nutrients

MDPI

Article

Biochemical and Hematological Correlates of Elevated Homocysteine in National Surveys and a Longitudinal Study of Urban Adults

May A. Beydoun [1,*,†], Hind A. Beydoun [2], Peter H. MacIver [1,3], Sharmin Hossain [1], Jose A. Canas [4], Michele K. Evans [1,‡] and Alan B. Zonderman [1,‡]

1 Laboratory of Epidemiology and Population Sciences, NIA/NIH/IRP, Baltimore, MD 21224, USA
2 Department of Research Programs, Fort Belvoir Community Hospital, Fort Belvoir, VA 22060, USA
3 Department of Psychology, University of Maryland Baltimore County, Catonsville, MD 21228, USA
4 Department of Pediatrics, Johns Hopkins Medical Institutions, Saint Petersburg, FL 33701, USA
* Correspondence: baydounm@mail.nih.gov; Fax: +1-410-558-8236
† M.A.B. had full access to the data used in this manuscript and completed all the statistical analyses.
‡ Co-senior authors.

Received: 12 March 2020; Accepted: 26 March 2020; Published: 30 March 2020

Abstract: Elevated blood homocysteine (Hcy) among middle-aged adults can increase age-related disease risk, possibly through other biochemical and hematological markers. We selected markers for hyperhomocysteinemia among middle-aged adults, studied time-dependent Hcy-marker associations and computed highly predictive indices of hyperhomocysteinemia, with cross-sectional and longitudinal validations. We used data from the National Health and Nutrition Examination Survey (NHANES III, phase 2, n_{max} = 4000), the NHANES 1999–2006 (n_{max} = 10,151) and pooled NHANES (cross-sectional validation). Longitudinal validation consisted of mixed-effects linear regression models (Hcy predicting markers' annual rates of change), applied to the Healthy Aging in Neighborhoods of Diversity Across the Life Span (HANDLS, n = 227–244 participants, k = 2.4 repeats/participant, Age_{base}: 30–65 years) data. Machine learning detected nine independent markers for Hcy > 14 µmol/L (NHANES III, phase 2): older age; lower folate and B-12 status; higher serum levels of creatinine, uric acid, alkaline phosphatase, and cotinine; mean cell hemoglobin and red cell distribution widths (RDW); results replicated in the 1999–2006 NHANES [AUC = 0.60–0.80]. Indices combining binary markers increased elevated Hcy odds by 6.9–7.5-fold. In HANDLS, first-visit Hcy predicted annual increase in creatinine, RDW and alkaline phosphatase, with third-visit index (2013–2018) directly predicting Hcy (2004–2009). We provide evidence of the internal and external validity of indices composed of several biomarkers that are strongly associated with elevated Hcy.

Keywords: homocysteine; hematological indices; biochemical indices; inflammation; predictive models; aging

1. Introduction

Homocysteine (Hcy) is a sulfur amino acid involved in the remethylation and transsulfuration metabolic pathways, with the first requiring folate and vitamin B-12 as coenzymes, while the second depends on a form of vitamin B6, pyridoxal 5-phosphate [1]. Epidemiological evidence indicates that elevated Hcy (>14 µmol/L) can increase risk for cardiovascular and cerebrovascular disease and may double the risk for Alzheimer's Disease (AD) [1–5]. Although biologically plausible, the causal nature of the AD–Hcy association remains a subject of debate. However, predicting Hcy from more commonly measured biochemical and hematological markers and creating a highly predictive index of elevated Hcy can be used in future cohort studies [6,7].

During one-carbon metabolism (OCM) cycles, one key enzymatic reaction involves re-methylation of Hcy, whereby a methyl group is acquired from N-5-methyl-tetrahydrofolate (MTHF) or from betaine to form methionine. While the former reaction requires folate and vitamin B12, the latter does not [8]. Adenosine triphosphate (ATP) is then used to convert methionine to S-adenosylemethionine (SAM), a universal methyl donor utilized by various acceptors including nucleic acids, hormones and neurotransmitters [8]. A methyl donation by-product, S-adenosylhomocysteine (SAH), is further hydrolyzed to regenerate Hcy, starting a new cycle of methyl group transfer [8]. The transsulfuration pathway then catabolizes excess Hcy not required for methyl donation into cysteine, using a vitamin-B-6-dependent enzyme, and cysteine is later oxidized to taurine and inorganic sulfates or excreted in urine [8]. Serum folate and vitamin B-6 and B-12 levels are strong inverse predictors of elevated blood Hcy [9–14], as are genetic polymorphisms associated with the OCM, such as MTHFR C667T, associated with reduced methylene tetrahydrofolate reductase (MTHFR) enzymatic activity [15]. Nevertheless, unexplained variability can be ascribed to kidney disease, explaining a positive association between Hcy and serum creatinine [5,16–19]. It is worth noting that men with higher muscle mass have been shown to have higher levels of both Hcy and creatinine, particularly when compared to women, given that ~70% of daily SAM-dependent methylation reactions are to produce creatine [20,21]. Hcy has been positively associated with red cell distribution width (RDW) [22]; with increased serum cotinine, a measure of active or passive recent smoking [23]; and with increased liver enzyme levels [24,25]. Other unexplored biochemical and hematological markers may also be predictive of elevated Hcy which could be reflecting other risk factors for age-related disease such as cardiovascular and neurodegenerative disorders. Generally, there is a paucity of research in the following areas: (1) discovering the most predictive measures of elevated Hcy out of selected biochemical and hematological markers; (2) creating indices that can be used as surrogates of elevated Hcy in studies which do not measure Hcy per se; (3) enhancing understanding as to why elevated Hcy may increase the risk of certain age-related diseases, including AD, by discovering novel markers that are highly predictive of elevated Hcy.

Thus, no study to date has examined and compared potential biochemical and hematological predictors of Hcy among middle-aged adults in a systematic and exploratory manner, by combining machine learning and receiver operating characteristic (ROC) techniques. This novel approach can be applied in other future studies examining the predictors of other clinical mediators of disease. Thus, our present study aimed at selecting a comprehensive yet parsimonious predictive model of elevated Hcy among middle-aged adults, using biochemical and hematological data from the third and most recent (1999–2006) National Health Nutrition Examination Survey, a model cross-validated in a longitudinal study of urban adults, from which an index reflecting elevated Hcy was also validated.

2. Materials and Methods

2.1. Databases

2.1.1. NHANES III, Phase 2 and 1999–2006

The National Health and Nutrition Examination Survey (NHANES) was conducted following guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects/patients were approved by the Institutional Review Board of the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC). Written or verbal informed consent was obtained from all participants; verbal consent was witnessed and formally recorded [26].

NHANES consists of cross-sectional surveys providing nationally representative data on the health and nutritional status of the U.S. civilian population. Initiated in the 1970s by the National Center for Health Statistics (NCHS), CDC, earlier waves of NHANES collected data in non-continuous fashion. Since 1999, NHANES became a continuous survey. The sampling design is stratified and multistage-probability-clustered. It includes an in-home interview for demographic and basic health information completed by trained staff and a health examination in a mobile examination center

(MEC), completed by physicians, medical/health technicians, and dietary and health interviewers [26]. Of interest are NHANES waves with complete blood Hcy data, namely NHANES III, phase 2 (1991–1994) [27] and the 1999–2006 wave [26]. Regulations for mandatory fortification of wheat flour with folic acid, currently in place in 53 countries, were implemented in the United States in 1998, adding 140 μg of folic acid per 100 g of enriched cereal grain product, and have been estimated to provide 100–200 μg of folic acid per day to women of childbearing age, ultimately reducing the incidence of neural tube defects [28]. This also resulted in a reduced prevalence of elevated Hcy over time and specifically between the two waves of NHANES used in this study [28].

NHANES specimen storage was consistent across waves. Upon arrival at the CDC or contract laboratories, the frozen specimens were sorted by vial type, and stored initially at -20 C. The refrigerated samples were stored at 4–8 C. Frozen specimens whose analysis might have been delayed were stored at -70 C or lower [29].

2.1.2. HANDLS 2004–2018

HANDLS is an ongoing prospective cohort study initiated in 2004. It focuses primarily on disparities in the cardiovascular and cognitive health of a socioeconomically diverse sample of Whites and African Americans aged 30–65 yo at baseline and living in selected neighborhoods of Baltimore, Maryland. In brief, HANDLS used an area probability sampling strategy of thirteen neighborhoods, with details provided elsewhere [30]. Phase 1 of Visit 1 (2004–2009) consisted of screening followed by recruitment, household interviews, while phase 2 of Visit 1 (also 2004–2009) consisted of in-depth examinations in a mobile Medical Research Vehicle (MRV), including measurements of blood pressure; anthropometrics and a fasting blood draw were also collected at the follow-up visits [Visit 2: 2009–2013; Visit 3: 2013–2018]. Although blood Hcy was measured only in a small subset of Visit 1 participants (i.e., at baseline), all other available hematological and biochemical indices had three repeats at Visits 1, 2 and 3 (2004–2009, 2009–2013 and 2013–2018). All clinical laboratory indices were obtained from Quest Diagnostics (Chantilly, VA). Mean follow-up times between visits ranged between 6 months and 8 years, with an average of 4–5 years.

Participants provided written informed consent after reviewing a protocol booklet written in layman's terms and watching a video detailing all procedures and future re-contacts. The HANDLS study was approved ethically by the Institutional Review Board of the National Institutes of Health, National Institute of Environmental Health Sciences (NIEHS/NIH).

2.2. Study Samples

We selected adults aged 30–65 years from the NHANES III (phase 2) and from the 1999–2006 waves. Similarly, by design, Visit 1 of HANDLS consisted of adults aged 30–65 years (Supplementary Figure S1). In the NHANES III, phase 2, biomarkers with >20% missing data compared to the sub-sample with complete Hcy measures were excluded. Out of 15,283 participants from phase 2 of NHANES III, 8585 had complete Hcy data, of whom 4008 were in the age range of interest. Of those, 3709–4000 had complete data on up to 82 biochemical and hematological markers. Similarly, for the NHANES 1999–2006, of an initial 41,474 participants, completeness on Hcy data was found for n = 28,449, of whom 10,151 were aged 30–65 years and the final analytical sample ranged between 9991 and 10,151, after biomarkers were selected with machine learning methods using NHANES III, phase 2 data. Pooling data from NHANES III, phase 2 and NHANES 1999–2006, 14,739–14,829 provided complete data on Hcy and the selected biochemical and hematological indices, within the age range 30–65 years. Finally, out of 3720 HANDLS participants, only 245 individuals had complete data on Hcy measured during the first MRV visit. All these participants had the target Visit 1 age of 30–65 years, and the final analytic samples for longitudinal analysis ranged between 227 and 244 individuals with multiple repeats (up to 3, mean repeats/participants, k = 2.4), depending on adjustment levels.

2.3. Serum Homocysteine

In NHANES III, phase 2, serum Hcy was measured at the Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, using the high-performance liquid chromatography method of Araki and Sako [31]. In the recent NHANES, serum Hcy was measured using "Abbott Homocysteine (HCY) assay", a fully automated technique [32,33], as was Hcy in the HANDLS sub-cohort. In all datasets, elevated Hcy was defined as >14 μmol/L, a cut-point of 2.639057 on the Log_e-transformed scale used to examine the association between elevated Hcy and AD in most previous studies [1–5].

2.4. Biochemical and Hematological Indices

Biochemical indices included nutritional biomarkers (e.g., folate, B-12, vitamin D, vitamin E, carotenoids, retinol, vitamin C, total calcium, iron, sodium, potassium etc.), metabolic parameters (e.g., serum insulin, glucose, cholesterol, triglycerides, creatinine, albumin, thyroid hormones, liver enzymes) and inflammatory markers (e.g., C-reactive proteins, Immunoglobulin G (IgG) against specific viruses and bacteria), and environmental indices of air pollution and smoking (e.g., blood lead and serum cotinine).

As stated earlier, lower serum folate and vitamin B-12 concentrations are among the highly predictive markers of elevated Hcy. In NHANES III, phase 2, serum folate and B-12 were measured using Bio-Rad Laboratories "Quantaphase Folate" radioassay kit [29,34], as was the case for more recent NHANES [35,36]. In HANDLS, these two measures were determined using enzyme immunoassay by Quest Diagnostics, Chantilly, VA [37], at Visits 1 through 3. Hematological indices consisted of markers of blood cell counts and characteristics (Supplemental methods 1–3).

2.5. Covariates

In all NHANES predictive models, the following covariates were considered: age, sex, race/ethnicity (1: NH white, 2: NH black, 3: Mexican American, 4: other Hispanic, 5: others), poverty status (0: >125% of poverty income ratio; 1: ≤125% of poverty income ratio), rural vs. urban area of residence, and region (Northeast, Midwest, South and West). Given that the last two factors are fixed in HANDLS (urban, Northeast: Baltimore city), only age, sex, race (African American vs. Whites) and poverty status were included, using a similar cut-point ≤125% of the federal poverty line.

2.6. Data Handling and Statistical Analysis

All analyses were conducted using Stata release 16.0 [38]. We first describe study characteristics [covariates (all datasets, waves); Log_e-transformed Hcy; Log_e-transformed biochemical and hematological indices (NHANES III, phase 2)], overall and by categorical Hcy (≤14 μmol/L vs. >14 μmol/L). Differences in means and proportions across these categories were tested using a design-based F-test accounting for sampling design complexity. Beyond this descriptive step, a multi-stage approach was implemented to select the key predictors of elevated Hcy in NHANES III, phase 2, validate them against the most recent NHANES, and cross-validate those predictors in a longitudinal study of urban adults (HANDLS). A flow diagram is used to summarize this approach (Figure 1).

To select predictive biomarkers of continuous Hcy (Log_e-transformed, z-scored), a statistical learning method known as least absolute shrinkage and selection operator (LASSO) was used. LASSO is a covariate selection methodology that is superior to both generalized linear models without covariate selection and the usually applied stepwise or backward elimination process [39]. In fact,

stepwise selection is often trapped into a local optimal solution and backward elimination can be time-consuming [39]. The LASSO, which does not ignore stochastic errors, is defined as follows:

$$\beta(\text{lasso}) = \text{arg}min_{\beta}\|\text{y} - \sum_{j=1}^{p} x_j\beta_j\|^2 + \lambda \sum_{j=1}^{p}|\beta_j| \quad (1)$$

with λ being a non-negative regularization parameter [39]. The second term of the equation, termed the "l1 penalty", is a key portion of this equation, ensuring the success of the LASSO method of covariate selection [39]. In fact, this method was shown to discover the right sparse representation of the model, given certain conditions [39]. More recently, several related methods have been developed and validated against each other, with an adaptive LASSO giving more consistent findings, particularly when compared with the non-negative garotte [39].

In our predictions, we used this convex optimization technique with an l_1 constraint, known as adaptive LASSO, as the main method to select the final linear regression model for prediction of Log_e-transformed Hcy with Log_e-transformed and z-scored biomarkers and socio-demographic factors, with the latter being force entered into all models. The model was trained on a random half sample of the total population (among the target age group: 30–65 years, sorting the sample by individual ID and fixing a random seed) and validated against the other half sample to check robustness of findings, by comparing R^2 between samples. Adaptive LASSO robustness is then compared to that of cross-validation (cvLASSO) and minimal Bayesian information criterion (minBIC) LASSO, and non-zero parameters were presented for each method. This parsimonious model, with Log_e-transformed Hcy as an outcome, was then run on the entire population accounting for survey design complexity (i.e., svy: reg) as a starting point for further backward elimination. Thus, beyond that point, additional terms were eliminated at a type I error of 0.10. This final model was applied to the binary outcome of elevated blood Hcy (>14 μmol/L), using svy: logit, and further backward elimination was carried out to obtain a short list of independent predictors for elevated Hcy. As a sensitivity analysis, additional markers identified with adaptive LASSO logistic regression on the same half-sample as for the adaptive LASSO linear regression were included in the reduced model to test their predictive value.

In the full NHANES III, phase 2, the selected Log_e-transformed biomarkers and continuous socio-demographic variables (e.g., age) in the final models were then entered into a series of ROC analyses to determine the most appropriate cut-point, which would have the largest sensitivity and specificity in predicting elevated Hcy (>14 μmol/L). Sensitivity (proportion of true positives, i.e., proportion of cases correctly identified as meeting the conditions of elevated Hcy) and specificity (proportion of true negatives, i.e., proportion of non-cases correctly identified as not meeting elevated Hcy) were calculated to evaluate accuracy of selected biochemical and hematological markers in depicting elevated Hcy, creating a series of ROC curves [40,41]. The ROC curve is a graphical plot of sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied. The area underneath of each ROC curve (AUC), a measure that is independent of classifier cut-points, can range between 0 and 1 and be computed with its 95% confidence interval (95% CI).

Step 1:Determine availability of Hcy	Step 2: Select out markers that have <20% missing on Hcy	Step 3: Determine sample sizes for complete Hcy and age range: 30-65y	Step 4: Descriptive by elevated Hcy (>14μmol/L) in age range: 30-65y	Step 5: Split sample into training and testing set for LASSO analysis.	Step 6: Backward elimination	Step 7: ROC curve analysis	Step 8: Indices and final logistic regression model	Step 9: Run mixed effects linear regression and LOWESS
NHANES III, phase 2, 1991-1994: >100 biomarkers N=8,585 (all ages: 1-90y)	**NHANES III, phase 2, 1991-1994:** 82 biomarkers of >100	**NHANES III, phase 2, 1991-1994:** 82 biomarkers N=3,709-4,000 (30-65y)	**NHANES III, phase 2, 1991-1994:** Socio-demographics and 82 biomarkers [Note: Hcy and biomarkers are Loge transformed]	**NHANES III, phase 2, 1991-1994:** • Run cvLASSO linear regression on training half sample. • Run minBIC LASSO on training half sample. • Run adaptive LASSO on training half sample. • Compare R2 between training and test sample. • List nonzero terms in descending order of effect size for each LASSO method. • Run adaptive LASSO logistic as sensitivity analysis.	**NHANES III, phase 2, 1991-1994:** • Run a linear model with Loge transformed Hcy as outcome and predictors selected from LASSO linear (adaptive method) • Use entire eligible sample and adjust for sampling design complexity. • Starting from this model, eliminate terms with p>0.10, using backward elimination. • Use the reduced model for Loge(Hcy) as starting point to eliminate the model further with logistic regression on binary elevated Hcy. • Determine final predictors for elevated Hcy for the next step.	**NHANES III, phase 2, 1991-1994:** • Run ROC curve analysis for Hcy>14 and compare areas under the curve between predictors selected in Step 6. • Determine if AUC>0.55. • Select those with AUC>0.55 as final predictors. • Present ROC curve, AUC and optimal cut-point with highest sensitivity and specificity. • Select only the common predictors that available in NHANES 1999-2006.	**Pooled NHANES III, phase and 1999-2006**: • Create several indices combining binary predictors available in the pooled sample and/or in the HANDLS sub-cohort. • Run ROC curve analysis for Hcy>14 to determine optimal cutoff of these ordinal indices. • Run multiple logistic regression model for Hcy>14 vs. selected binary predictors and another set of models with binary indices.	**HANDLS sub-cohort**: • Mixed models with outcome being Hcy and predictors being continuous predictors selected in steps 7-8. • Adjust for other covariates, including biomarkers, in a serial manner. • Determine association between visit 1 Hcy and visit 1 markers; and visit 1 Hcy vs. change in these markers over time. • Sample sizes N=227-244, k=2.4 repeats/participant. • Determine association between visit 1 Hcy and ordinal index at visit 3 using LOWESS and Pearson's correlation, r (N=81)
NHANES 1999-2006: N=28,449 (all ages: 0-85y)	**NHANES 1999-2006:** >100 biomarkers with commonalities with NHANES III, phase 2	**NHANES 1999-2006:** N=10,151 (30-65y)	**NHANES, 1999-2006**: Socio-demographics			**NHANES, 1999-2006**: • Run ROC curve analysis using the predictors selected in Step 7 for NHANES III, phase 2. • Compare optimal cut-points and AUC.	The results of this analysis can used in other cohorts to create indices for elevated Hcy, with high sensitivity and specificity.	The results of this analysis is a cross-validation of the results from the national surveys.
HANDLS sub-cohort: visit 1 (Hcy) N=245 (30-65y)	**HANDLS sub-cohort**: visits 1-3: >100 biomarkers with commonalities with NHANES III, phase 2.	**HANDLS sub-cohort**: N=245 (30-65y)	**HANDLS sub-cohort, visit 1, 2004-2009**: Socio-demographics			**Pooled NHANES III, phase and 1999-2006**: • Re-run ROC curve analysis on the pooled sample of NHANES III, phase and 1999-2006, with age: 30-65y and complete Hcy.		

