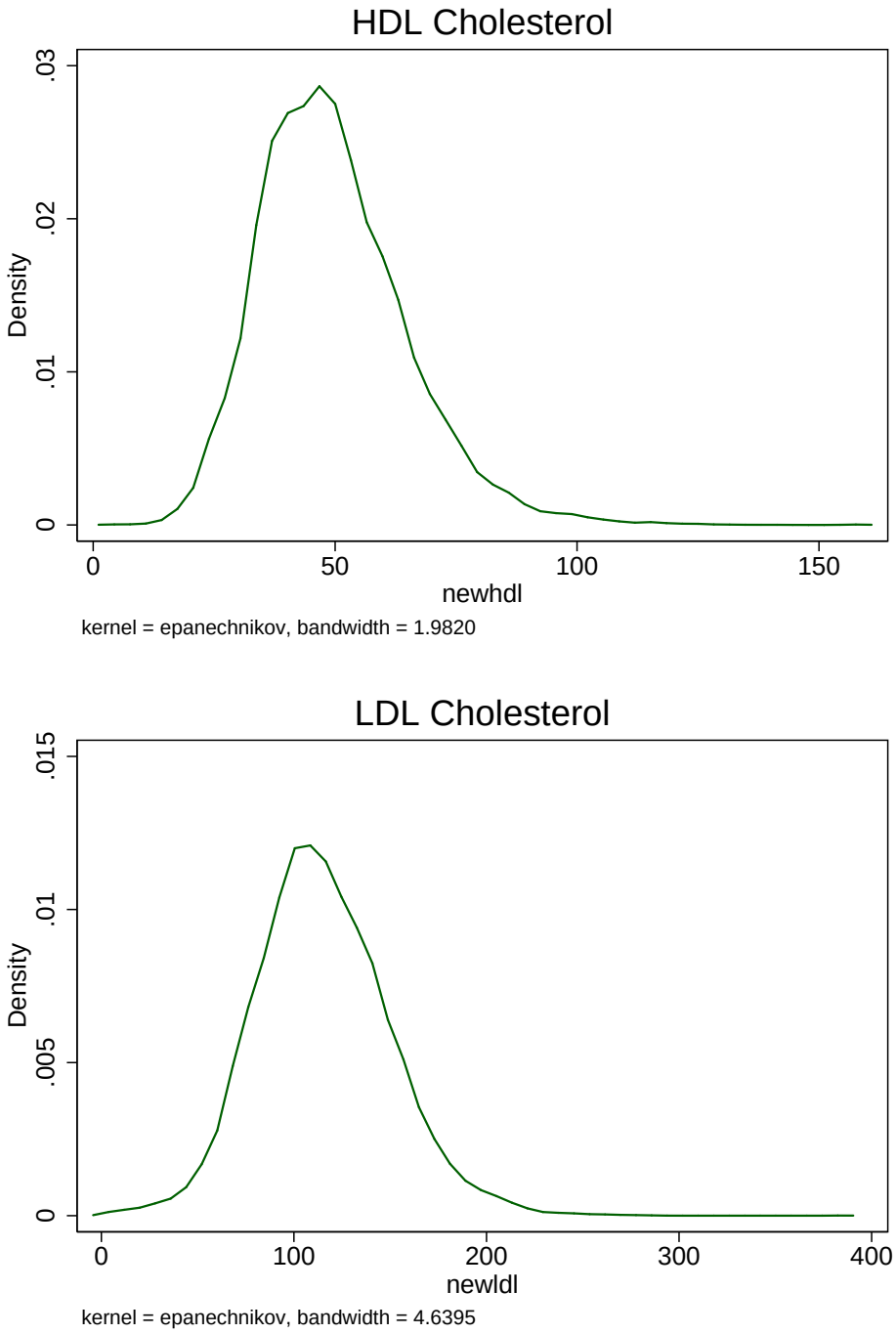


Figure 6: Density Plots for Triglycerides and BUN (Weighted)