Figure 1. Flow diagram of predictive modeling using LASSO, ROC curves and multivariable regression modeling. Abbreviations: cvLASSO = cross-validation LASSO; LASSO = least absolute shrinkage and selection operator; LOWESS = Locally weighted regression; HANDLS = Healthy Aging in Neighborhoods of Diversity Across the Life Span; minBIC LASSO = minimum Bayesian information criterion LASSO; NHANES = National Health and Nutrition Examination Surveys; R2 = coefficient of determination.

A Log_e-transformed biomarker positively associated with elevated Hcy would yield an AUC between 0.5 and 1.0. An area of 0.70, for instance, has the following interpretation: if we randomly select a biomarker from the Hcy^+ and Hcy^- groups, the value of that biomarker will be greater in the Hcy^+ group than in the Hcy^- group, 70% of the time. The ROC curves and their associated AUC are presented, with biomarkers inversely linked to elevated Hcy having their values inverted (i.e., multiplying them by −1). Subsequently, biomarkers retained in the final logistic regression model (selected with machine learning and backward elimination), were further pruned out when ROC AUC was <0.55. Thus, only biomarkers with AUC ≥ 0.55 were retained and their AUC and optimal cut-points presented for NHANES III, phase 2, validated against NHANES 1999–2006 and presented for the pooled NHANES.

An index combining all selected binary biomarkers was computed with a potential to range between 0 and m' [number of selected biomarkers with positive criterion: ≥optimal cut-point]. The final index summing categorical biomarkers was computed using revised cut-points from pooled NHANES data. Two ordinal indices were obtained, namely, Index I summing all selected binary biomarkers reflecting elevated Hcy that were available in all selected NHANES waves, and Index II, sub-set of Index I, using only commonly measured biomarkers available in the HANDLS study. Similarly, a ROC analysis was conducted on the pooled NHANES to determine the optimal cut-point for Indices I and II. To determine potential use of those biomarkers as surrogates for elevated Hcy in other large epidemiological studies, a logistic regression model was conducted with each Index and with Index I components entered simultaneously in the pooled NHANES data.

Importantly, the cross-sectional and longitudinal associations of Visit 1 Hcy (2004–2009) on repeated measures of the biochemical and hematological markers (Visits 1 through 3: 2004–2018) that were selected for NHANES (continuous) were tested in a sub-set of HANDLS using multiple mixed-effects linear regression analysis (See Supplemental methods 4). All models were adjusted for Visit 1 age (Model 1), with further adjustment for sex, race and poverty status applied to Model 2, while the full model (Model 3) additionally adjusted for all remaining biomarkers. Finally, Index II, computed using the NHANES cut-point, was computed at Visit 3 and correlated with Hcy at Visit 1 of HANDLS using Pearson's correlation and locally weighted regression (LOWESS) smoother, to assess external validity of the association. Type I error was set at 0.05 with $p < 0.10$ considered as a trend or tendency towards an association.

3. Results

Table 1 present study sample characteristic distributions, namely Hcy and socio-demographic factors for both NHANES waves and for HANDLS Visit 1, as well as biochemical and hematological correlates for NHANES III, phase 2, overall and stratified by Hcy status. In all samples, age and sex (men vs. women) were consistently associated with elevated Hcy, while poverty was directly associated with elevated Hcy only in NHANES 1999–2006, and both race/ethnicity and poverty status trended towards an association with this binary outcome in the HANDLS sub-cohort. Numerous biochemical and hematological indices were significantly associated with elevated Hcy, including serum cotinine, mean cell hemoglobin (MCH), RDW, blood lead, serum uric acid (SUA), serum creatinine, serum alkaline phosphatase, while others were inversely linked to elevated Hcy, namely serum vitamin E, most serum carotenoids, serum retinyl esters, serum folate and vitamin B-12 ($p < 0.05$). It is worth noting that those associations are crude, not adjusted for socio-demographic factors such as age and sex. Among the known predictive factors, serum folate was shown to have an unadjusted mean of 7.6 in the elevated Hcy group compared to 13.6 in the normal group, suggesting a strong inverse relationship.

Table 1. Participant characteristics distribution by hyperhomocysteinemia status: NHANES III, phase 2; NHANES 1999–2006 and HANDLS 2004–2018 [1].

	Overall				**Hcy ≤ 14 μmol/L**				**Hcy > 14 μmol/L**				P_{Hcy}
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	
NHANES III, phase 2: 1991–94	4000		100.0	0.0	3663		91.5	0.7	337		8.4	0.7	
Age (y)	4000	44.3		0.42	3663	44.0		0.45	337	47.1		0.76	0.001
Sex, % men	4000		49.0	1.2	3663		47.8	1.2			62.8	4.9	0.008
Race/ethnicity	4000				3663				337				0.51
NH white			74.8	2.1			74.6	2.2			75.8	3.7	
NH black			11.1	1.0			10.9	1.1			13.0	1.9	
MA			5.2	0.7			5.4	0.7			3.7	0.7	
Other			8.8	1.6			9.0	1.6			7.6	2.6	
Poverty status													
PIR ≥ 125%	3714		14.0	1.9	3388		14.0	1.9	326		14.6	2.3	0.79
PIR < 125%													
Region	4000		18.9	2.2	3663		18.4	2.5	337		23.8	4.8	0.30
Northeast			22.6	4.4			22.8	4.5			20.7	4.4	
Midwest			36.0	7.3			35.8	7.4			36.9	7.6	
South			22.6	7.5			22.9	7.6			18.3	6.9	
West													
Urban/Rural	4000				3663				337				0.59
Urban			50.5	7.5			50.7	7.6			48.2	8.2	
Rural			49.6	7.5			49.3	7.6			51.8	8.2	
Hcy, Log_e	4000	+2.17		0.01	3663	+2.10		0.01	337	+2.93		0.04	<0.001
Selected biochemical and hematological indices, Log_e		**Mean, Log_e**	**Mean, exp**	**SE, Log_e**		**Mean, Log_e**	**Mean, exp**	**SE, Log_e**		**Mean, Log_e**	**Mean, exp**	**SE, Log_e**	
Serum cotinine, ng/mL	3966	+0.33	1.39	0.12	3630	+0.17	1.190	0.11	336	+2.08	8.0	0.27	<0.001
Serum vitamin D, nmol/L	3997	+4.21	67.4	0.02	3660	+4.21	67.40	0.02	337	+4.11	60.9	0.04	0.015
Serum thyroxine, nmol/L	3997	+4.7	109.9	0.01	3660	+4.70	109.9	0.01	337	+4.63	102.5	0.03	0.12
Serum TSH, mU/L	3925	+0.42	1.52	0.03	3594	+0.42	1.522	0.03	331	+0.41	1.506	0.06	0.94
Serum antimicrosomal Ab, U/mL	3927	−0.61	0.54	0.06	3596	−0.61	0.543	0.06	331	−0.55	0.576	0.18	0.73
Serum anti-thyroglobulin Ab, U/mL	3927	−0.06	0.94	0.04	3596	−0.06	0.942	0.04	331	−0.03	0.970	0.09	0.76
White blood cell count	3998	+1.94	6.96	0.01	3661	+1.93	6.890	0.01	337	+2.01	7.463	0.03	0.015
Lymphocyte percent	3998	+3.46	31.81	0.01	3661	+3.47	32.14	0.01	337	+3.40	29.96	0.03	0.026
Mononuclear percent	3920	+1.67	5.310	0.04	3584	+1.67	5.312	0.04	336	+1.69	5.419	0.05	0.77
Granulocyte percent	3920	+4.10	60.34	0.01	3584	+4.10	60.34	0.01	336	+4.13	62.17	0.02	0.12
Lymphocyte number	3998	+0.80	2.23	0.01	3661	+0.79	2.20	0.01	337	+0.81	2.247	0.03	0.67
Mononuclear number	3905	−0.99	0.370	0.03	3572	−1.00	0.368	0.03	333	−0.91	0.402	0.05	0.039

Table 1. *Cont.*

	Overall				Hcy ≤ 14 µmol/L				Hcy > 14 µmol/L				P_{Hcy}
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	
Granulocyte number	3920	+1.43	4.18	0.02	3584	+1.43	4.18	0.02	336	+1.53	4.618	0.03	0.017
Red blood cell count, SI	3997	+1.55	4.71	0.00	3660	+1.55	4.71	0.00	337	+1.55	4.711	0.00	0.74
Hemoglobin, g/L	3998	+4.96	142.5	0.00	3661	+4.95	141.2	0.00	337	+4.97	144.0	0.01	0.001
Hematocrit, L/L = 1	3997	−0.87	0.420	0.00	3660	−0.87	0.419	0.00	337	−0.85	0.427	0.01	0.001
Mean cell volume, fL	3998	+4.49	89.12	0.00	3661	+4.49	89.12	0.00	337	+4.51	90.92	0.00	<0.001
Mean cell hemoglobin, pg	3997	+3.41	30.3	0.00	3660	+3.41	30.27	0.00	337	+3.43	30.87	0.00	<0.001
Mean cell hemoglobin conc., SI	3997	+5.82	337.0	0.00	3660	+5.82	337.0	0.00	337	+5.83	340.3	0.00	0.30
Red cell distribution width, %	3998	−2.05	0.130	0.00	3661	−2.05	0.129	0.00	337	−2.02	0.132	0.00	<0.001
Platelet count: SI	3998	+5.54	254.7	0.01	3661	+5.53	252.1	0.01	337	+5.54	254.7	0.02	0.87
Platelet distribution width, %	3973	+2.81	16.61	0.00	3640	+2.80	16.44	0.00	333	+2.80	16.44	0.00	0.94
Mean platelet volume, fL	3997	+2.13	8.41	0.00	3660	+2.13	8.41	0.00	337	+2.12	8.331	0.01	0.30
Lead, µmol/L	3999	−2.13	0.118	0.03	3662	−2.15	0.116	0.04	337	−1.82	0.162	0.06	<0.001
Erythrocyte protoporphyrin, SI	3999	−0.19	0.83	0.01	3662	−0.18	0.84	0.01	337	−0.25	0.779	0.03	0.029
Serum iron, µmol/L	4000	+2.71	15.02	0.01	3663	+2.71	15.03	0.01	337	+2.73	15.33	0.04	0.67
Serum TIBC, µmol/L	3997	+4.16	64.07	0.01	3660	+4.15	63.43	0.01	337	+4.18	65.37	0.01	0.067
Serum ferritin, µmol/L	3998	+4.43	83.93	0.03	3661	+4.41	82.27	0.03	337	+4.55	94.63	0.09	0.14
Serum folate, nmol/L	4000	+2.57	13.07	0.04	3663	+2.61	13.60	0.04	337	+2.03	7.614	0.07	<0.001
RBC folate, nmol/L	3952	+6.03	415.7	0.03	3615	+6.05	424.1	0.02	337	+5.73	307.9	0.06	<0.001
Serum vitamin B-12, pmol/L	3999	+5.79	327.0	0.01	3662	+5.81	333.6	0.01	337	+5.52	249.6	0.04	<0.001
Serum vitamin C, nmol/L	3841	+3.50	33.11	0.04	3510	+3.54	34.47	0.04	331	+3.11	22.42	0.09	<0.001
Serum normalized calcium, mmol/L	3709	+0.21	1.233	0.00	3410	+0.21	1.234	0.00	308	+0.21	1.234	0.00	0.55
Serum total calcium, nmol/L	3993	+0.84	2.316	0.00	3657	+0.84	2.316	0.00	336	+0.84	2.314	0.00	0.029
Serum selenium, nmol/L	3977	+0.47	1.599	0.01	3642	+0.47	1.600	0.01	335	+0.51	1.665	0.02	0.004
Serum vitamin A, µmol/L	3993	+0.66	1.934	0.01	3656	+0.66	1.935	0.01	337	+0.67	1.954	0.02	0.57
Serum vitamin E, µmol/L	3993	+3.26	26.05	0.01	3656	+3.27	26.31	0.01	337	+3.16	23.571	0.02	<0.001
Serum alpha carotene, µmol/L	3993	−2.62	0.073	0.028	3622	−2.60	0.074	0.030	322	−2.87	0.057	0.08	0.003
Serum beta carotene, µmol/L	3991	−1.24	0.289	0.02	3656	−1.22	0.295	0.02	335	−1.56	0.210	0.07	<0.001
Serum beta-cryptoxanthin, µmol/L	3991	−1.94	0.143	0.02	3655	−1.92	0.147	0.02	336	−2.17	0.114	0.05	<0.001
Serum lutein/zeaxanthin, µmol/L	3992	−1.07	0.343	0.01	3656	−1.07	0.343	0.01	336	−1.15	0.317	0.03	0.003
Serum lycopene, µmol/L	3990	−0.91	0.402	0.02	3655	−0.90	0.407	0.02	335	−1.00	0.368	0.03	0.012
Serum retinyl esters, µmol/L	3974	−1.78	0.169	0.02	3641	−1.75	0.174	0.02	333	−2.09	0.124	0.05	<0.001

Table 1. *Cont.*

	Overall				Hcy ≤ 14 μmol/L				Hcy > 14 μmol/L				P_{Hcy}
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	
Serum cholesterol, mmol/L	3994	+1.65	5.206	0.01	3657	+1.65	5.207	0.01	337	+1.66	5.259	0.02	0.56
Serum triglycerides, mmol/L	3994	+0.33	1.391	0.02	3657	+0.32	1.377	0.02	337	+0.39	1.476	0.05	0.30
Serum HDL-cholesterol, mmol/L	3972	+0.20	1.221	0.01	3640	+0.21	1.234	0.01	332	+0.17	1.185	0.02	0.12
Serum C-reactive protein, mg/dL	3983	−1.20	0.301	0.02	3646	−1.21	0.298	0.02	337	−1.19	0.304	0.04	0.65
Serum hepatitis A Ab	4000	+0.47	1.600	0.01	3663	+0.47	1.600	0.01	337	+0.48	1.616	0.03	0.67
Serum hepatitis B core Ab	4000	+0.65	1.915	0.00	3663	+0.65	1.916	0.00	337	+0.64	1.896	0.02	0.66
Serum hepatitis C Ab	4000	+0.67	1.954	0.00	3663	+0.68	1.974	0.00	337	+0.67	1.954	0.01	0.46
Serum rubella Ab, IU	3885	+4.31	74.44	0.04	3555	+4.31	74.44	0.04	330	+4.26	70.81	0.13	0.74
Serum sodium, mmol/L	3997	+4.95	141.1	0.00	3641	+4.94	139.77	0.00	336	+4.95	141.1	0.00	0.81
Serum potassium, mmol/L	3977	+1.41	4.095	0.00	3641	+1.41	4.096	0.00	336	+1.40	4.055	0.01	0.30
Serum chloride, mmol/L	3977	+4.64	103.5	0.00	3641	+4.64	103.5	0.00	336	+4.64	103.5	0.00	0.35
Serum bicarbonate, mmol/L	4000	+3.30	27.11	0.01	3663	+3.30	27.11	0.01	337	+3.30	27.11	0.02	0.91
Serum total calcium, mmol/L	3977	+0.83	2.293	0.00	3641	+0.83	2.290	0.00	336	+0.84	2.320	0.00	0.021
Serum phosphorus, mmol/L	3977	+0.09	1.094	0.01	3641	+0.08	1.083	0.01	336	+0.10	1.105	0.01	0.10
Serum uric acid, μmol/L	3977	+5.72	304.9	0.01	3641	+5.71	301.87	0.01	336	+5.82	336.97	0.02	<0.001
Serum glucose, mmol/L	3974	+1.68	5.366	0.01	3639	+1.68	5.366	0.01	335	+1.69	5.419	0.03	0.66
Serum blood urea nitrogen, SI	3977	+1.55	4.711	0.01	3641	+1.55	4.711	0.01	336	+1.54	4.664	0.02	0.49
Serum total bilirubin, μmol/L	3977	+2.25	9.487	0.02	3641	+2.25	9.487	0.02	336	+2.30	9.974	0.04	0.18
Serum creatinine, μmol/L	3977	+4.52	91.83	0.00	3641	+4.52	91.83	0.00	336	+4.61	100.48	0.02	<0.001
Serum iron, μmol/L	3977	+2.66	14.29	0.01	3641	+2.66	14.30	0.01	336	+2.68	14.58	0.04	0.65
Serum cholesterol, mmol/L	3977	+1.68	5.365	0.01	3641	+1.68	5.366	0.01	336	+1.69	5.419	0.02	0.40
Serum triglycerides, mmol/L	3977	+0.29	1.336	0.02	3641	+0.29	1.336	0.02	336	+0.35	1.419	0.06	0.34
Aspartate aminotransferase, U/L	3977	+3.02	20.49	0.01	3641	+3.02	20.49	0.01	336	+3.02	20.49	0.05	0.90
Alanine aminotransferase, U/L	3977	+2.86	17.46	0.02	3641	+2.87	17.63	0.02	336	+2.73	15.33	0.07	0.058
Gamma glutamyl transferase, U/L	3976	+3.17	23.80	0.02	3640	+3.16	23.57	0.02	336	+3.31	27.38	0.07	0.056
Serum lactate dehydrogenase, U/L	3976	+5.10	164.0	0.01	3641	+5.10	164.02	0.01	335	+5.10	164.02	0.01	0.86
Serum alkaline phosphatase, U/L	3977	+4.37	79.04	0.01	3641	+4.36	78.26	0.01	336	+4.52	91.83	0.02	<0.001
Serum total protein, g/L	3977	+4.29	72.96	0.00	3641	+4.29	72.97	0.00	336	+4.29	72.97	0.01	0.80
Serum albumin, g/L	3977	+3.71	40.85	0.00	3641	+3.71	40.85	0.00	336	+3.72	41.26	0.01	0.049
Serum globulin, g/L	3977	+3.46	31.81	0.01	3641	+3.46	31.82	0.01	336	+3.45	31.50	0.01	0.22
Serum osmolality, mmol/kg	3977	+5.64	281.4	0.00	3641	+5.64	281.46	0.00	336	+5.64	281.46	0.00	0.68
Glycated hemoglobin, %	3995	+1.68	5.365	0.01	3658	+1.68	5.365	0.01	337	+1.69	5.419	0.02	0.45
Plasma glucose, mmol/L	3996	+1.67	5.312	0.01	3659	+1.67	5.312	0.01	337	+1.68	5.365	0.02	0.66
Urinary cadmium, nmol/L	3964	+1.21	3.353	0.05	3637	+1.19	3.287	0.05	327	+1.47	4.349	0.10	0.003

Table 1. *Cont.*

	Overall				Hcy ≤ 14 μmol/L				Hcy > 14 μmol/L				P_{Hcy}
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	
Urinary creatinine, mmol/L	3960	+2.18	8.846	0.02	3635	+2.17	8.758	0.02	325	+2.30	9.974	0.06	0.048
Urinary albumin, μg/L	3960	+1.58	4.854	0.06	3635	+1.55	4.711	0.06	325	+1.91	6.753	0.16	0.041
Urinary iodine, μg/L	3956	+2.54	12.67	0.05	3631	+2.54	12.68	0.05	325	+2.51	12.30	0.07	0.68
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	P_{Hcy}
NHANES 1999–2006	10,151		100.0	0.0	9704		95.9	0.3	447		4.1	0.3	
Age (y)	10,151	45.8		0.20	9704	45.6		0.20	447		50.4	0.6	<0.001
Sex, % men	10,151		48.6	0.4	9704		48.2	0.4	447		59.9	0.4	0.005
Race/ethnicity	7605				7260								0.50
NH white			73.0	2.1			73.1	2.0	345		72.1	4.1	
NH black			10.9	1.1			10.7	1.0			16.5	2.5	
MA			6.8	1.0			7.0	1.0			3.0	0.6	
Other			9.2	1.1			9.3	1.1			8.4	2.6	
Poverty status	9471				9060								<0.001
PIR ≥ 125%			81.2	0.7			84.6	0.7	411		75.1	2.4	
PIR < 125%			15.8	0.7			15.4	0.7			24.9	2.4	
Hcy, Log_e	10,151	+2.08		0.01	9704	+2.04		0.01	447	+2.94		0.02	<0.001
	n	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	***n***	**Mean**	**%**	**SE**	P_{Hcy}
HANDLS 2004–2018	245		100.0		220		89.8		25		10.2		
Age (y)	245	49.2		0.56	220	48.6		0.59	25	54.3		1.5	0.002
Sex, % men	245		51.0		220		48.6		25		72.0		0.032
Race/ethnicity	245				220				25				
Whites			29.8				31.8				12.0		0.052
AA			70.2				68.2				88.0		
Poverty status	245				220				25				0.069
PIR ≥ 125%			37.1				39.1				20.0		
PIR < 125%			62.9				60.9				80.0		
Hcy, Log_e	245		+2.26	0.02	220	+2.19		0.02	25	+2.90		0.05	<0.001

Abbreviations: HANDLS = Healthy Aging in Neighborhoods of Diversity Across the Life Span: Hcy = Homocysteine; NH = non-Hispanic; NHANES = National Health and Nutrition Examination Surveys; PIR = Poverty Income Ratio; ROC = Receiver Operating Characteristics analysis. [1] All analyses, except for HANDLS, were adjusted for sampling design complexity, to obtain corrected standard errors for means and proportions. Means and proportions of study variables were compared across categories of Hcy (0 = normal, 1 = hyperhomocysteinemic), using simple linear regression for continuous variables and logistic regression in which Hcy category was the outcome for categorical variables. *p* value presented is associated with the regression coefficient. In NHANES, the regression models were also adjusted for sampling design complexity. Biochemical and hematological markers were compared by Hcy categories on their Log_e-transformed scale (Mean, SE). However, the exponentiated mean is also presented for better clinical interpretation.