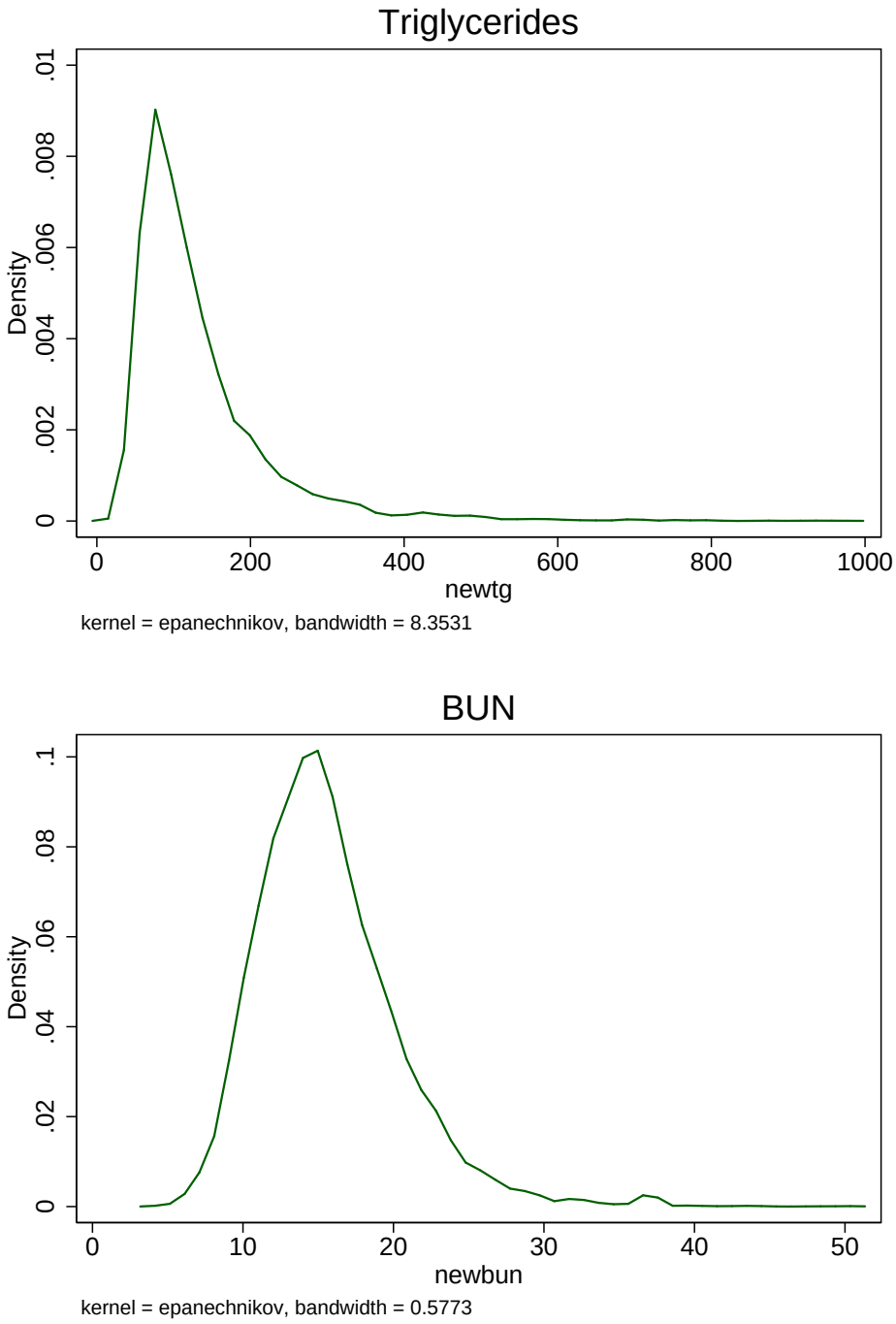


Figure 7: Density Plots for Creatinine and Glucose (Weighted)