Cross-validation (cv), adaptive and minBIC LASSO results are presented in Supplementary Table S1, using a random half sample of NHANES III, phase 2 and allowing for replication by sorting the sample by ID and setting a random seed to fixed value. Our findings indicated that Log_e-transformed Hcy was associated with a number of biochemical and hematological indices, forcing adjustment for socio-demographic factors, most of which were shown to associate with elevated Hcy in Table 1. The result of the adaptive LASSO (initial model of choice) followed by backward elimination process is shown in Supplementary Table S2 for both continuous and binary Hcy outcomes, while accounting for survey design complexity. In the reduced logistic regression model, the finally selected predictors included: serum folate (−), creatinine (+), age (+), serum vitamin B-12 (−), aspartate aminotransferase (+), alanine aminotransferase (−), SUA (+), mean cell hemoglobin, MCH (+), serum albumin (+), serum vitamin C (+), RDW (+), alkaline phosphatase (+), retinyl esters (−) and serum cotinine (+). Additional control for five markers identified by LASSO logistic (adaptive method, Supplemental Table S1) and not by the LASSO linear (adaptive method) did not alter this finding. Of those, only 10 markers survived the selection criteria of AUC > 0.55, and one (retinyl esters) was excluded due to missingness in recent waves. The remaining nine components were retained, with related ROC curves; estimated optimal cut-points for highest sensitivity/specificity are presented in Supplementary Figure S2. For NHANES 1999–2006, the ROC curves were similar for each of the nine components and cut-offs were comparable using the same criterion for elevated Hcy. The pooled NHANES similarly yielded a mid-range value of cut-points as indicated in Supplementary Figure S2.

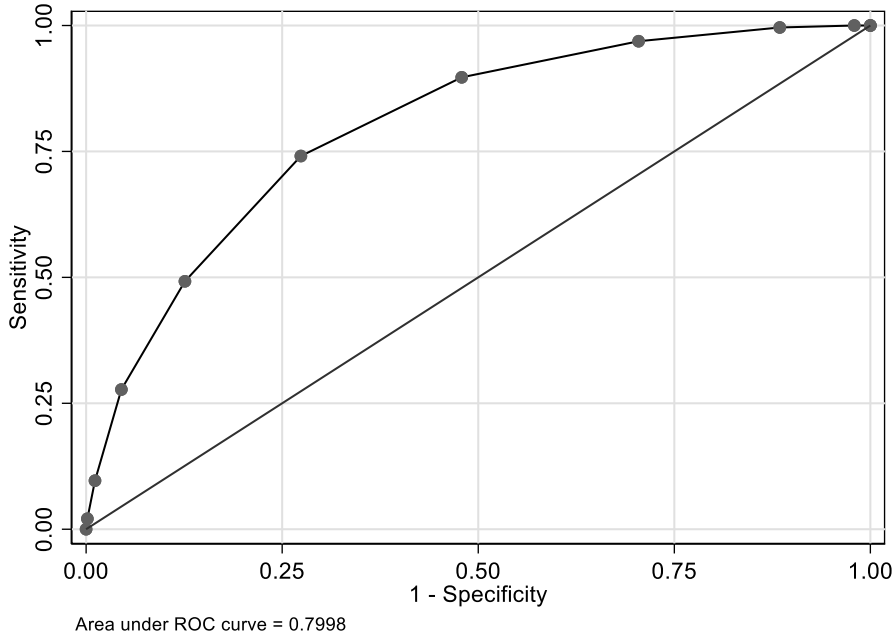


Figure 2. Nine-marker index (Index I) and its predictive value of elevated Hcy: ROC curves for pooled NHANES [1]. [1] n = 13,822; optimal cut-point was at 5; AUC = 0.799, 95% CI: 0.785,0.814. Index I included binary biomarkers of elevated Hcy selected using LASSO and backward elimination. The full list of the nine components of Index I are: Age, serum folate, serum vitamin B-12, serum creatinine, red cell distribution width, mean cell hemoglobin, serum cotinine, serum uric acid and alkaline phosphatase. Cut-points for individual components are: serum folate, Log_e, in nmol/L, <2.83; serum creatinine, Log_e, in μmol/L ≥4.481; older age, in years, ≥49; serum vitamin B-12, Log_e, in pmol/L, <5.74; mean cell hemoglobin, Log_e, in *pg*, ≥3.422; red cell distribution width, Log_e, in %, ≥2.553; serum uric acid, Log_e, in μmol/L, ≥5.826; serum alkaline phosphatase, Log_e, in μmol/L, ≥4.356 U/L; serum cotinine, Log_e, in ng/mL, −0.579.

The nine-component Index I predicted elevated Hcy with an AUC of 0.798 (95% CI: 0.783,0.812) in the pooled NHANES data (Figure 2). Similarly, Index II, which excluded serum cotinine, thus

including eight components, exhibited an AUC of 0.794; 95% CI: 0.780,0.809 (Figure 3), indicating that when two values of Indices I and II were chosen at random, the lower value corresponded to a ≤14 value of Hcy and the higher value to >14 value of Hcy, ~80% of the time, suggesting a high predictive value of both indices for elevated Hcy. Optimal cut-point for both indices was 5. For each binary index (≥5 vs. <5) and in the pooled NHANES, (Table 2), the adjusted odds of elevated Hcy were increased 6.9–7.4-fold. Each of the nine binary components of Index I, when included into the model, simultaneously predicted elevated Hcy, independently increasing the odds by >24% (higher RDW) up to 3.5-fold (lower serum folate).

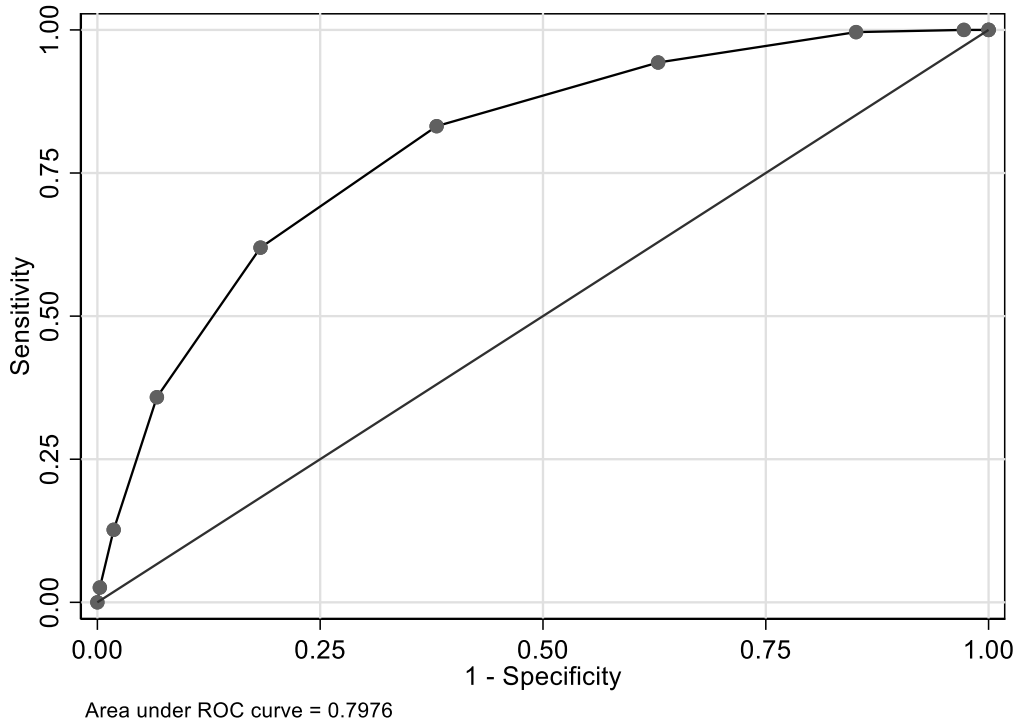


Figure 3. Eight-marker index (Index II) and its predictive value of elevated Hcy: pooled NHANES [1]. [1] n = 13,920; optimal cut-point at 5; AUC = 0.798; 95% CI: 0.783,0.812. The index consisted of a summation of all binary Index I biomarkers (See Figure 2 footnotes for cut-points), excluding blood cotinine, which was not available in HANDLS.

In the HANDLS sub-cohort analyses (Table 3), we found that Visit 1 Hcy was cross-sectionally associated, after multivariable adjustment, with lower serum folate, higher serum creatinine, lower serum vitamin B-12, and increases levels of SUA and alkaline phosphatase. Longitudinally, the multivariable adjusted mixed-effects regression model indicated that Visit 1 Hcy was linked to faster rate of increase in serum creatinine and serum vitamin B -12, with a trend towards a direct association ($p < 0.10$) with rates of increase in RDW and alkaline phosphatase. Finally, Visit tHcy (Log_e-transformed) was found to be moderately and positively correlated with Visit 3 Index II, computed using pooled NHANES cut-points, with a Pearson's correlation r = +0.34 (n = 81). The smoothed positive association is depicted in Figure 4, indicating a linear relationship for the range of the data.

Table 2. Selected independent binary correlates and indices (I and II) of elevated homocysteine: Reduced multiple logistic regression models and model-specific area under the ROC curve; pooled NHANES III, phase 2 and 1999–2006 [1].

	Elevated Homocysteine		
	OR	**95% CI**	***p*-Value**
Model 1: Binary predictors, (n = 14,739)			
Lower serum folate [2]	3.49	(2.63,4.63)	<0.001
Higher serum creatinine [3]	1.86	(1.51,2.29)	<0.001
Older age [4]	1.95	(1.56,2.44)	<0.001
Lower serum vitamin B-12 [5]	2.52	(1.98,3.21)	<0.001
Higher MCH [6]	1.60	(1.28,2.02)	<0.001
Higher RDW [7]	1.24	(1.01,1.54)	0.044
Higher SUA [8]	1.67	(1.35,2.06)	<0.001
Higher alkaline phosphatase [9]	1.71	(1.35,2.15)	<0.001
Higher serum cotinine [10]	1.77	(1.44,2.17)	<0.001
Model 2: Index I ≥ 5 (n = 14,739)	7.43	(5.75,9.61)	<0.001
Model 3: Index II ≥ 5, (n = 14,829)	6.90	(5.37,8.84)	<0.001

Abbreviations: HANDLS = Healthy Aging in Neighborhoods of Diversity Across the Life Span: Hcy = Homocysteine; NHANES = National Health and Nutrition Examination Surveys; ROC = Receiver Operating Characteristics analysis. [1] All elements of Indices I and II were Log_e transformed. Cut-points are determined using ROC analysis, using highest sensitivity/specificity combinations. Binary components are entered simultaneously in Model 1. Model 2 includes Index I which sums binary variables: "lower serum folate" (1 = yes, 0 = no), "Higher serum creatinine", "Older age", "Lower serum vitamin B-12", "Higher MCH", "Higher RDW", "Higher SUA", "Higher alkaline phosphatase" and "Higher serum cotinine", with a possible range of 0–9. Model 3 included Index II which sums binary variables of Index I excluding "Higher blood lead" and "Higher serum cotinine". Possible range: 0–7. Cut-points for Indices I and II were also determined using ROC curve analysis, with an optimal cut-point selected using the highest sensitivity/specificity combination. [2] Serum folate, Log_e, in nmol/L, < 2.83; [3] Serum creatinine, Log_e, in μmol/L ≥4.481; [4] Older age, in years, ≥49, [5] Serum vitamin B-12, Log_e, in pmol/L, <5.74; [6] Mean cell hemoglobin, Log_e, in pg, ≥3.422; [7] Red cell distribution width, Log_e, in %, ≥2.553; [8] Serum Uric Acid, Log_e, in μmol/L, ≥5.826; [9] Serum alkaline phosphatase, Log_e, in μmol/L, ≥4.356 U/L; [10] Serum cotinine, Log_e, in ng/mL, −0.579.

Table 3. Baseline serum homocysteine as a predictor of selected biochemical and hematological parameters at baseline and their change over time: mixed-effects linear regression models; HANDLS 2004–2018 [1].

Outcome	Intercept		Time		Hcy		Hcy × Time		(*n*) k
	γ_{00} ± SE	*p*	γ_{10} ± SE	*p*	γ_{0a} ± SE	*p*	γ_{1a} ± SE	*p*	
Serum folate, nmol/L									
Model 1: Age-adjusted	+33.6 ± 1.2	<0.001	+0.11 ± 0.16	0.47	−0.21 ± 0.32	0.51	−0.05 ± 0.05	0.36	(243) k = 2.4
Model 2: Socio-demographic adjusted	+29.8 ± 6.0	<0.001	+2.07 ± 0.83	0.012	−0.25 ± 0.34	0.45	−0.05 ± 0.05	0.37	(243) k = 2.4
Model 3: Multivariable-adjusted	+35.1 ± 6.2	<0.001	+1.39 ± 0.87	0.11	−1.41 ± 0.43	0.001	+0.03 ± 0.07	0.69	(227) k = 2.4
Serum creatinine, µmol/L									
Model 1: Age-adjusted	+107.9 ± 7.1	<0.001	−1.18 ± 0.30	<0.001	+18.3 ± 1.89	<0.001	+0.11 ± 0.10	0.24	(243) k = 2.4
Model 2: Socio-demographic adjusted	+100.1 ± 37.1	0.007	−3.40 ± 1.50	0.024	+18.3 ± 2.00	<0.001	+0.14 ± 0.10	0.16	(243) k = 2.4
Model 3: Multivariable-adjusted	+117.7 ± 35.1	0.001	−3.93 ± 1.60	0.014	+19.8 ± 2.01	<0.001	+0.24 ± 0.12	0.045	(227) k = 2.4
Serum vitamin B-12, µmol/L									
Model 1: Age-adjusted	+472 ± 13	<0.001	−11.2 ± 1.6	<0.001	−4.64 ± 3.69	0.20	+0.85 ± 0.53	0.11	(243) k = 2.4
Model 2: Socio-demographic adjusted	+334 ± 68	<0.001	+6.1 ± 8.6	0.48	−5.19 ± 3.80	0.17	+0.87 ± 0.55	0.11	(243) k = 2.4
Model 3: Multivariable-adjusted	+307 ± 69	<0.001	−8.0 ± 8.9	0.37	−11.0 ± 4.9	0.023	+1.66 ± 0.69	0.015	(227) k = 2.4
Mean cell hemoglobin, pg									
Model 1: Age-adjusted	+29.6 ± 0.2	<0.001	−0.003 ± 0.002	0.56	+0.03 ± 0.05	0.53	−0.003 ± 0.005	0.56	(244) k = 2.4
Model 2: Socio-demographic adjusted	+30.9 ± 0.9	<0.001	−0.04 ± 0.10	0.67	+0.02 ± 0.05	0.72	−0.004 ± 0.005	0.49	(244) k = 2.4
Model 3: Multivariable-adjusted	+29.2 ± 0.97	<0.001	+0.03 ± 0.10	0.78	+0.04 ± 0.06	0.50	+0.01 ± 0.01	0.42	(227) k = 2.4
Red cell distribution width, %									
Model 1: Age-adjusted	+13.8 ± 0.11	<0.001	+0.12 ± 0.01	<0.001	*+0.05 ± 0.03*	*0.079*	+0.007 ± 0.003	0.035	(244) k = 2.4
Model 2: Socio-demographic adjusted	+12.7 ± 0.57	<0.001	+0.18 ± 0.06	0.002	+0.04 ± 0.03	0.16	+0.008 ± 0.004	0.021	(244) k = 2.4
Model 3: Multivariable-adjusted	+12.7 ± 0.5	<0.001	+0.19 ± 0.06	0.002	−0.04 ± 0.04	0.20	*+0.009 ± 0.005*	*0.056*	(227) k = 2.4
Serum uric acid, µmol/L									
Model 1: Age-adjusted	+314.8 ± 5.3	<0.001	+3.79 ± 0.7	<0.001	+7.48 ± 1.45	<0.001	−0.20 ± 0.22	0.36	(243) k = 2.4
Model 2: Socio-demographic adjusted	+226.8 ± 26.7	<0.001	+2.80 ± 3.51	0.42	+5.30 ± 1.48	<0.001	−0.12 ± 0.23	0.62	(243) k = 2.4
Model 3: Multivariable-adjusted	+228.6 ± 27.0	<0.001	+2.01 ± 3.61	0.58	+8.11 ± 1.88	<0.001	−0.01 ± 0.29	0.97	(227) k = 2.4
Serum alkaline phosphatase, U/L									
Model 1: Age-adjusted	+90.8 ± 2.1	<0.001	−0.88 ± 0.22	<0.001	+2.66 ± 0.56	<0.001	−0.03 ± 0.07	0.70	(243) k = 2.4
Model 2: Socio-demographic adjusted	+70.7 ± 10.5	<0.001	−0.39 ± 1.19	0.74	+2.54 ± 0.58	<0.001	−0.002 ± 0.08	0.98	(243) k = 2.4
Model 3: Multivariable-adjusted	+75.2 ± 10.7	<0.001	−0.27 ± 1.23	0.83	+2.72 ± 0.74	<0.001	*+0.16 ± 0.10*	*0.095*	(227) k = 2.4

Abbreviations: HANDLS = Healthy Aging in Neighborhoods of Diversity Across the Life Span; Hcy = Homocysteine; k = mean number of observations/participant; *n* = Number of participants. [1] All selected biochemical and hematological markers were measured in SI units. Measures were not Log_e transformed. Model 1 was adjusted for age, centered at 48.8 years. Model 2 was additionally adjusted for sex, race (African Americans vs. Whites), and poverty status (above vs. below poverty). Model 3 was additionally adjusted for all remaining biochemical and hematological measures that were selected. In Model 3, Folate was centered at 32.4, creatinine at 105.8, vitamin B-12 at 494.76, Mean cell hemoglobin at 29.46, red cell distribution width at 13.64, serum uric acid at 306.2, and alkaline phosphatase at 82.

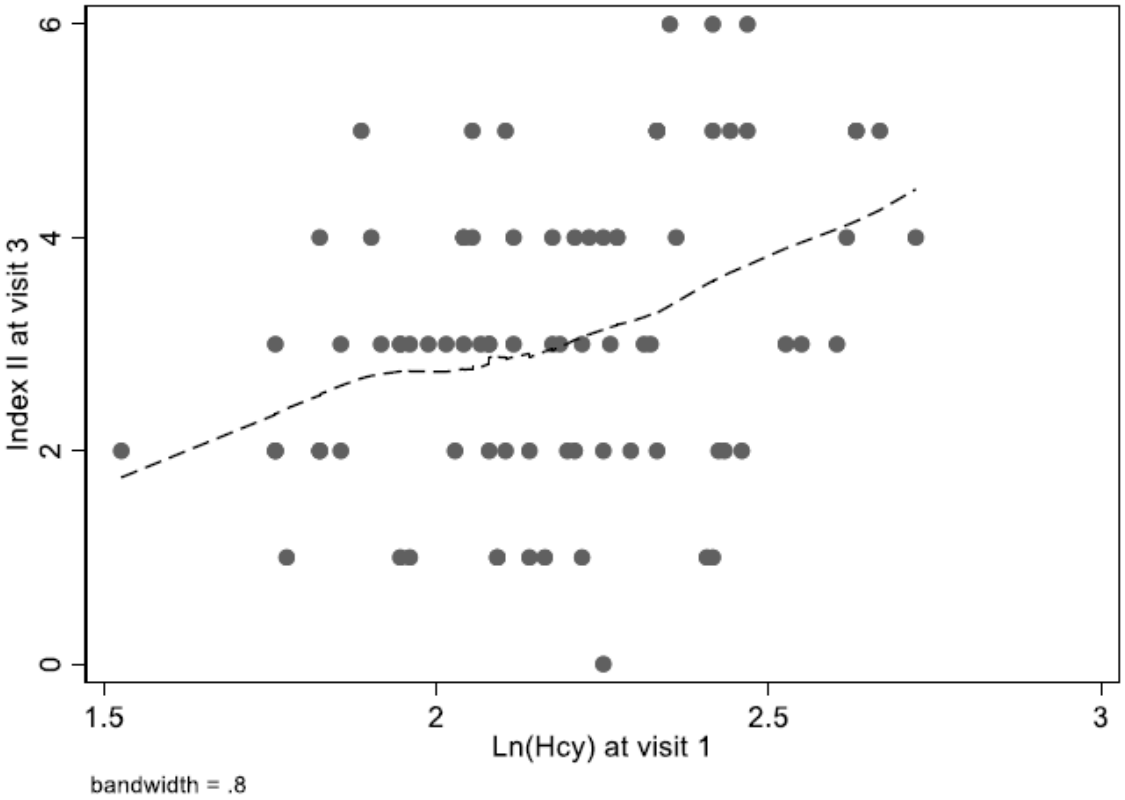


Figure 4. Lowess smoother of total Visit 1 homocysteine vs. Visit 3 Index II (r = 0.34, n = 81): HANDLS 2004–2009 (Hcy) and 2013–2018 (Index II) [1]. Abbreviations: HANDLS = Healthy Aging in Neighborhoods of Diversity Across the Life Span: Hcy = Homocysteine; NHANES = National Health and Nutrition Examination Surveys; ROC = Receiver Operating Characteristics analysis. [1] Blood total Hcy, in μmol/L, is Log_e-transformed and uncategorized. Index II was computed using cut-points obtained from the pooled NHANES ROC analysis. See Table 2 footnotes for details. Index II may range from 0 to 8, and no cut-point was used in this analysis.

4. Discussion

Here we present data among middle-aged adults, validating a predictive index for Hcy > 14 μmol/L derived from independent biochemical and hematological correlates using modern techniques. The study uncovered up to nine independent predictors for elevated Hcy, some of which have been found to be correlated with each other (e.g., serum folate, B-12, MCH, and RDW) in previous studies as well as with hyperhomocysteinemia. Both adjusted and unadjusted associations indicated that serum folate was the most predictive factor that was inversely related to elevated Hcy. Our findings of inverse associations of serum folate and cobalamin with elevated Hcy concentrations conform with earlier research using national data from pre-folate and post-folate fortification eras [17,19,42]. In fact, the two previous NHANES studies (III, phase 2:1991–1994 and 1999–2004) reported comparable findings, despite examining pre-selected factors, rather than exploring all available biochemical and hematological markers. The NHANES III study concluded that serum creatinine and cobalamin concentrations showed the strongest and weakest association with blood Hcy, respectively [19]. Notably, folate and vitamin B-12 were inversely related to Hcy. Men had higher mean Hcy than women, along with lower concentrations of serum folate, red blood cell (RBC) folate, and serum vitamin B12. The NHANES 1999–2004 study concluded that blood Hcy concentration was ~9.7% higher in men vs. woman [17], and was directly related to systolic blood pressure, serum creatinine, and serum cotinine, while being inversely correlated with serum folate levels, RBC folate, and serum vitamin B-12, and positively correlated with methylmalonic acid (MMA) concentration [17]. While those associations were largely replicated in our study, age and not sex was retained in the model upon

backward elimination. Moreover, our study excluded other lifestyle or health-related factors, selecting only biochemical and hematological indices measured in NHANES III, phase 2 (1991–1994) along with Hcy. Consequently, MMA was excluded due to its unavailability in the NHANES III phase 2 data from among the 82 selected biomarkers.

Among retained correlates, SUA was strongly related to Hcy. In a retrospective cohort study (n = 16,477, age: 20–80 years), elevated SUA was previously directly associated with hyperhomocysteinemia, whereby the fully adjusted association remained significant only among men (OR = 1.5; 95% CI, 1.3,1.7; $p < 0.001$) [43]. These results are comparable to our findings, whereby SUA ≥ 339 μmol/L was associated with a 1.67-fold increase in the odds of hyperhomocysteinemia (>14 μmol/L), with a 95% CI: 1.35,2.06.

Hyperuricemia can be modified with diet as well, including reduced alcohol, red meat and sugar consumption [44,45].

High MCH is commonly a sign of macrocytic anemia (i.e., enlarged RBCs) subsequent to folate or vitamin B-12 deficiency [46], though it may also result from liver diseases [46]. Thus, Hcy may well be a correlate of higher MCH, resulting from any or both vitamin deficiencies; and a higher Hcy may result in higher MCH over time. Nevertheless, our findings indicated that although an independent correlate of elevated Hcy, MCH was not among the strongest predictors based on ROC analyses. Furthermore, our longitudinal analyses did not indicate that baseline Hcy was associated with faster increase in MCH over time. Nevertheless, larger studies are needed to corroborate these findings. Elevated RDW, reflecting RBC size variability (i.e., anisocytosis), independently predicted chronic disease morbidity and mortality [47–51]. Unlike MCH, RDW was previously studied in relation to hyperhomocysteinemia [22,52], and was found to be directly related to elevated Hcy in one cross-sectional study of middle-aged Chinese adults, independently of age, neutrophil count, mean corpuscular volume, and hemoglobin [22]. Another larger cross-sectional study of 5554 adults (18–64 years), however, failed to detect this independent RDW-Hcy [52]. Our findings indicated that elevated MCH was in fact more strongly associated with elevated Hcy (OR = 1.60, 95% CI: 1.28,2.02) when compared with RDW (OR = 1.24, 95% CI:1.01,1.54). Thus, elevated Hcy may be a stronger marker of enlarged RBCs than of anisocytosis. Nevertheless, our longitudinal analysis has shown that Hcy in its continuous form predicted RDW to a greater extent than MCH, and was associated with a faster rate of increase in RDW over time.

Moreover, serum bone alkaline phosphatase, a marker of biliary inflammation and cholelithiasis [53], was shown to be up-regulated in vitamin B-6 deficiency [25,54]. As stated earlier, Hcy was previously inversely corelated with vitamin B-6 status [9–14], as the latter is directly involved in OCM. Our study is to date the first to show that higher blood Hcy is associated with elevated alkaline phosphatase, both cross-sectionally and longitudinally. Thus, although the main modifiers of Hcy are B-vitamins, particularly folate and vitamin B-12, liver enzymes are correlates of Hcy that can be modulated with reduced alcohol consumption [55], and alkaline phosphatase in particular is a key mediator in the reported association between Hcy and reduction in bone mineral density associated with osteoporosis among postmenopausal women [56,57]. Thus, Hcy may be merely a marker of certain health outcomes, while liver enzyme elevations can act as the main causal pathway.

Several studies have indicated that smokers had more elevated blood Hcy than non-smokers, independently of other factors, while having lower serum levels of folate and vitamin B-12 [23]. Among self-reported never smokers >20 years of age [NHANES III, n = 3232], serum cotinine quartiles were independently and linearly associated with blood Hcy, as were age, being male, being non-white, and having lower sum folate or serum B-12 [58]. Serum cotinine was also among key predictors of elevated Hcy in studies examining multiple correlates in earlier and more recent NHANES [17,19]. Thus, stopping cigarette smoking may have an effect of reducing the risk of elevated Hcy. Nevertheless, controlled randomized trials are needed to ascertain a causal association. It is worth noting that both serum cotinine and urinary cadmium were linked to recent smoking [59,60].

Several notable study strengths include the novel coupled use of machine learning and ROC analyses to select independent predictors for elevated Hcy and subsequently create combined indices

and conduct multivariable regression models. The initial analysis screened over 82 biochemical and hematological biomarkers, and our confidence in the predictive indices was enhanced by validation between cross-sectional national data and an independent longitudinal study of urban adults. The LASSO linear model was used to obtain a first set of predictors for continuous Log_e-transformed Hcy, which were then applied to the binary outcome, given that the 14 μmol/L cutoff to define elevated Hcy might be considered arbitrary for some health outcomes, aside from AD. Thus, our goal was to limit the set of markers to those that independently predicted Hcy, both in its continuous and categorical form.

Among the limitations, the threshold used for Hcy of 14 μmol/L in our study, while being used by others previously, may be sub-optimal in some samples, given their different levels of mean Hcy, particularly given the decreasing prevalence rates between pre- and post-folate fortification. Thus, even though cutoffs for predictors were comparable at optimal sensitivity and specificity between NHANES waves, the predictive values may have differed between those two waves, with expected higher positive predictive value at higher prevalence of elevated Hcy (i.e., NHANES III, phase 2) and vice versa for the negative predictive value. Second, Hcy measurement, while comparable between waves, used different techniques between NHANES III, phase 2 and the more recent NHANES, potentially affecting the validity of the cut-point used between those two waves. Nevertheless, given that comparable biomarker optimal cut-points were obtained between NHANES waves through ROC analyses for Hcy > 14 μmol/L, measurement errors ascribed to differential use of techniques (HPLC vs. immunoassay) was assumed minimal. Attempts to calibrate those two methods are needed in future studies with repeat measures using both methods. Finally, our key findings and the indices derived from ROC analyses may be applicable only at mid-life, a time window whereby cardiovascular and neurodegenerative diseases can be prevented through Hcy-reducing interventions. Nevertheless, future studies should examine those relationships and validate those indices among older adults aged ≥ 65 years.

In sum, we provide evidence of internal and external validity of indices composed of several biochemical and hematological markers that are strongly associated with elevated Hcy, which may be used as proxies in future longitudinal studies. Components of those indices that are amenable to intervention (e.g., folate and B-12 supplementation, alcohol consumption which affects both liver enzymes and uric acid, cigarette smoking) should also be studied as alternative pathways for which elevated Hcy can affect cardiovascular and neurodegenerative disease trajectories.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6643/12/4/950/s1, Supplementary Figure S1. Participant flowchart. Supplementary Figure S2. ROC curves of selected correlates of elevated homocysteine (>14 μmol/L) and cut-point determination: NHANES III, phase 2; NHANES 1999–2006 and pooled NHANES. Supplemental Methods 1: Biochemical and Hematological Indices: NHANES III, phase 2: 1991–1994. Supplemental Methods 2: Biochemical and Hematological Indices Selected for NHANES 1999–2006. Supplemental Methods 3: Biochemical and Hematological Indices Selected for HANDLS 2004–2018. Supplemental Methods 4: Description of Time-Interval Mixed-effects linear regression models, HANDLS 2004–2018. Supplementary Table S1. Predictive models of serum homocysteine (Loge transformed): CV, minimum BIC and adaptive linear LASSO for NHANES III, phase 2: training half-sample results and consistency with testing half-sample, 30–65 y. Supplementary Table S2. Reduced selected LASSO model for biochemical and hematological predictors (Loge transformed and z-scored) of Hcy (Loge transformed and z-scored) and elevated Hcy (equivalent to Hcy > 14 μmol/L), adjusting for socio-demographic factors: NHANES III, phase 2; ages 30–65 y.

Author Contributions: M.A.B. contributed to the study concept, planned analysis, conducted data management and statistical analysis, conducted literature review, wrote and revised the manuscript. H.A.B. planned analysis, assisted in data management and statistical analysis, conducted literature review, wrote and revised parts of the manuscript. P.H.M. planned analysis, conducted literature search and review, wrote and revised parts of the manuscript. S.H. planned analysis, conducted literature search and review, wrote parts of the manuscript, revised the manuscript. J.A.C. conducted literature review, wrote and revised parts of the manuscript. M.K.E. acquired data, wrote and revised parts of the manuscript. A.B.Z. acquired data, planned analysis, wrote and revised parts of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Intramural Research Program of the NIH, National institute on Aging.

Acknowledgments: We would like to thank Megan Williams and Nicolle Mode (NIA/NIH/IRP) for internally reviewing our manuscript. The authors declare no conflict of interest.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AUC	Area under the curve
BIC	Bayesian Information Criterion
CDC	Centers for Disease Control and Prevention
cv	Cross-validation
HANDLS	Healthy Aging in Neighborhood of Diversity Across the Life Span
Hcy	Homocysteine
LASSO	least absolute shrinkage and selection operator
LOWESS	Locally weighted regression
MCH	Mean cell hemoglobin
MEC	mobile examination center
MMA	methylmalonic acid
MRV	Medical Research Vehicles
MTHF	N-5-methyl-tetrahydrofolate
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Surveys
RDW	Red cell distribution width
ROC	Receiver Operating Characteristic
SAH	S-adenosylhomocysteine
SAM	S-adenosylemethionine
SUA	Serum Uric Acid

References

1. Selhub, J. Public health significance of elevated homocysteine. *Food Nutr. Bull.* **2008**, *29*, S116–S125. [CrossRef] [PubMed]
2. Seshadri, S.; Beiser, A.; Selhub, J.; Jacques, P.F.; Rosenberg, I.H.; D'Agostino, R.B.; Wilson, P.W.; Wolf, P.A. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *N. Engl. J. Med.* **2002**, *346*, 476–483. [CrossRef] [PubMed]
3. Clarke, R.; Smith, A.D.; Jobst, K.A.; Refsum, H.; Sutton, L.; Ueland, P.M. Folate, vitamin B12, and serum total homocysteine levels in confirmed Alzheimer disease. *Arch. Neurol.* **1998**, *55*, 1449–1455. [CrossRef]
4. Beydoun, M.A.; Beydoun, H.A.; Gamaldo, A.A.; Teel, A.; Zonderman, A.B.; Wang, Y. Epidemiologic studies of modifiable factors associated with cognition and dementia: Systematic review and meta-analysis. *BMC Public Health* **2014**, *14*, 643. [CrossRef] [PubMed]
5. Selhub, J.; Jacques, P.F.; Rosenberg, I.H.; Rogers, G.; Bowman, B.A.; Gunter, E.W.; Wright, J.D.; Johnson, C.L. Serum total homocysteine concentrations in the third National Health and Nutrition Examination Survey (1991–1994): Population reference ranges and contribution of vitamin status to high serum concentrations. *Ann. Intern. Med.* **1999**, *131*, 331–339. [CrossRef] [PubMed]
6. Booth, G.L.; Wang, E.E. Preventive health care, 2000 update: Screening and management of hyperhomocysteinemia for the prevention of coronary artery disease events. The Canadian Task Force on Preventive Health Care. *CMAJ* **2000**, *163*, 21–29.
7. Marti-Carvajal, A.J.; Sola, I.; Lathyris, D.; Dayer, M. Homocysteine-lowering interventions for preventing cardiovascular events. *Cochrane Database Syst. Rev.* **2017**, *8*, CD006612. [CrossRef]
8. Selhub, J.; Miller, J.W. The pathogenesis of homocysteinemia: Interruption of the coordinate regulation by S-adenosylmethionine of the remethylation and transsulfuration of homocysteine. *Am. J. Clin. Nutr.* **1992**, *55*, 131–138. [CrossRef]
9. Hatzis, C.M.; Bertsias, G.K.; Linardakis, M.; Scott, J.M.; Kafatos, A.G. Dietary and other lifestyle correlates of serum folate concentrations in a healthy adult population in Crete, Greece: A cross-sectional study. *Nutr. J.* **2006**, *5*, 5. [CrossRef]

10. Manavifar, L.; Nemati Karimooy, H.; Jamali, J.; Talebi Doluee, M.; Shirdel, A.; Nejat Shokohi, A.; Fatemi Nayyeri, M. Homocysteine, Cobalamin and Folate Status and their Relations to Neurocognitive and Psychological Markers in Elderly in Northeasten of Iran. *Iran J. Basic Med. Sci.* **2013**, *16*, 772–780.
11. Song, J.H.; Park, M.H.; Han, C.; Jo, S.A.; Ahn, K. Serum Homocysteine and Folate Levels are Associated With Late-life Dementia in a Korean Population. *Osong Public Health Res. Perspect.* **2010**, *1*, 17–22. [CrossRef] [PubMed]
12. Yang, X.; Gao, F.; Liu, Y. Association of homocysteine with immunological-inflammatory and metabolic laboratory markers and factors in relation to hyperhomocysteinaemia in rheumatoid arthritis. *Clin. Exp. Rheumatol.* **2015**, *33*, 900–903. [PubMed]
13. Cheng, C.H.; Huang, Y.C.; Chen, F.P.; Chou, M.C.; Tsai, T.P. B-vitamins, homocysteine and gene polymorphism in adults with fasting or post-methionine loading hyperhomocysteinemia. *Eur. J. Nutr.* **2008**, *47*, 491–498. [CrossRef]
14. Selhub, J.; Jacques, P.F.; Wilson, P.W.; Rush, D.; Rosenberg, I.H. Vitamin status and intake as primary determinants of homocysteinemia in an elderly population. *JAMA* **1993**, *270*, 2693–2698. [CrossRef] [PubMed]
15. Rozen, R. Genetic predisposition to hyperhomocysteinemia: Deficiency of methylenetetrahydrofolate reductase (MTHFR). *Thromb. Haemost.* **1997**, *78*, 523–526. [CrossRef] [PubMed]
16. Ohishi, T.; Fujita, T.; Suzuki, D.; Nishida, T.; Asukai, M.; Matsuyama, Y. Serum homocysteine levels are affected by renal function during a 3-year period of minodronate therapy in female osteoporotic patients. *J. Bone Miner. Metab.* **2019**, *37*, 319–326. [CrossRef] [PubMed]
17. Ganji, V.; Kafai, M.R. Demographic, lifestyle, and health characteristics and serum B vitamin status are determinants of plasma total homocysteine concentration in the post-folic acid fortification period, 1999–2004. *J. Nutr.* **2009**, *139*, 345–352. [CrossRef]
18. Francis, M.E.; Eggers, P.W.; Hostetter, T.H.; Briggs, J.P. Association between serum homocysteine and markers of impaired kidney function in adults in the United States. *Kidney Int.* **2004**, *66*, 303–312. [CrossRef]
19. Ganji, V.; Kafai, M.R.; Third National, H.; Nutrition Examination, S. Demographic, health, lifestyle, and blood vitamin determinants of serum total homocysteine concentrations in the third National Health and Nutrition Examination Survey, 1988–1994. *Am. J. Clin. Nutr.* **2003**, *77*, 826–833. [CrossRef]
20. Fukagawa, N.K.; Martin, J.M.; Wurthmann, A.; Prue, A.H.; Ebenstein, D.; O'Rourke, B. Sex-related differences in methionine metabolism and plasma homocysteine concentrations. *Am. J. Clin. Nutr.* **2000**, *72*, 22–29. [CrossRef]
21. Sadre-Marandi, F.; Dahdoul, T.; Reed, M.C.; Nijhout, H.F. Sex differences in hepatic one-carbon metabolism. *BMC Syst. Biol.* **2018**, *12*, 89. [CrossRef] [PubMed]
22. Peng, Y.F.; Pan, G.G. Red blood cell distribution width predicts homocysteine levels in adult population without vitamin B12 and folate deficiencies. *Int. J. Cardiol.* **2017**, *227*, 8–10. [CrossRef] [PubMed]
23. Haj Mouhamed, D.; Ezzaher, A.; Neffati, F.; Douki, W.; Najjar, M.F. Effect of cigarette smoking on plasma homocysteine concentrations. *Clin. Chem. Lab. Med.* **2011**, *49*, 479–483. [CrossRef]
24. Lippi, G.; Salvagno, G.L.; Targher, G.; Montagnana, M.; Guidi, G.C. Plasma gamma-glutamyl transferase activity predicts homocysteine concentration in a large cohort of unselected outpatients. *Intern. Med.* **2008**, *47*, 705–707. [CrossRef]
25. Tanaka, T.; Scheet, P.; Giusti, B.; Bandinelli, S.; Piras, M.G.; Usala, G.; Lai, S.; Mulas, A.; Corsi, A.M.; Vestrini, A.; et al. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am. J. Hum. Genet.* **2009**, *84*, 477–482. [CrossRef] [PubMed]
26. Center for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey. Available online: http://www.cdc.gov/nchs/nhanes.htm (accessed on 25 October 2019).
27. NCHS. *Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988–1994*; NCHS: Highlandsville, MD, USA, 1994.
28. Crider, K.S.; Bailey, L.B.; Berry, R.J. Folic acid food fortification-its history, effect, concerns, and future directions. *Nutrients* **2011**, *3*, 370–384. [CrossRef]
29. Center for Disease Control and Prevention (CDC). Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey (NHANES III), 1988–1994. Available online: https://www.cdc.gov/nchs/data/nhanes/nhanes3/cdrom/nchs/manuals/labman.pdf (accessed on 25 October 2019).