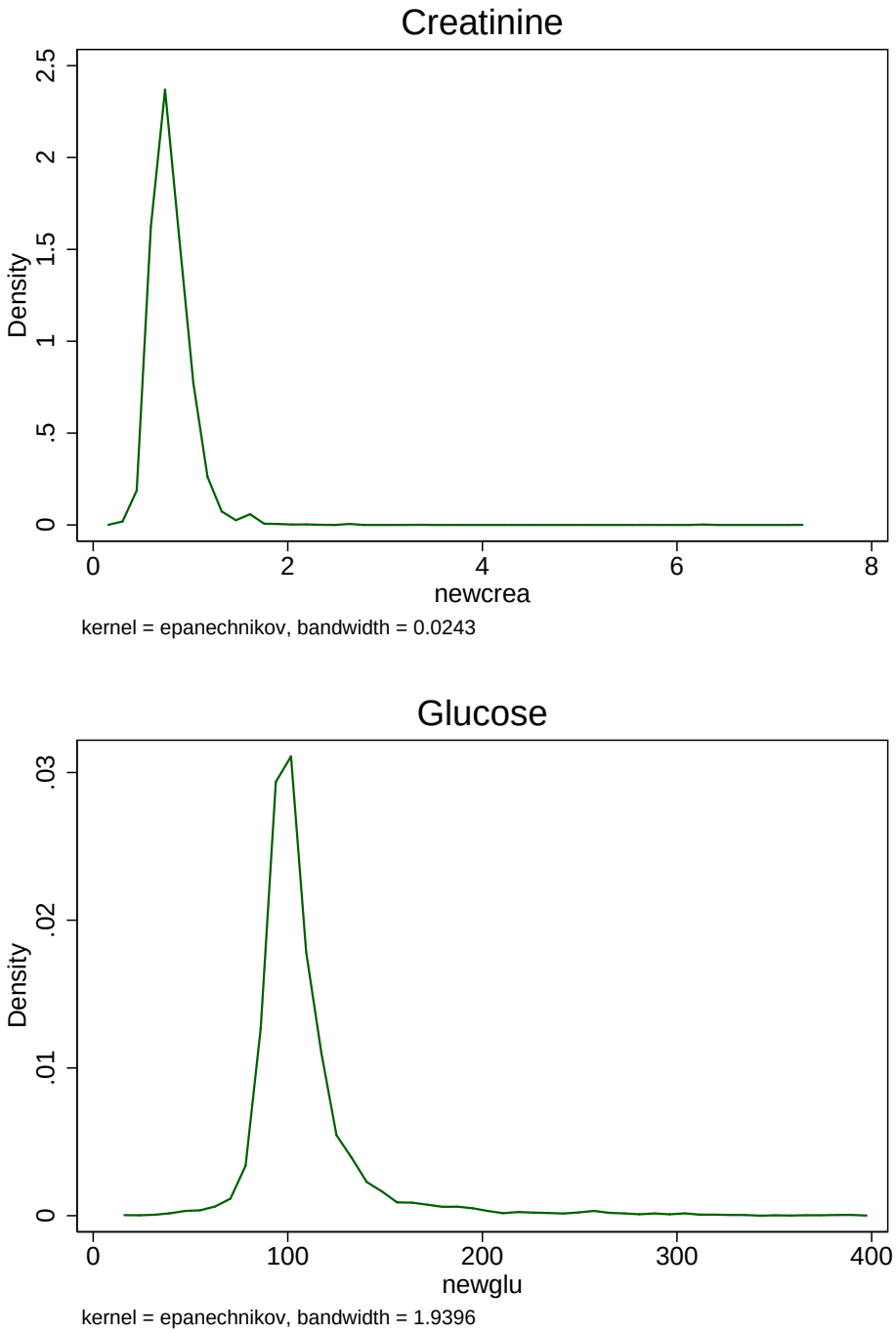


Figure 8: Density Plot for Uric Acid (Weighted)