30. Evans, M.K.; Lepkowski, J.M.; Powe, N.R.; LaVeist, T.; Kuczmarski, M.F.; Zonderman, A.B. Healthy aging in neighborhoods of diversity across the life span (HANDLS): Overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethn. Dis.* **2010**, *20*, 267–275.
31. Araki, A.; Sako, Y. Determination of free and total homocysteine in human plasma by high-performance liquid chromatography with fluorescence detection. *J. Chromatogr.* **1987**, *422*, 43–52. [CrossRef]
32. Abbott Homocysteine (HCY) assay package insert fo IMX Analyzer. Available online: https://www.cdc.gov/nchs/data/nhanes/nhanes_01_02/l06_b_met_homocysteine_IMX.pdf (accessed on 25 October 2019).
33. Pernet, P.; Lasnier, E.; Vaubourdolle, M. Evaluation of the AxSYM homocysteine assay and comparison with the IMx homocysteine assay. *Clin. Chem.* **2000**, *46*, 1440–1441. [CrossRef]
34. Laboratories, B.R. *Instruction Manual, Bio-Rad Quantaphase Folate Radioassay Kit*; Bio-Rad Laboratories: Hercules, CA, USA, 1987.
35. Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Surveys (NHANES 2005–06): Description of Laboratory Methodology: Vitamin B-12. Available online: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/B12_D.htm (accessed on 15 February 2019).
36. Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Surveys (NHANES 2005–06): Description of Laboratory Methodology: Folate. Available online: https://wwwn.cdc.gov/Nchs/Nhanes/2005-2006/FOLATE_D.htm (accessed on 15 February 2019).
37. Diagnostics, Q. Vitamin B-12 (Cobalamin) and Folate Panel. Available online: https://testdirectory.questdiagnostics.com/test/test-detail/7065/vitamin-b12-cobalamin-and-folate-panel-serum?cc=MASTER (accessed on 21 October 2019).
38. STATA. *Statistics/Data Analysis: Release 16.0*; Stata Corporation: College Station, TX, USA, 2019.
39. Zou, H. The adaptive Lasso and it oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1428. [CrossRef]
40. Albeck, M.J.; Borgesen, S.E. ROC-curve analysis. A statistical method for the evaluation of diagnostic tests. *Ugeskr Laeger* **1990**, *152*, 1650–1653. [PubMed]
41. Soreide, K. Receiver-operating characteristic (ROC) curve analysis in diagnostic, prognostic and predictive biomarker research. *J. Clin. Pathol.* **2008**. [CrossRef]
42. Jacques, P.F.; Bostom, A.G.; Wilson, P.W.; Rich, S.; Rosenberg, I.H.; Selhub, J. Determinants of plasma total homocysteine concentration in the Framingham Offspring cohort. *Am. J. Clin. Nutr.* **2001**, *73*, 613–621. [CrossRef] [PubMed]
43. Cohen, E.; Levi, A.; Vecht-Lifshitz, S.E.; Goldberg, E.; Garty, M.; Krause, I. Assessment of a possible link between hyperhomocysteinemia and hyperuricemia. *J. Investig. Med.* **2015**, *63*, 534–538. [CrossRef] [PubMed]
44. Beydoun, M.A.; Canas, J.A.; Fanelli-Kuczmarski, M.T.; Tajuddin, S.M.; Evans, M.K.; Zonderman, A.B. Genetic risk scores, sex and dietary factors interact to alter serum uric acid trajectory among African-American urban adults. *Br. J. Nutr.* **2017**, *117*, 686–697. [CrossRef]
45. Beydoun, M.A.; Fanelli-Kuczmarski, M.T.; Canas, J.A.; Beydoun, H.A.; Evans, M.K.; Zonderman, A.B. Dietary factors are associated with serum uric acid trajectory differentially by race among urban adults. *Br. J. Nutr.* **2018**, *120*, 935–945. [CrossRef]
46. Aslinia, F.; Mazza, J.J.; Yale, S.H. Megaloblastic anemia and other causes of macrocytosis. *Clin. Med. Res.* **2006**, *4*, 236–241. [CrossRef]
47. Li, N.; Zhou, H.; Tang, Q. Red Blood Cell Distribution Width: A Novel Predictive Indicator for Cardiovascular and Cerebrovascular Diseases. *Dis. Markers* **2017**, *2017*, 7089493. [CrossRef]
48. Tajuddin, S.M.; Nalls, M.A.; Zonderman, A.B.; Evans, M.K. Association of red cell distribution width with all-cause and cardiovascular-specific mortality in African American and white adults: A prospective cohort study. *J. Transl. Med.* **2017**, *15*, 208. [CrossRef]
49. Hoffmann, J.J. Red cell distribution width and mortality risk. *Clin. Chim. Acta* **2012**, *413*, 824–825. [CrossRef]
50. Perlstein, T.S.; Weuve, J.; Pfeffer, M.A.; Beckman, J.A. Red blood cell distribution width and mortality risk in a community-based prospective cohort. *Arch. Intern. Med.* **2009**, *169*, 588–594. [CrossRef] [PubMed]
51. Patel, K.V.; Semba, R.D.; Ferrucci, L.; Newman, A.B.; Fried, L.P.; Wallace, R.B.; Bandinelli, S.; Phillips, C.S.; Yu, B.; Connelly, S.; et al. Red cell distribution width and mortality in older adults: A meta-analysis. *J. Gerontol. A Biol. Sci. Med. Sci.* **2010**, *65*, 258–265. [CrossRef] [PubMed]

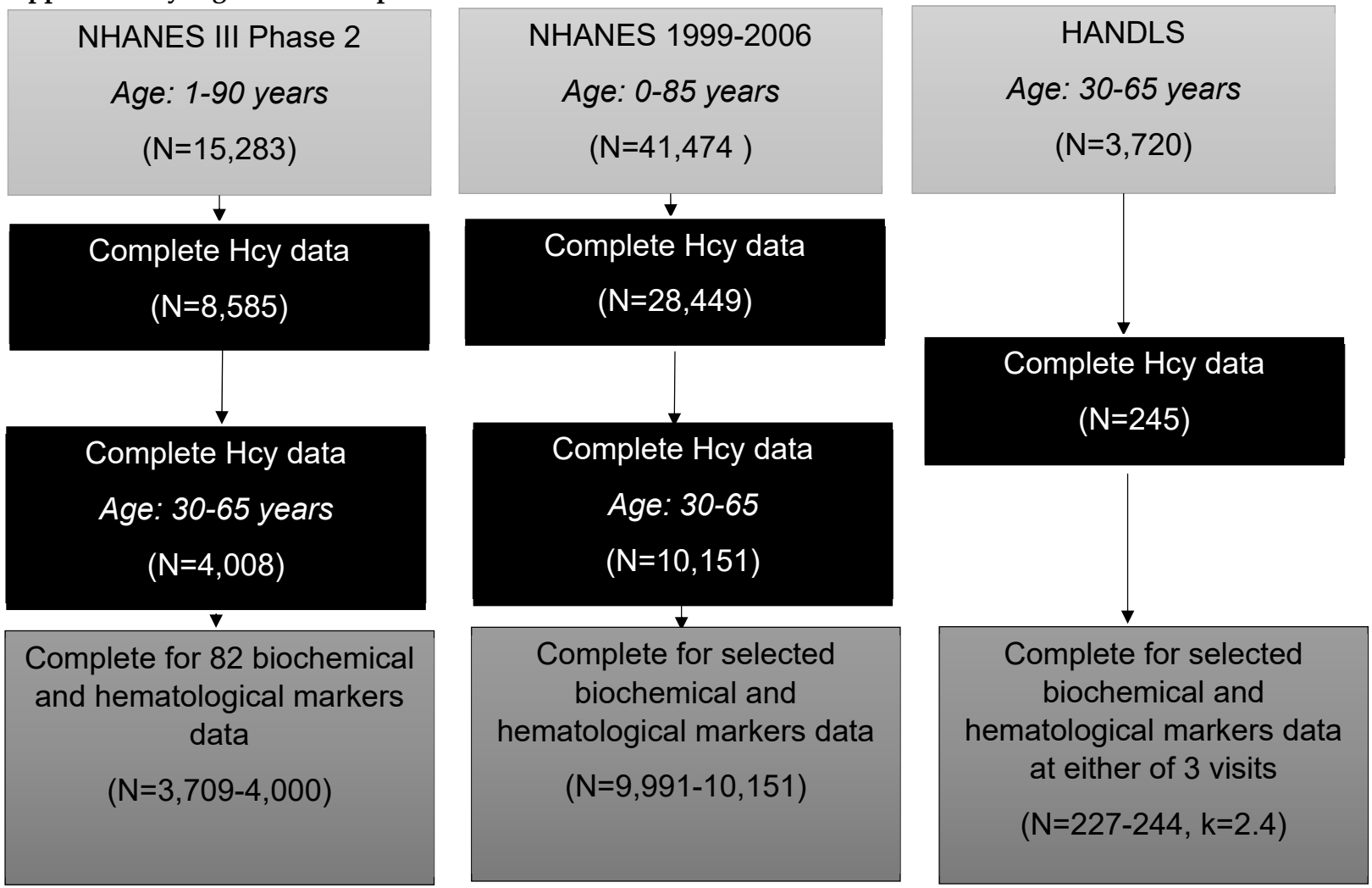
52. Margalit, I.; Cohen, E.; Goldberg, E.; Krause, I. Reconsidering the relation between serum homocysteine and red blood cell distribution width: A cross-sectional study of a large cohort. *Biomarkers* **2018**, *23*, 483–486. [CrossRef] [PubMed]
53. Poupon, R. Liver alkaline phosphatase: A missing link between choleresis and biliary inflammation. *Hepatology* **2015**, *61*, 2080–2090. [CrossRef]
54. Loohuis, L.M.; Albersen, M.; de Jong, S.; Wu, T.; Luykx, J.J.; Jans, J.J.M.; Verhoeven-Duif, N.M.; Ophoff, R.A. The Alkaline Phosphatase (ALPL) Locus Is Associated with B6 Vitamer Levels in CSF and Plasma. *Genes (Basel)* **2018**, *10*, 8. [CrossRef]
55. Wannamethee, S.G.; Shaper, A.G. Cigarette smoking and serum liver enzymes: The role of alcohol and inflammation. *Ann. Clin. Biochem.* **2010**, *47*, 321–326. [CrossRef]
56. Bailey, R.L.; Looker, A.C.; Lu, Z.; Fan, R.; Eicher-Miller, H.A.; Fakhouri, T.H.; Gahche, J.J.; Weaver, C.M.; Mills, J.L. B-vitamin status and bone mineral density and risk of lumbar osteoporosis in older females in the United States. *Am. J. Clin. Nutr.* **2015**, *102*, 687–694. [CrossRef]
57. Nakamura, Y.; Suzuki, T.; Kato, H. Serum bone alkaline phosphatase is a useful marker to evaluate lumbar bone mineral density in Japanese postmenopausal osteoporotic women during denosumab treatment. *Ther. Clin. Risk Manag.* **2017**, *13*, 1343–1348. [CrossRef]
58. Kim, D.B.; Oh, Y.S.; Yoo, K.D.; Lee, J.M.; Park, C.S.; Ihm, S.H.; Jang, S.W.; Shim, B.J.; Kim, H.Y.; Seung, K.B.; et al. Passive smoking in never-smokers is associated with increased plasma homocysteine levels. *Int. Heart J.* **2010**, *51*, 183–187. [CrossRef]
59. Lee, W.; Lee, S.; Roh, J.; Won, J.U.; Yoon, J.H. The Association between Involuntary Smoking Exposure with Urine Cotinine Level and Blood Cadmium Level in General Non-Smoking Populations. *J. Korean Med. Sci.* **2017**, *32*, 568–575. [CrossRef]
60. Sanchez-Rodriguez, J.E.; Bartolome, M.; Canas, A.I.; Huetos, O.; Navarro, C.; Rodriguez, A.C.; Arribas, M.; Esteban, M.; Lopez, A.; Castano, A. Anti-smoking legislation and its effects on urinary cotinine and cadmium levels. *Environ. Res.* **2015**, *136*, 227–233. [CrossRef] [PubMed]

ON-LINE SUPPLEMENTARY MATERIAL

Biochemical and hematological correlates of elevated homocysteine in national surveys and a longitudinal study of urban adults

Beydoun et. al.

Supplementary Figure 1. Participant flowchart



Abbreviations: k=Average number of repeats/participant; HANDLS=Healthy Aging in Neighborhoods of Diversity Across the Life Span; Hcy=Homocysteine; N=number of participants; NHANES=National Health and Nutrition Examination Surveys.

SUPPLEMENTAL METHODS 1: BIOCHEMICAL AND HEMATOLOGICAL INDICES: NHANES III, phase 2: 1991-1994

BIOMARKER	DESCRIPTION (UNITS)
hopsi	Serum homocysteine: SI (umol/L)
cop	Serum cotinine (ng/mL)
vdpsi	Serum vitamin D: SI (nmol/L)
t4psi	Serum throxine: SI (nmol/L)
thpsi	Serum thyroid stim hormone: SI (mU/L)
tmp	Serum antimicrosomal antibody (U/mL)
tap	Serum anti-thyroglobulin antibody (U/mL)
wcpsi	White blood cell count: SI
lmppcnt	Lymphocyte percent (Coulter)
moppcnt	Mononuclear percent (Coulter)
grppcnt	Granulocyte percent (Coulter)
lmp	Lymphocyte number (Coulter)
mop	Mononuclear number (Coulter)
grp	Granulocyte number (Coulter)
rcpsi	Red blood cell count: SI
hgpsi	Hemoglobin: SI (g/L)
htpsi	Hematocrit: SI (L/L = 1)
mvpsi	Mean cell volume: SI (fL)
mcpsi	Mean cell hemoglobin: SI (pg)
mhpsi	Mean cell hemoglobin concentration: SI
rwpsi	Red cell distribution width: SI (fraction)
plpsi	Platelet count: SI
dwp	Platelet distribution width (%)
pvpsi	Mean platelet volume: SI (fL)
pbpsi	Lead: SI (umol/L)
eppsi	Erythrocyte protoporphyrin: SI (umol/L)
fepsi	Serum iron: SI (umol/L)
tipsi	Serum TIBC: SI (umol/L)
pxp	Serum transferrin saturation (%)
frpsi	Serum ferritin: SI (ug/L)
fopsi	Serum folate: SI (nmol/L)
rbpsi	RBC folate: SI (nmol/L)
vbpsi	Serum vitamin B12: SI (pmol/L)
vcpsi	Serum vitamin C: SI (mmol/L)
icpsi	Serum normalized calcium: SI (mmol/L)
capsi	Serum total calcium: SI (mmol/L)

sepsi	Serum selenium: SI (nmol/L)
vapsi	Serum vitamin A: SI (umol/L)
vepsi	Serum vitamin E: SI (umol/L)
acpsi	Serum alpha carotene: SI (umol/L)
bcpsi	Serum beta carotene: SI (umol/L)
bxpsi	Serum beta cryptoxanthin: SI (umol/L)
lupsi	Serum lutein/zeaxanthin: SI (umol/L)
lypsi	Serum lycopene: SI (umol/L)
repsi	Serum sum retinyl esters: SI (umol/L)
tcpsi	Serum cholesterol: SI (mmol/L)
tgpsi	Serum triglycerides: SI (mmol/L)
hdpsi	Serum HDL cholesterol: SI (mmol/L)
crp	Serum C-reactive protein (mg/dL)
ahp	Serum hepatitis A antibody
hbp	Serum hepatitis B core antibody
hcp	Serum hepatits C antibody
rupunit	Serum rubells antibody (IU)
napsi	Serum sodium: SI (mmol/L)
skpsi	Serum potassium: SI (mmol/L)
clpsi	Serum chloride: SI (mmol/L)
c3psi	Serum bicarbonate: SI (mmol/L)
scpsi	Serum total calbicarcium: SI (mmol/L)
pspsi	Serum phosphorus: SI (mmol/L)
uapsi	Serum uric acid: SI (umol/L)
sgpsi	Serum glucose: SI (mmol/L)
bupsi	Serum blood urea nitrogen: SI (mmol/L)
tbpsi	Serum total bilirubin: SI (umol/L)
cepsi	Serum creatinine: SI (umol/L)
sfpsi	Serum iron: SI (umol/L)
chpsi	Serum cholesterol: SI (mmol/L)
trpsi	Serum triglycerides: SI (mmol/L)
aspsi	Aspartate aminotransferase: SI (U/L)
atpsi	Alanine aminotransferase: SI (U/L)
ggpsi	Gama glutamyl transferase: SI (U/L)
ldpsi	Serum lactate dehydrogenase: SI (U/L)
appsi	Serum alkaline phosphatase: SI (U/L)
tppsi	Serum total protein: SI (g/L)
ampsi	Serum albumin: SI (g/L)
gbpsi	Serum globulin: SI (g/L)
ospsi	Serum osmolality: SI (mmol/Kg)
ghp	Glycated hemoglobin: (%)

g1p	Plasma glucose (mg/dL)
g1psi	Plasma glucose: SI (mmol/L)
udpsi	Urinary cadmium: SI (nmol/L)
urpsi	Urinary creatinine: SI (mmol/L)
ubp	Urinary albumin (ug/mL)
uip	Urinary iodine (ug/dL)

Homocysteine

Homocysteine concentrations were measured as part of an NHANES III surplus sera project on serum samples from participants 12 years of age or older who were seen during phase II of this survey (1991–1994). This project is described in greater detail elsewhere[1]. Homocysteine concentrations were measured at the Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University by using the high-performance liquid chromatography method of Araki and Sako[2]. The interassay coefficient of variation for this assay was 6%.

Hematological indices

Red blood, White blood and platelet cell indices

Microscopic examination (manual differential) of the peripheral blood spread on a glass slide utilized a stained blood film to perform a differential leukocyte count, evaluate red cell morphology, and estimate number of platelets. Manual differential variables include segmented neutrophils, lymphocytes, monocytes, eosinophils, basophils, blasts, promyelocytes, metamyelocytes, myelocytes, bands, atypical lymphocytes, anisocytosis, basophilic stippling, hypochromia, poikilocytosis, polychromatophilia, macrocytosis, microcytosis, sickle cells, spherocytosis, target cells, toxic granulation, and vacuolated cells (GRPDIF, LMPDIF, MOPDIF, EOP, BAP, BOP, BLP, PRP, MEP, MLP, BAP, LAP, ANP, BSP, HZP, PKP, POP, MRP, MIP, SIP, SHP, TTP, TXP, and VUP).

In NHANES III, a manual differential was performed on a special subsample of examinees aged one year and older. This manual differential was used for internal quality control purposes and to confirm abnormal hematology results. This subsample was defined as a random 10-percent sample of all examined persons plus all examinees who had a predetermined high or low value for one or more of the following hematologic assessments: white blood cell count (WBC), red blood cell count (RBC), hemoglobin, hematocrit, mean cell volume (MCV), mean cell hemoglobin (MCH), mean cell hemoglobin concentration (MCHC), red blood cell distribution width (RDW), platelet count, mean platelet volume (MPV), lymphocyte percentage, mononuclear percentage, or granulocyte percentage. A table of predetermined high and low values for WBC, RBC, hemoglobin, hematocrit, MCV, MCH, MCHC, RDW, platelet count, MPV, lymphocyte percentage, mononuclear percentage, and granulocyte percentage is located in the Manual for Medical Technicians[3].

RCP: Red blood cell count
See notes for HGP and GRP.

VUP: Vacuolated cells

See note for ANP.

WCP: White blood cell count
See note for HGP and GRP.

(A) Biochemical indices

Blood micronutrient and electrolyte biomarkers

Serum 25-hydroxyvitamin D (25(OH)D)

The INCSTAR 25(OH)D assay consists of a two-step procedure, with the first being a rapid extraction of 25(OH)D and other hydroxylated metabolites from serum or plasma with acetonitrile, and the second being the assay of the treated sample by using an equilibrium RIA procedure. [4]

Serum ferritin

Ferritin, like hemoglobin, is a major iron storage protein. Serum ferritin levels increase as a result of iron overload, aging, infection, inflammation, liver disease, juvenile rheumatoid arthritis, leukemia, and Hodgkin's disease; and decrease as a result of iron deficiency. Ferritin is measured by using the Bio-Rad Laboratories' "QuantImune Ferritin IRMA" kit, a single-incubation two-site immunoradiometric assay (IRMA) based on the general principles of assays. [4,5]

Serum folate and vitamin C

Folate is required in cellular metabolism and hematopoiesis, and prolonged deficiency leads to megaloblastic anemia. In NHANES III, serum folate is measured by using the Bio-Rad Laboratories "Quantaphase Folate" radioassay kit.[4,6]

Serum normalized calcium

To measure the total calcium concentration, protein-bound and complexed calcium must be released.[4] Within the pH range of 6.9-8.0, the NOVA 7 analyzer can predict the normalized calcium level within 2% of the theoretical value.[7]

Serum selenium

Serum selenium is measured by atomic absorption spectrometry[4] in a procedure based on Lewis et al.[8] and Paschal and Kimberly.[9]

Serum vitamin E, retinol, retinyl esters, carotenoids

Serum levels of vitamin A (retinol), vitamin E (α-tocopherol), retinyl esters and carotenoids are measured by isocratic high performance liquid chromatography with detection at three wavelengths, namely 300, 325, and 450 nm. Quantitation is accomplished by comparison of peak heights with a standard solution.[4]

LUP: Serum lutein/zeaxanthin

The lower limit of detection (LOD) for lutein/zeaxanthin was 0.43 ug/dL. Using the LOD coding formula (detection limit divided by the square root of two), the calculated value indicating that the serum lycopene results were below the level of detection would be 0.30. After rounding, the value of 0 (zero) was placed in the results field to indicate that the serum lutein/zeaxanthin was below 0.43 ug/dL.

LYP: Serum lycopene

The lower limit of detection (LOD) for lycopene was 0.63 ug/dL. Using the LOD coding formula (detection limit divided by the square root of two), the calculated value indicating that the serum lycopene results were below the level of detection would be 0.44. After rounding, the value of 0 (zero) was placed in the results field to indicate that the serum lycopene was below 0.63 ug/dL.

FEP: Serum iron

Laboratory methods differed between NHANES III and previous surveys. Therefore, results may not be comparable between surveys. Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996).

PXP: Serum transferrin saturation

This value was calculated as (FEP/ TIP) * 100.

RBP: RBC folate

See note for FOP.

SEP: Serum selenium

Selenium values were measured on two Perkin-Elmer graphite furnace atomic absorption spectrophotometers (model 3030 and model 5100) during the six-year study. Based on a comparability study using linear models, the results generated using the Model 5100 instrument (from 12/07/90 to 1/13/95) were on average 4.3 percent higher than those from the Model 3030 instrument (used from 10/1/88 to 12/06/90).Since the Model 5100 represented more precise measurements, the model 3030 data were adjusted to make them comparable to the Model 5100. Perkin-Elmer Model 5100 Zeeman-corrected graphite furnace atomic absorption spectrophotometer testing began on 12/07/90. All selenium values measured prior to 12/07/90 were adjusted to the AA5100 values. The formula used was:
New value = 16.795 + 0.902 * original value.

SFP: Serum iron

This value was obtained from the standard battery of biochemical assessments. Use of the laboratory test result from the reference method (FEP), rather than the SFP value, is generally recommended. For most analyses of serum iron, the appropriate variable to use will be FEP. The value from the biochemistry profile (SFP) should not be used routinely. Consult the Laboratory Procedures Used for NHANES III

(U.S. DHHS, 1996) for details. Laboratory test results for SFP were added to the protocol after NHANES III began. This result field was blank-filled for examinees who were examined prior to the start of testing.

VCP: Serum vitamin C

For NHANES III, serum concentrations of vitamin C were measured using a total vitamin C, fully reduced method using high-performance liquid chromatography with electrochemical detection (HPLC-EC) analysis.

VEP: Serum vitamin E
The vitamin E value of 9999 was confirmed.

Blood metabolic and inflammatory indices

Thyroid hormones

Total (protein-bound and free) circulating thyroxine (T_4) concentrations were determined using an enzyme-based homogeneous immunoassay on the Hitachi 704. The reference range is 4.5 to 12.5 μg/dL. The coefficient for inter-assay coefficient of variation (CV) was less than 10%. [4]

The thyroid stimulating hormone (TSH, in mu/L) assay is a chemiluminescence immunometric assay utilizing a mouse monoclonal antibody to TSH immobilized on a polystyrene bead and a goat polyclonal antibody to TSH conjugated with an acridinium ester. The reference range is 0.3 to 5.0 mu/L. The coefficient for inter-assay CV was less than 5%. [4]

Serum lipids: total cholesterol (TC), triglycerides (TG) and high-density lipoprotein-Cholesterol (HDL-C)

Cholesterol is measured enzymatically in serum or plasma in a series of coupled reactions that hydrolyze cholesterol esters and oxidize the 3-OH group of cholesterol. Triglycerides are measured using a series of coupled reactions in which triglycerides are hydrolyzed to produce glycerol. HDL-cholesterol is measured following the precipitation of the other lipoproteins with a polyanion/divalent cation mixture. [4]

CHP: Serum cholesterol
This value was obtained from the standard battery of biochemical assessments. Use of the laboratory test result from the reference method (TCP), rather than the CHP value, is generally recommended.For most analyses of serum cholesterol, the appropriate variable to use will be TCP. The value from the biochemistry profile (CHP) should not be used routinely. Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996) for the details.

LCP: Serum LDL cholesterol calculation
The value for LDL was calculated by the Friedewald equation as follows:
LDL = total cholesterol - high density cholesterol - triglyceride/5.
Because the equation is not valid when serum triglyceride values exceed 400 mg/dL, the LDL is missing when serum triglyceride (TGP) exceeds 400 mg/dL.
Serum LDL was calculated on examinees who were instructed to fast

(ages 12 and older) and who did fast at least nine hours, were examined in the morning, and were randomly assigned to the morning fasting sample (WTPFHSD6 > 0). Therefore, LDL would be blank if examinees were aged less than 12 years, fasted fewer than nine hours, were examined in an afternoon or evening session, or were not randomly assigned to the morning session. For the purpose of this calculation, the number of hours fasted was rounded to the nearest whole integer.
For more information regarding this equation, refer to the Laboratory
Procedures Used for NHANES III (U.S. DHHS, 1996).

TGP: Serum triglycerides
Serum triglyceride levels were measured regardless of the examinee's fasting status. Mean serum triglycerides and the distribution of serum triglycerides should be estimated only on examinees who did fast at least nine hours, were examined in the morning, and were randomly assigned to the morning fasting sample (WTPFHSD6 > 0). For the purpose of this calculation, the number of hours fasted was rounded to the nearest whole integer. Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996) for details.

TRP: Serum triglycerides
This value was obtained from the standard battery of biochemical assessments. Use of the laboratory test result from the reference method (TGP), rather than the TRP value, is generally recommended. For most analyses, the appropriate variable to use is TGP. The value from the biochemistry profile (TRP) should not be used routinely. Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996) for details. Results for TRP were added to the protocol after NHANES III began. This result field was blank-filled for examinees who were examined prior to the start of testing.

Glucose biomarkers

Serum C-peptide (first venipuncture)

The specimen for this assay was obtained at the time of the initial venipuncture. This result is available for all six years of the survey. Examinees aged 40-74 years who used insulin were excluded from the OGTT. A first venipuncture was obtained, but the glucose challenge and second venipuncture were canceled. In these instances, the variables G1P, C1P and I1P have a value, but the results G2P, C2P and I2P from the second venipuncture are blank-filled to indicate a medical exclusion.

C2P: Serum C-peptide (second venipuncture)

Post-glucose challenge levels of C-peptide and insulin for examinees who had an OGTT were measured only during 1991-1994.

G1P: Plasma glucose (first venipuncture)

Plasma glucose was measured using the reference method on examinees aged 20 years and older. Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996) for details. During NHANES III, OGTT testing was conducted on MEC examinees aged
40-74 years. A random assignment was made prior to conducting the OGGT to determine who should receive a morning examination[3,10]. As a result, approximately half of the OGGT examinees received the morning OGTT after an overnight fast. This subsample most closely conformed to the World Health Organization (WHO) criteria for OGTT testing to identify diabetes[11]. Therefore, this morning subsample is the NHANES III subsample that should be used to estimate the prevalence of diabetes and impaired glucose tolerance. People who reported a medical history of diabetes but who were not using insulin therapy were asked to conform to the fasting instructions for their examination session and were eligible for an OGTT if the age criteria were satisfied. The morning sample weights (WTPFHSD6) for total NHANES III weights for the morning OGTT subsample should be used when weighting these data to generate national estimates. Data from the afternoon and evening OGTTs do not conform to the WHO protocol for diagnosing diabetes or IGT and should not be used for these purposes.

If an examinee was given an OGTT during an examination session other than the session assigned, that examinee's sample weight for the assigned session will be zero. For example, if an examinee was selected for a morning OGTT but was tested in the afternoon, the examinee's morning sample weight for the OGTT will be zero.

GHP: Glycated hemoglobin (HbA1c)
Glycohemoglobin measurements for NHANES III were performed by the Diabetes Diagnostic Laboratory at the University of Missouri -- Columbia using the Diamat Analyzer System (Bio-Rad Laboratories, Hercules, CA). This ion-exchange HPLC system measures HbA1c (a specific glycohemoglobin) and has demonstrated excellent, long-term precision (interassay CV's 2.0). It was standardized to the reference method that was used for the Diabetes Control and Complications Trial (DCCT). Variant hemoglobins, including hemoglobin C, D, F, and elevated HbF, can interfere with HbA1c measurement by the Diamat HPLC. Hemoglobin S in its heterozygous state does not interfere with this assay. Although interferences usually can be detected by an abnormal Diamat chromatogram, HbA1c results for these specimens were not considered valid. Therefore, samples containing hemoglobin variants or elevated HbF or samples that produce chromatograms indicating hemoglobin degradation were analyzed by an alternate method that used affinity chromatography to separate glycohemoglobin. Affinity chromatographic methods were not affected by the presence of hemoglobin variants and were less sensitive to hemoglobin degradation due to improper sample handling. The affinity method used also was standardized to the DCCT reference method. Reasons for using the affinity method for an examinee's test included an extra peak on the chromatogram, hemoglobin C, elevated hemoglobin F, or other abnormal hemoglobin.

I1P: Serum insulin (first venipuncture)

This is the adjusted insulin value for examinees. Most of the Insulin values in NHANES III (1988-1991) were adjusted because the manufacturer of the laboratory testing kits changed during that period. An indicator of the kit number is located in the I1P2PFLG field (i.e., 1 = Kit 1, 2 = Kit 2, and 3 = Kit 3). All insulin values from Kit 1 and Kit 2 assays were adjusted linearly to match the Kit 3 numbers. Further information on this adjustment procedure is available in the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996).

The equations used to adjust the data were:
Kit 3 = 0.787 (Kit 1) + 0.832 Equation 1
Kit 3 = 0.597 (Kit 2) + 1.746 Equation 2

The following steps were used to make the adjustment:

1. Equation 1 was applied to group 1 (Kit 1) data
2. Equation 2 was applied to group 2 (Kit 2) data
3. Group 3 data (Kit 3) were left unchanged.

The time periods for the insulin kits were as follows:

Group	Assay Period	Assay Method
1	10/88-01/05/90	Kit 1
2	01/06/90-09/06/90	Kit 2
3	11/01/90-end of study	Kit 3

I2P: Serum insulin (second venipuncture)

SGP: Serum glucose
This value was obtained from the standard battery of biochemical assessments. Use of the laboratory test result for plasma
glucose from the reference method (G1P), rather than the SGP value, is generally recommended. For most analyses, the appropriate variable to use will be G1P. The value from the biochemistry profile (SGP) should not be used routinely.Consult the Laboratory Procedures Used for NHANES III (U.S. DHHS, 1996) for details.

Protein biomarkers

GBP: Serum globulin
Globulin results were added to the protocol after NHANES III began. This result field was blank-filled for examinees who were examined prior to the start of testing.

HGP: Hemoglobin

In NHANES I, NHANES II, and HHANES, determinations of red and white blood cell counts were made using a semiautomated cell counter Coulter model FN). Determinations of hemoglobin concentration (Hb) were made using a Coulter hemoglobinometer, and determinations of packed cell volume (PCV) were made using the microhematocrit centrifuge method. The hematologic indices MCH, MCHC, and MCV were calculated as follows:

MCH = Hb/RBC
MCHC= Hb/PCV
MCV = PCV/RBC

In NHANES III, these hematologic parameters were determined by using a fully automated Coulter S+JR hematology analyzer. These analyzers measured the mean (red) cell volume (MCV) directly, utilizing a process of continuous integration of pulse heights divided by the pulse number; PCV values were calculated through the multiplication of MCV and RBC.

Although it has been shown that identified errors in the microhematocrit method caused by plasma trapping and red cell dehydration approximately compensate each other,[12] packing errors can occur in macrocytic anemia and can be considerable in sickle cell anemia, spherocytosis, and thalassemias[13]. Therefore, individual values for MCV, PCV ("hematocrit"), and MCHC from NHANES III cannot be compared directly to values from the previous NHANES.

HTP: Hematocrit

See note for HGP.

MCPSI: Mean cell hemoglobin
See note for HGP.
MHP: Mean cell hemoglobin concentration
See note for HGP.

CEP: Serum Creatinine

The Cleveland Clinic Foundation (CCF) laboratory analyzed the serum creatinine specimens using a Roche coupled enzymatic assay (creatininase, creatinase, sarcosine oxidase, kits # 1775677 and 1775766) performed on a Roche P Module instrument. The Roche method calibrators were traceable to an isotope dilution mass spectrometric method for serum creatinine using standard references methods (NIST SRM 967) and confirmed by analysis of CAP LN-24 linearity set based on NIST assigned values. Serum creatinine by the Roche method was then compared to the original NHANES III measurements which used the Jaffe kinetic alkaline picrate method performed on a Roche Hitachi 737 analyzer. There were significant differences in results between these two measurements. The comparison of values revealed the mean (SD) serum creatinine at NHANES, CCF, and their difference were 1.177 (0.315), 0.947 (0.302), and 0.231(0.066) mg/dL, respectively (paired t-test, $p<0.0001$). The Deming regression (adjusting for errors in measurement) for the correction is Standard Creatinine (Y) = 0.960*NHANES Creatinine (X) – 0.184 (r = 0.978).

IgG to infections

RUP: Serum rubella antibody

Rubella antibody data are reported both as an optical density index
and in International Units. The index was calculated by subtracting the absorbance of the control well from the absorbance of the antigen well (AG-NS) and dividing the difference by the cut-off value. The

cut-off value was calculated as the mean AG-NS value of duplicate 10 IU standards. The equation used was: O.D. index = (AG-NS)/Cut-off value
An O.D. index greater than or equal to one indicates the presence of antibody.

RUPUNIT: Serum rubella antibody (IU)

Rubella antibody data are reported both as an optical density index and in International Units. International Units were calculated based on a standard curve using a regression analysis of duplicate AG-NS values of 10, 40, & 100 IU standards and their squares. An International Unit value greater than or equal to 10 indicates the presence of antibody.

SAP: Serum hepatitis B surface antigen
See note for HBP.

SSP: Serum hepatitis B surface antibody
See note for HBP.

TOP: Serum toxoplasmosis antibody
The presence and quantity of antibody to Toxoplasma gondii in the test sample were determined by comparing the optical density of the test sample to a standard curve. A standard curve was constructed using optical density readings from positive control sera obtained from a kit; these readings were calibrated to WHO Toxo 60 serum and read as International Units (IU/mL). Those test samples exhibiting titer below 7 IU/mL indicated a non-significant level of antibody according to this technique; thus, they were considered to be negative, indicating no infection. Those test samples with results greater than 6 IU/mL were considered to be positive, indicating infection at some undetermined time.

VRP: Serum varicella antibody
Varicella antibody data were reported as an optical density index. See note RUP for the index calculation. The equation used was:
O.D. index = (AG-NS)/Cut-off value
The cut-off value was 0.1. An O.D. index equal to or greater than one indicates the presence of antibody.

Other inflammatory markers

MEP: Metamyelocyte cells
See note for ANP.
MIP: Microcytosis
See note for ANP.
MLP: Myelocyte cells

See note for ANP.
MOP: Mononuclear number
See note for GRP.
MOPDIF: Monocyte cells
See note for ANP.

MRP: Macrocytosis
See note for ANP.
MVPSI: Mean cell volume
See note for HGP.

PLP: Platelet count
See note for GRP.
POP: Polychromatophilia
See note for ANP.
PRP: Promyelocyte cells
See note for ANP.

Liver enzymes

The kinetic assay analyzer computer uses absorbance measurements to calculate alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), BUN, creatinine, gamma glutamyl transaminase (G-GT), and lactate dehydrogenase (LDH) concentrations.

(a) Alkaline phosphatase (ALP)
The method is linear up to 1000 U/L. When reanalyzing any specimen with a concentration greater than 1000 U/L, prepare a 10-fold (1+9) dilution of the specimen with saline.
The minimum detection limit, based on a linear regression curve of certified material analyzed 20 times, is 2 U/L. Results below the detection limit are reported as <2 U/L.

(b) Alanine aminotransferase (ALT)
The method is linear up to 400 U/L. When reanalyzing any specimen with a concentration greater than 400 U/L, prepare a 10-fold (1+9) dilution of the specimen with saline. The minimum detection limit, based on a linear regression curve of certified material analyzed 20 times, is 1 U/L. Results below the detection limit are reported as <1 U/L.