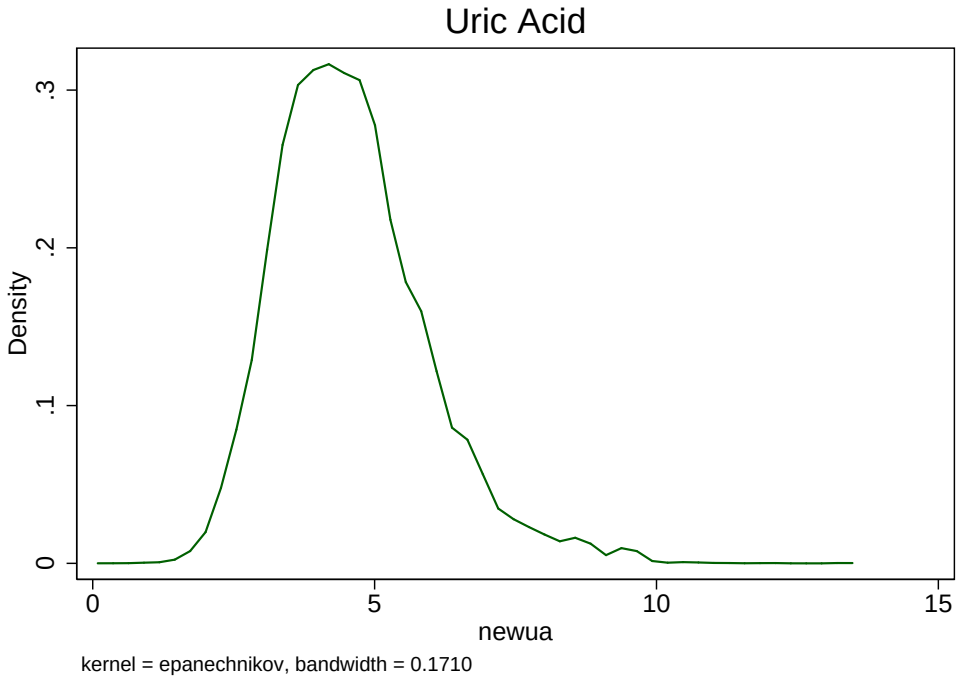
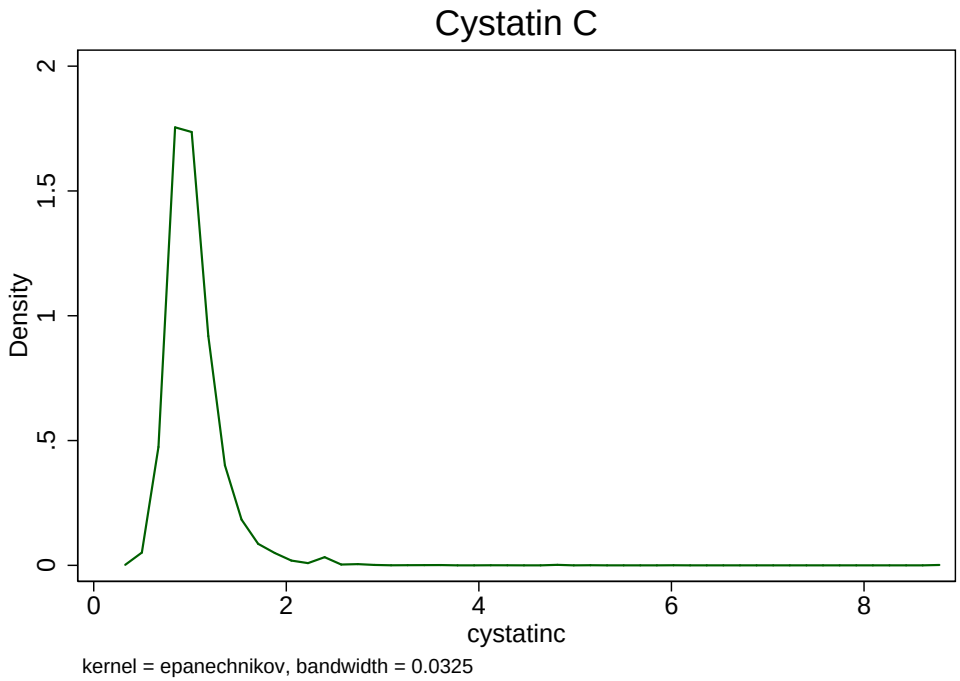


Figure 9: Density Plots for Cystatin C (Weighted)



Literature Cited

[1] Crimmins, Eileen M., and Teresa Seeman. 2001. "Integrating biology into demographic research on health and aging (with a focus on the MacArthur Study of Successful Aging)." *Cells and surveys: Should biological measures be included in social science research*, (2001):9-41.

[2] Crimmins, E. and T. Seeman. 2004. "Integrating biology into the study of health disparities." *Population and Development Review*, 30:89-107.

[3] Liotta L, Di Franco A, Pazzaqli M, Luconi M. (2013). "Glycated hemoglobin (HbA1c) measurement in frozen whole blood depends on baseline values of fresh samples." *Anal Bioanal Chem*, 405(1):429-434.

[4] Little RR, Rohlfing CL, Tennill AL, Connolly S, Hanson S. "Effects of sample storage conditions on glycated hemoglobin measurement: evaluation of five different high performance liquid chromatography methods." *Diabetes Technol Ther*. 2007 Feb; 9(1):36-42.

[5] Norlund L, Fex G, Lanke J, et al. "Reference intervals for the glomerular filtration rate and cell-proliferation markers: serum cystatin C and serum B2-microglobulin/cystatin C-ratio." *Scand J Clin Lab Invest*. 1997; 57:463-470.

[6] Steffes M, Cleary P, Goldstein D, Little R, Wiedmeyer HM, Rohlfing C, England J, Bucksa J, Nowicki M. "Hemoglobin A1c measurements over nearly two decades: sustaining comparable values throughout the Diabetes Control and Complications Trial and the Epidemiology of Diabetes Interventions and Complications study." *Clin Chem*. 2005 Apr; 51(4):753-8.

[7] Weir, D. (2008). "Elastic powers: The integration of biomarkers into the Health and Retirement Study". In Weinstein M, Vaupel J., Wachter K. (Eds), *Biosocial Surveys*. (pp.78-95). Washington, D.C.: National Academies Press.

[8] Wooldridge J. 2002. *Econometric Analysis for Cross Section and Panel Data*, Cambridge: MIT Press

[9] Zhao Y, Strauss J, Yang G, et al. (2013). China Health and Retirement Longitudinal Study 2011-2012 National Baseline Users' Guide.