(c) Aspartate aminotransferase (AST)
The method is linear up to 800 U/L. When reanalyzing any specimen with a concentration greater than 800 U/L, prepare a 10-fold (1+9) dilution of the specimen with saline. The results must then be multiplied by 10 to account for this dilution. The minimum detection limit, based on linear a regression curve of certified material analyzed 20 times, is 1 U/L. Results below the detection limit are reported as <1 U/L.

(d) ASPSI: Aspartate aminotransferase

α-Ketoglutarate reacts with L-aspartate in the presence of AST to form L-glutamate plus oxaloacetate. The indicator reaction uses the oxaloacetate for a kinetic determination of NADH consumption. The International Federation of Clinical Chemistry (IFCC) has now recommended standardized procedures for ALT determination, including 1) optimization of substrate concentrations, 2) the use of Tris buffers, 3)

preincubation of a combined buffer and serum solution to allow side reactions with NADH to occur, 4) substrate start (α -ketoglutarate), and 5) optimal pyridoxal phosphate activation.

(e) Gamma glutamyl transaminase (γ-GT)

The method is linear up to 1200 U/L. When reanalyzing any specimen with a concentration greater than 1200 U/L, prepare a 10-fold (1+9) dilution of the specimen with saline. The minimum detection limit, based on a linear regression curve of certified material analyzed 20 times, is 2 U/L. Results below the detection limit are reported as <2 U/L.

(f) Lactate dehydrogenase (LDH)
The method is linear up to 1000 U/L. When reanalyzing any specimen with a concentration greater than 1000 U/L, prepare a 10-fold (1+9) dilution of the specimen with saline. The minimum detection limit, based on a linear regression curve of certified material analyzed 20 times, is 2 U/L. Results below the detection limit are reported as <2 U/L.

Urinary biomarkers

Urinary iodine (IU) is measured by using the reduction-oxidation reaction between ceric and arsenite catalyzed by iodide. Therefore, UI is proportional to its catalytic activity. Urine samples, controls, and iodate standard are digested with chloric acid, then measured spectrometrically at 420 nm with a Technicon AutoAnalyzer. Calculations are based on an iodine standard curve. The final concentration of UI is expressed in μg/dL and is easily converted to μg UI per gram urinary creatinine.[4,14,15]

Urinary creatinine was analyzed with a Jaffé rate reaction using an ASTRA analyzer.[4] Creatinine is released into the plasma at a relatively constant rate and thus has a constant amount per unit muscle mass. Consequently, creatinine is the best indicator of impaired kidney function.

A solid-phase fluorescent immunoassay (FIA) was used for the measurement of urinary albumin.[16] Increased microalbuminuria is a sign of renal disease and is predictive of nephropathy risk in type 2 diabetes patients. [4,16]

Cadmium analysis is used to identify toxicity. Occupational exposure is the most common cause of elevated cadmium levels.[4] Cadmium is measured in urine by atomic absorption spectrometry by using a modification of the method described by Pruszkowska et al. [17]

URP: Urinary creatinine
Although the laboratory method detection limit for urinary creatinine is 1 mg/dL, all values below 10 mg/dL were considered "statistically suspect" and were coded as "below level of detection".

SUPPLEMENTAL METHODS 2: BIOCHEMICAL AND HEMATOLOGICAL INDICES SELECTED FOR NHANES 1999-2006

Total homocysteine (tHcy)

Total homocysteine (tHcy) in plasma is measured by the "Abbott Homocysteine (HCY) assay",[18] a fully automated fluorescence polarization immunoassay (FPIA) from Abbott Diagnostics performed on the AxSYM® platform.[19] In brief, dithiothreitol (DTT) reduces homocysteine bound to albumin and to other small molecules, homocystine, and mixed disulfides, to free thiol. S-adenosyl-homocysteine (SAH) hydrolase catalyzes conversion of homocysteine to SAH in the presence of added adenosine. In the subsequent steps, the specific monoclonal antibody and the fluoresceinated SAH analog tracer constitute the FPIA detection system.[20] Plasma total homocysteine concentrations are calculated by the Abbott AxSYM® using a machine-stored calibration curve.

An international round robin performed in 1998[21] demonstrated that the Abbott method is fully equivalent to other most frequently used methods in this field (i.e., HPLC-FD, HPLC-ED, GC/MS). Thus, the Abbott Homocysteine (HCY) assay will be used as primary method for the determination of plasma total homocysteine in NHANES 1999+. For NHANES 1999-2001, the Abbott IMx® analyzer was used, starting NHANES 2002, the Abbott AxSYM® analyzer is used. The IMx® and the AxSYM® platforms are both using the same reagent kit, but the AxSYM® is a newer fully-automated analyzer that can measure multiple analytes during one run. Pernet et al. showed that the two platforms agree well. [19]

This method is linear for homocysteine in the range 0.8-50 μmol/L. Samples with results <2 μmol/L or >15 μmol/L are reanalyzed for confirmation before results are released. Samples with total homocysteine concentrations >50 μmol/L are diluted 10-fold with PBS or FPIA buffer and reanalyzed. This method has a total coefficient of variation in the range of 3-6%.

Folate/Vitamin B12 (Serum and whole blood)

Both vitamins are measured by using the Bio-Rad Laboratories "Quantaphase II Folate/vitamin B12" radioassay kit [22,23]. The assay is performed by combining serum or a whole blood hemolysate sample with ^{125}I-folate and ^{57}Co-vitamin B12 in a solution containing dithiothreitol (DTT) and cyanide. The mixture is boiled to inactivate endogenous folate-binding proteins and to convert the various forms of vitamin B12 to cyanocobalamin. The reduced folate and its analogs are stabilized by DTT during the heating. The mixture is cooled and then combined with immobilized affinity-purified porcine intrinsic factor and folate-binding proteins. The addition of these substances adjusts and buffers the pH of the reaction mixture to 9.2. The reaction mixture is then incubated for 1 hour at room temperature.

Creatinine

The LX20 modular chemistry side uses the Jaffe rate method (kinetic alkaline picrate) to determine the concentration of creatinine in serum, plasma, or urine. A precise volume of sample is introduced into a reaction cup containing an alkaline picrate solution. Absorbance readings are taken at both 520 nm and

560 nm. Creatinine from the sample combines with the reagent to produce a red color complex. The observed rate measurement at 25.6 seconds after sample introduction has been shown to be a direct measure of the concentration of the creatinine in the sample.

Uric Acid

The LX20 uses a timed endpoint method to measure the concentration of uric acid in serum, plasma, or urine. Uric acid is oxidized by uricase to produce allatoin and hydrogen peroxide. The hydrogen peroxide reacts with 4-aminoantipyrine (4-AAP) and 3.5-dichloro-2-hydroxybenzene sulfonate (DCHBS) in a reaction catalyzed by peroxidase to produce a colored product. The system monitors the change in absorbance at 520 nm at a fixed time interval. The change in absorbance is directly proportional to the concentration of uric acid in the sample.

MCH and RDW

Complete Blood Count (CBC) with Five-Part Differential NHANES 2005–2006

The Beckman Coulter method of sizing and counting particles uses measurable changes in electrical

resistance produced by nonconductive particles suspended in an electrolyte.

A suspension of blood cells passes through a small orifice simultaneously with an electric current. A small opening (aperture) between electrodes is the sensing zone through which suspended particles pass. In the sensing zone, each particle displaces its volume of electrolyte. Beckman Coulter measures the displaced volume as a voltage pulse, the height of each pulse being proportional to the volume of the particle.

The quantity of suspension drawn through the aperture is for an exact reproducible volume. Beckman Coulter counts and sizes individual particles at a rate of several thousand per second. This method is independent of particle shape, color, and density.

The MAXM is a quantitative, automated, differential cell counter for in vitro diagnostic use.

The MAXM measures these parameters in whole blood: Cell	Parameter	Measured	Pulse size wavelength calculation	Reported units
WBC	white blood cell count	WBC bath	≥35 fL	n × 103 cells/µL
RBC	red blood cell count	RBC bath	36–360 fL	n × 106 cells/µL

Hgb	Hemoglobin concentration	WBC bath	525 nm	g/dL
Hct	hematocrit	computed	RBC x MCV/10	%
MCV	mean cell volume	derived from RBC histogram	# × size of RBC/total RBC	fL
MCH	mean cell hemoglobin	computer	Hgb/RBC × 10	pg
MCHC	mean cell hemoglobin concentration	computed	Hgb/Hct × 100	g/dL
RDW	red cell distribution width	derived from RBC histogram	CV expressed in % of the RBC size distribution	%
Plt	Platelet count	RBC bath	2 to 20 fL	*n* × 103 cells/μL
MPV	Mean platelet volume	derived from Plt histogram	Mean volume of Plt population under the fitted curve × constant	fL
NE%	neutrophil percent	derived from scatterplot	# cells inside NEarea/# cells inside total cell area × 100	%
NE #	neutrophil number	Absolute number	NE%/100 × WBC count	103 cells/μL

9. Beckman Coulter Reportable Range of Results

Parameter	**Linearity**	**Limits: The greater of**
WBC x 10^3 cells/ L	0 to 99.9	0.2 or 3.0%
RBC x 10^6 cells/ L	0 to 7.00	0.05 or 2.0%
Hgb g/dL	0 to 25.0	0.2 or 3.0%
MCV fL	50.0 to 150.0	2.0 or 3.0%
Plt x 10^3 cells/ L	0 to 999	10.0 or 7.0%
MPV fL	5.0 to 20.0	5.0%

Alkaline phosphatase

The DxC800 system uses a kinetic rate method using a 2-Amino-2-Methyl-1-Propanol (AMP) buffer to measure ALP activity in serum or plasma. In the reaction, the ALP catalyzes the hydrolysis of the colorless organic phosphate ester substrate, p-Nitrophenylphosphate, to the yellow colored product p-Nitrophenol and phosphate. This reaction occurs at an alkaline pH of 10.3. The system monitors the rate of change in absorbance at 410 nm over a fixed-time interval. This rate of change in absorbance is directly proportional to the ALP activity in the serum. (URL: https://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BIOPRO_G.htm)

Cotinine

Serum specimens and urine specimens are processed, stored, and shipped to the Division of Laboratory Sciences, National Center for Environmental Health, and Centers for Disease Control and Prevention for analysis.

Vials are stored under appropriate frozen (–20°C) conditions until they are shipped to National Center for Environmental Health for testing.

Cotinine is a major metabolite of nicotine that may be used as a marker for both active smoking, and as an index to Environmental Tobacco Smoke (ETS) exposure, or "passive smoking". Cotinine is generally preferred over nicotine for such assessments because of its substantially longer half-life. The half-life of cotinine in plasma has been estimated to be about 15–20 hrs; by contrast, the half-life of nicotine is only 0.5–3 hrs. Cotinine may be measured in serum, urine or saliva – the half-life of cotinine in all three fluids is essentially the same. Cotinine concentrations tend to be higher (3–8x) in urine than in serum; however, for studies requiring a quantitative assessment of exposure, plasma or serum is regarded as the fluid of choice. Therefore, serum was chosen for NHANES cotinine analyses.

Serum cotinine is measured by an isotope dilution-high performance liquid chromatography / atmospheric pressure chemical ionization tandem mass spectrometry (ID HPLC-APCI MS/MS). Briefly, the serum sample is spiked with methyl-D3 cotinine as an internal standard, and after an equilibration period, the sample is applied to a basified solid-phase extraction column. Cotinine is extracted off the column with methylene chloride, the organic extract is concentrated, and the residue is injected onto a short, C18 HPLC column. The eluant from these injections is monitored by APCI-MS/MS, and the m/z 80 daughter ion from the m/z 177 quasi-molecular ion is quantitated, along with additional ions for the internal standard, external standard, and for confirmation. Cotinine concentrations are derived from the ratio of native to labeled cotinine in the sample, by comparisons to a standard curve.

NNAL is measured by using liquid chromatography linked to tandem mass spectrometry (LC/MS/MS). For "total" NNAL assays, the urine sample is fortified with an NNAL-$^{13}C_6$ internal standard, and then hydrolyzed using β-glucuronidase in incubations for at least 24 hours. The samples are then extracted and cleaned up on a specially-designed solid-phase molecularly-imprinted polymer (MIP) column, after which the analyte is eluted and analyzed by LC/MS/MS, monitoring the m/z 210->180 native, and m/z 216->186 internal standard transition ions. NNAL concentrations are derived from the ratio of the integrated peaks of native to labeled ions by comparison to a standard calibration curve. Free NNAL measurements are conducted in a similar manner, but with the omission of prior enzymatic

hydrolysis. Bound NNAL (i.e. NNAL-Gluc) may be estimated from the difference of (Total NNAL – Free NNAL). This method has been described previously[24].

There were no changes (from the previous 2 years of NHANES) to equipment, lab methods, or lab site. Detailed instructions on specimen collection and processing can be found in the NHANES Laboratory/Medical Technologists Procedures Manual (LPM).

SUPPLEMENTAL METHODS 3: BIOCHEMICAL AND HEMATOLOGICAL INDICES SELECTED FOR HANDLS 2004-2018

Serum homocysteine

Serum homocysteine was measured on a sub-sample of HANDLS at visit 1 using a standard enzyme immunoassay conducted at the National Institute on Aging, Clinical Research Branch Core Laboratory, that is comparable to the method used for NHANES 1999 onwards.

Serum folate and vitamin B-12

Fasting blood was collected from each participant during MRV visits. Specimens in volumes of 2 mL serum were collected in small tubes and refrigerated. Serum folate and cobalamin were measured using chemiluminescence immunoassay by Quest Diagnostics, Chantilly, VA [25].

Serum creatinine

Using participant fasting venous blood specimens, baseline serum creatinine was measured at the National Institute on Aging, Clinical Research Branch Core Laboratory, using a modified kinetic Jaffe method (CREA method, Dade Dimension X-Pand Clinical Chemistry System, Siemens Healthcare Diagnostics Inc., Newark, DE) for a small group of participants (n=88); while the majority of participants (n=1,528) had baseline serum creatinine analyzed at Quest Diagnostics, Inc. by isotope dilution mass spectrometry (IDMS) (Olympus America Inc., Melville, NY) and standardized to the reference laboratory, Cleveland Clinic. While inter-assay coefficients of variation (CV) for this sample could not be calculated due to the use of only one or the other measurement of creatinine at baseline, only intra-assay CVs (mean/SD) could be estimated and those were 0.192 and 0.187 for the CREA and the IDMS methods, respectively. Serum creatinine was measured by Quest diagnostics in remaining waves.

RDW

RDW was measured by automated Coulter DXH 800 hematology analyzer as part of peripheral complete blood count (Beckman Coulter, Brea, CA), and was expressed as coefficient of variation (%) of red blood cell volume distribution. Regular calibration was performed every 3 months on the hematology analyzer and quality control was performed according to the manufacturer's recommendations.[26] There are usually two RDW measurements used for clinical purposes, namely the RDW-coefficient of variation (CV, unit: %), which we used in this study, and the RDW-Standard Deviation (SD, unit: fL) from which RDW-CV is derived. In fact, RDW-CV=RDW-SD×100/MCV, where MCV is the mean cell volume. The normal range for RDW-CV is 11.0 - 15.0%. Thus, the RDW-CV (%) depends on both the width of the distribution (normal range: 40-55 fL) curve and the MCV.[27]

MCH

The hematologic index MCH was calculated as follows: MCH = Hb/RBC. Using electronic cell sizing/cytometry/microscopy, Hb was assayed from a sample of 1 ml of blood drawn from participants after overnight fast and refrigerated up to 6 days (Quest diagnostics).

Serum uric acid (SUA)

SUA measurements are useful in the diagnosis and treatment of renal and metabolic disorders, including renal failure, gout, leukemia, psoriasis, starvation or other wasting conditions, and in patients receiving cytotoxic drugs. Using 1 ml of fasting blood serum, uric acid was measured using a standard spectrophotometry method. The reference range for adult men is 4.0-8.0 mg/dL, whereas for women, this range is cited as 2.5-7.0 mg/dL. (http://www.questdiagnostics.com/testcenter/TestDetail.action?ntc=905) Other reference ranges were also recently suggested and depend on the menopausal status of women. Those reference ranges are based on predictive value for gout outcomes among healthy individuals and do not necessarily predict other pathologies. Thus, based on recent research evidence, a "normal" SUA value is suggested to be <6.0 mg/dL for all healthy adult individuals.

Alkaline phosphatase

This liver enzyme was measured at Quest diagnostics using spectrophotometry. URL: https://testdirectory.questdiagnostics.com/test/test-detail/234/alkaline-phosphatase?cc=MASTER

SUPPLEMENTAL METHODS 4. DESCRIPTION OF TIME-INTERVAL MIXED-EFFECTS LINEAR REGRESSION MODELS, HANDLS 2004-2018

The main multiple mixed-effects regression models can be summarized as follows:

Multi-level models vs. **Composite models**

Eq.

1.1-1.4

$$Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \varepsilon_{ij}$$

$$\pi_{0i} = \gamma_{00} + \gamma_{0a} X_{aij} + \sum_{k=1}^{l} \gamma_{0k} Z_{ik} + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{1a} X_{aij} + \sum_{m=1}^{n} \gamma_{1m} Z_{im} + \zeta_{1i}$$

$$Y_{ij} = \gamma_{00} + \gamma_{0a} X_{aij} + \sum_{k=1}^{l} \gamma_{0k} Z_{ik} + \gamma_{10} Time_{ij} + \gamma_{1a} X_{aij} Time_{ij} + \sum_{m=1}^{n} \gamma_{1m} Z_{im} Time_{ij} + (\zeta_{0i} + \zeta_{1i} Time_{ij} + \varepsilon_{ij})$$

Where Y_{ij} is the outcome (Each biochemical or hematological marker measured at visits 1, 2 and 3) for each individual "i" and visit "j"; π_{0i} is the level-1 intercept for individual i; π_{1i} is the level-1 slope for individual i; γ_{00} is the level-2 intercept of the random intercept π_{0i} ; γ_{10} is the level-2 intercept of the slope π_{1i} ; Z_{ik} is a vector of fixed covariates for each individual *i* that are used to predict level-1 intercepts and slopes and included baseline age (Age_{base}) among other covariates. $X_{ija,}$ represents the main predictor variable (Serum Hcy at visit 1); ζ_{0i} and ζ_{1i} are level-2 disturbances; ε_{ij} is the within-person level-1 disturbance. Of primary interest are the main effects of each exposure X_a (γ_{0a}) and their interaction with *TIME* (γ_{1a}), as described in a previous methodolgical paper.[28]

Supplementary Table 1. Predictive models of serum homocysteine (Log_e transformed): CV, minimum BIC and adaptive linear LASSO for NHANES III, phase 2: training half-sample results and consistency with testing half-sample, 30-65 y.

	cvLASSO	Min BIC LASSO	Adaptive LASSO	Between sample R^2: cvLASSO	Between sample R^2: Min BIC	Between sample R^2: Adaptive LASSO
λ						
Training Sample, N=2004	0.0067622	0.047706	0.4496467	0.47	0.51	0.47
Testing Sample, N=2004	**--**	**--**	**--**	0.47	0.49	0.48
Non-zero parameter predictors (Log_e transformed, z-scored, SI units), training set (descending order of effect size)						
Serum folate	**x**	**x**	**X[a]**			
Serum creatinine	**x**	**x**	**X[a]**			
Age, y	**x**	**x**	**X[a]**			
Serum vitamin B-12	**x**	**x**	**X[a]**			
Aspartate aminotransferase	**x**	**x**	**x**			
Alanine aminotransferase	**x**		**X[a]**			
Serum uric acid	**x**	**x**	**X[a]**			
Mean cell hemoglobin	**x**		**X[a]**			

Serum albumin	**x**	**x**	**X[a]**
Census region: South vs. NorthEast	**x**	**x**	**X[a]**
Serum total calcium	**x**	**x**	**X[a]**
Lead	**x**	**x**	**x**
Serum triglycerides	**x**		**x**
Platelet count	**x**		**X[a]**
Gamma glutamyl transferase	**x**		**x**
Serum iron	**x**		**x**
Serum sodium	**x**		**X[a]**
Race/ethnicity: Other vs. NH white	**x**	**x**	**X[a]**
RBC folate	**x**	**x**	**x**
Serum vitamin A	**x**		**x**
Serum glucose	**x**	**x**	**X[a]**
Serum potassium	**x**		**x**
Serum total bilirubin	**x**		**x**
Serum vitamin C	**x**		**x**
White blood cell count	**x**		**X[a]**
Serum total calcium	**x**	**x**	**x**
Serum phosphorus	**x**		**X[a]**
Urinary albumin	**x**		**X[a]**

Serum selenium	**x**		**x**
Red cell distribution width	**x**		**x**
Serum normalized calcium	**x**	**x**	**x**
Below vs. Above poverty	**x**	**x**	**X[a]**
Serum alkaline phosphatase	**x**	**x**	**X[a]**
Serum thyroxine	**x**		**x**
Serum lactate dehydrogenase	**x**		**X[a]**
Serum sum retinyl esters	**x**	**x**	**X[a]**
Serum cotinine, ng/mL	**x**	**x**	**X[a]**
Serum bicarbonate	**x**		**X[a]**
Urinary iodine	**x**		**x**
Urinary creatinine	**x**		**x**
Rural vs. Urban	**x**	**x**	**X[a]**
Sex	**x**	**x**	**X[a]**
Serum vitamin D	**x**		**x**
Glycated hemoglobin, %	**x**		
Serum cholesterol	**x**		
Serum hepatitis C antibody	**x**		
Mean cell volume	**x**	**x**	
West vs. NorthEast	**x**	**x**	**X[a]**
Serum globulin	**x**		

Serum thyroid stim hormone	x		
Serum lutein/zeaxanthin	x		
Mexican-American vs. NH whites	x	x	X[a]
Serum HDL cholesterol	x		[a]
Serum chloride	x		[a]
Serum rubella antibody	x		
Serum anti-thyroglobulin antibody	x		
Serum hepatitis A antibody	x		
Platelet distribution width	x		
Vitamin A			[a]
Serum antimicrosomal antibody			[a]
RBC count			[a]
Intercept	x	x	x

Abbreviations: BIC=Bayesian Information Criterion; cv=cross-validation; Hcy=Homocysteine; LASSO= least absolute shrinkage and selection operator ; NH=Non-Hispanic; NHANES=National Health and Nutrition Examination Surveys.

[a] Non-zero terms identified in the same half sample using adaptive LASSO logit for binary Hcy outcome.

Supplementary Table 2. Reduced selected LASSO model for biochemical and hematological predictors (Log_e transformed and z-scored) of Hcy (Log_e transformed and z-scored) and elevated Hcy (equivalent to Hcy>14 *μmol/L*), adjusting for socio-demographic factors: NHANES III, phase 2; ages 30-65y [1]

	Homocysteine, Log_e transformed, z-score			**Elevated homocysteine: >14 μmol/L vs. ≤14 μmol/L**		
	β	(SE)	P-value	OR	95% CI	P-value
Reduced model	(N=3,784)			(N=3,794)		
Serum folate	**-0.36**	**(0.04)**	**<0.001**	**0.29**	**(0.18,0.45)**	**<0.001**
Serum creatinine	**+0.26**	**(0.03)**	**<0.001**	**1.85**	**(1.30,2.65)**	**0.002**
Age	**+0.19**	**(0.03)**	**<0.001**	**1.26**	**(1.02,1.57)**	**0.037**
Serum vitamin B-12	**-0.23**	**(0.06)**	**0.001**	**0.50**	**(0.31,0.83)**	**0.009**
Aspartate aminotransferase	**+0.11**	**(0.02)**	**<0.001**	**1.80**	**(1.25;2.59)**	**0.003**
Alanine aminotransferase	**-0.16**	**(0.02)**	**<0.001**	**0.39**	**(0.27,0.57)**	**<0.001**
Serum uric acid	**+0.11**	**(0.02)**	**<0.001**	**1.48**	**(1.10,2.00)**	**0.013**
Mean cell hemoglobin	**+0.11**	**(0.03)**	**0.004**	**1.69**	**(1.22,2.34)**	**0.003**
Serum albumin	**+0.17**	**(0.02)**	**<0.001**	**1.77**	**(1.33,2.36)**	**<0.001**
Total calcium	**+0.06**	**(0.02)**	**0.006**	**…** [2]		
Platelet count	**-0.04**	**(0.02)**	**0.096**	**…** [2]		
Gamma glutamyl transferase	**+0.04**	**(0.02)**	**0.081**	**…** [2]		
Serum glucose	**-0.05**	**(0.02)**	**0.009**	**…** [2]		
Serum vitamin C	**+0.05**	**(0.02)**	**0.058**	**…** [2]		
Red cell distribution width	**+0.11**	**(0.02)**	**<0.001**	**1.32**	**(1.09,1.61)**	**0.008**
Alkaline phosphatase	**+0.08**	**(0.03)**	**0.013**	**2.29**	**(1.52,3.44)**	**<0.001**
Serum retinyl esters	**-0.05**	**(0.02)**	**0.054**	**0.79**	**(0.65,0.96)**	**0.009**
Serum cotinine	**+0.07**	**(0.02)**	**0.012**	**1.27**	**(1.07,1.50)**	**<0.001**

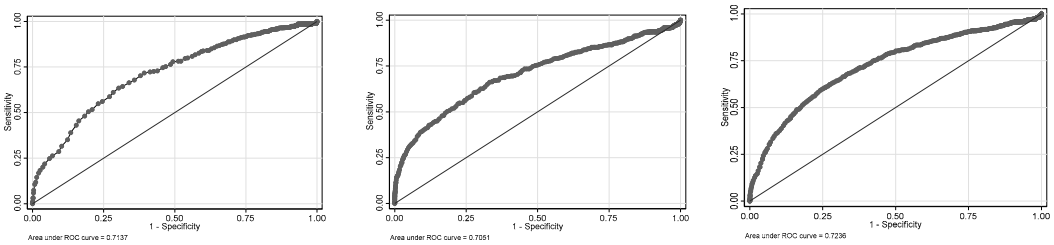
Abbreviations: Hcy=Homocysteine; LASSO= least absolute shrinkage and selection operator ; NH=Non-Hispanic; NHANES=National Health and Nutrition Examination Surveys; RBC=Red blood cells.

[1] Reduced models were determined using backward elimination starting from the initially selected LASSO model, applied to both linear and logistic regression models. At each step of the backward elimination, only parameters with $p<0.10$ were retained. The final step include parameters with $p<0.10$. [2] Predictor eliminated as compared to the linear regression reduced model. See units in Table 1.

Supplementary Figure 2. ROC curves of selected correlates of elevated homocysteine (>14 *μmol/L*) and cut-point determination: NHANES III, phase 2; NHANES 1999-2006 and pooled NHANES

(A) Lower Serum Folate

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

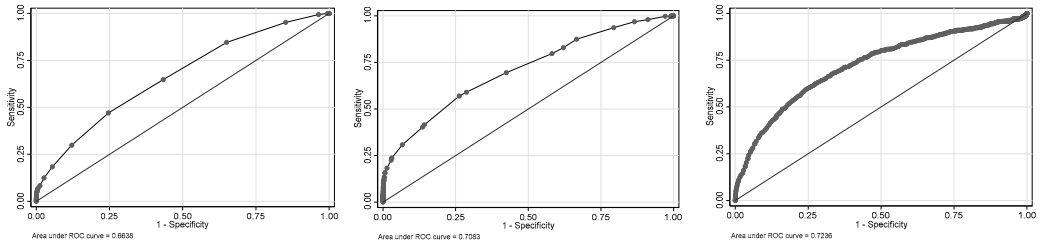


[1] Optimal cut-point at -2.20; [2] Optimal cut-point at -3.07; [3] Optimal cut-point at -2.83.

Note: Serum folate is in SI units (*nmol/L*), was Log_e transformed and multiplied by -1. Thus a value ≥-2.83 on this scale, corresponds to a Log_e transformed value ≤+2.83 and an untransformed value ≤16.93 *nmol/L.*

(B) Higher Serum Creatinine

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

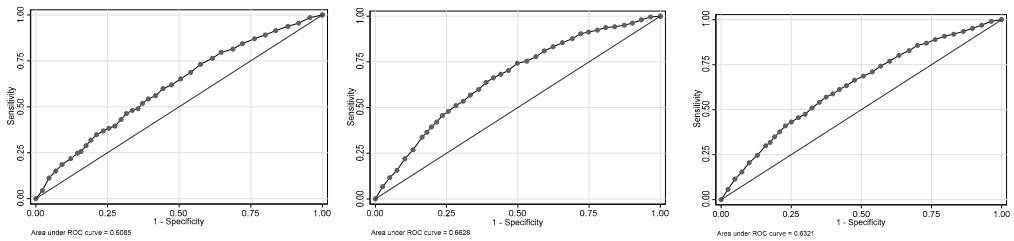


[1] Optimal cut-point at +4.576; [2] Optimal cut-point at +4.376; [3] Optimal cut-point at +4.481.

Note: Serum creatinine is Log_e transformed. Thus, a value ≥4.481 corresponds to an untransformed value ≥88.3. The same applies to all other biomarkers, except for age which not transformed. See units in Table 1.

(C) Older Age

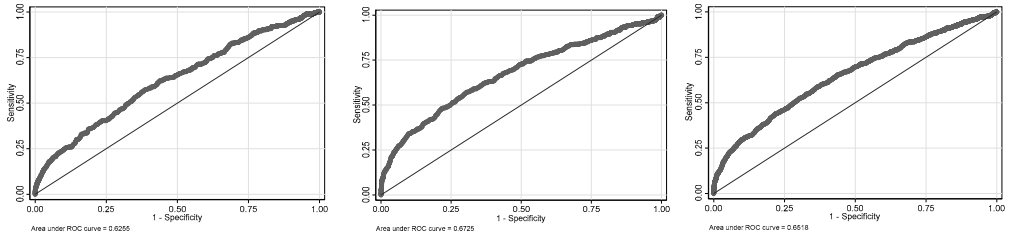
NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**



[1] Optimal cut-point at 46 years; [2] Optimal cut-point at 50; [3] Optimal cut-point at 49.

(D)Lower Serum B-12

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

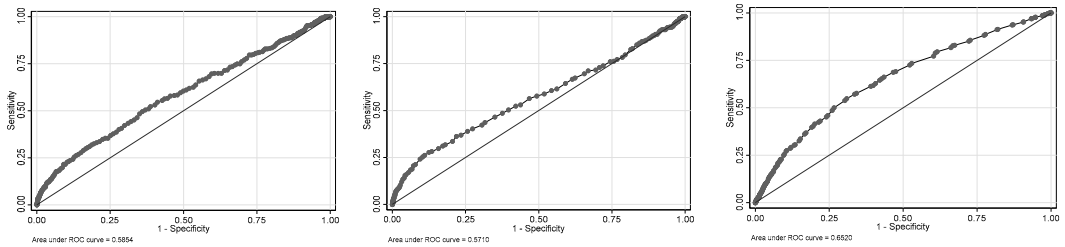


[1] Optimal cut-point at -5.76; [2] Optimal cut-point at -5.72; [3] Optimal cut-point at -5.74.

Note: Serum vitamin B-12 is in SI units (*pmol/L*), was Log_e transformed and multiplied by -1. Thus a value ≥-5.74 on this scale, corresponds to a Log_e transformed value ≤+5.74 and an untransformed value ≤311 *pmol/L*.

(E) Higher MCH

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

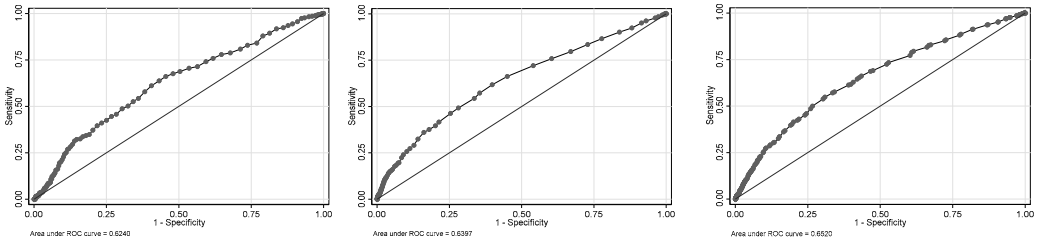


[1] Optimal cut-point at +3.409; [2] Optimal cut-point at +3.430; [3] Optimal cut-point at +3.422.

Note: MCH is in SI units (pg) and was Log_e transformed.

(F) Higher RDW

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

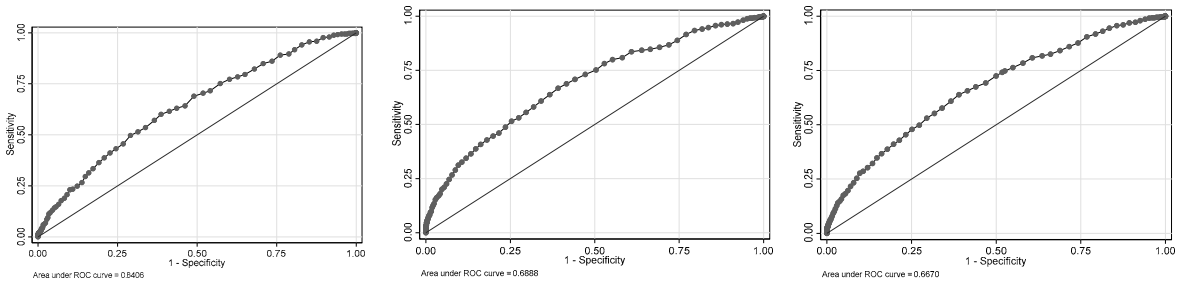


[1] Optimal cut-point at +2.576; [2] Optimal cut-point at +2.541; [3] Optimal cut-point at +2.553.

Note: RDW is in SI units (%) and was Log_e transformed.

(G) Higher Serum Uric acid

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

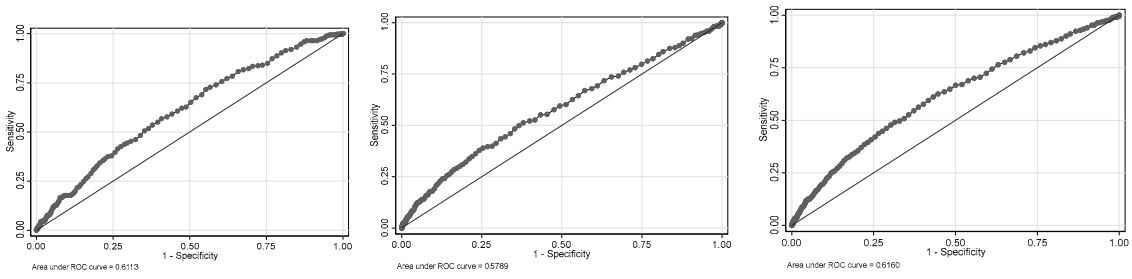


[1] Optimal cut-point at 5.808; [2] Optimal cut-point at 5.826; [3] Optimal cut-point at 5.826.

Note: Serum uric acid is in SI units ($\mu mol/L$) and was Log_e transformed.

(H) Higher Alkaline phosphatase

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**

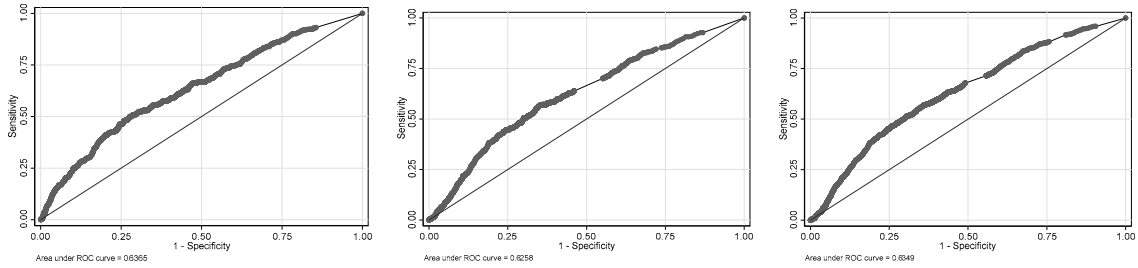


[1] Optimal cut-point at 4.477; [2] Optimal cut-point at 4.290; [3] Optimal cut-point at 4.356.

Note: Alkaline phosphatase is in SI units (*U/L*) and was Log_e transformed.

(I) Higher serum Cotinine

NHANES III, phase 2[1] **NHANES 1999-2006[2]** **Pooled NHANES[3]**



[1] Optimal cut-point at -0.349; [2] Optimal cut-point at -0.778; [3] Optimal cut-point at -0.579.

Note: Serum cotinine is in SI units *(ng/mL)* and was Log_e transformed.

Supplementary references

1. Jacques, P.F.; Rosenberg, I.H.; Rogers, G.; Selhub, J.; Bowman, B.A.; Gunter, E.W.; Wright, J.D.; Johnson, C.L. Serum total homocysteine concentrations in adolescent and adult Americans: results from the third National Health and Nutrition Examination Survey. *Am J Clin Nutr* **1999**, *69*, 482-489, doi:10.1093/ajcn/69.3.482.
2. Araki, A.; Sako, Y. Determination of free and total homocysteine in human plasma by high-performance liquid chromatography with fluorescence detection. *J Chromatogr* **1987**, *422*, 43-52, doi:10.1016/0378-4347(87)80438-3.
3. (DHHS)., U.S.D.o.H.a.H.S. National Center for Health Statistics. NHANES III reference manuals and reports. Hyattsville, MD: Centers for Disease Control and Prevention, 1996. . **1996**.
4. Gunter, E.W., Lewis, B. G., Koncikowski, S. M., . *Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994*, http://www.cdc.gov/nchs/data/nhanes/nhanes3/cdrom/nchs/manuals/labman.pdf; US DHHS, Public Health Service, Centers for Disease Control and Prevention, National Center for Environmental Health and National Center for Health Statistics: Atlanta, GA and Hyattsville, MD, 2010.
5. Addison, G.M.; Beamish, M.R.; Hales, C.N.; Hodgkins, M.; Jacobs, A.; Llewellin, P. An immunoradiometric assay for ferritin in the serum of normal subjects and patients with iron deficiency and iron overload. *Journal of clinical pathology* **1972**, *25*, 326-329.

6. Laboratories, B.-R. *Instruction Manual, Bio-Rad Quantaphase Folate Radioassay Kit.* ; Hercules, CA, 1987.
7. 7+7, N. *Electrolyte Analyzer instruction manual.*; Waltham (MA), 1986.
8. Lewis, S.A.; Hardison, N.W.; Veillon, C. Comparison of isotope dilution mass spectrometry and graphite furnace atomic absorption spectrometry with Zeeman background correction for determination of plasma selenium. *Analytical chemistry* **1986**, *58*, 1272-1273, doi:10.1021/ac00297a070.
9. Paschal D. C., K.M.M. Automated direct determination of selenium in serum by electrothermal atomic absorption spectroscopy. *At Spectrosc* **1986**, *7*, 75-78.
10. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Series 1: programs and collection procedures. *Vital and health statistics. Ser. 1, Programs and collection procedures* **1994**, 1-407.
11. World Health Organization. Diabetes Mellitus: Report of a WHO study group. In *WHO Technical Report Series* Geneva, Switzerland, 1995; p 727.
12. Bull, B.S.; Rittenbach, J.D. A proposed reference haematocrit derived from multiple MCHC determinations via haemoglobin measurements. *Clinical and laboratory haematology* **1990**, *12 Suppl 1*, 43-53.
13. National Committee for Clinical Laboratory Standards, N.d.H.-W., PA: NCCLS. ; 1993. *Procedure for determining packed cell volume by the microhematocrit method -- second edition: approved standard.*; NCCLS: Wayne, PA, 1993.
14. Benotti, J.; Benotti, N. Protein-Bound Iodine, Total Iodine, and Butanol-Extractable Iodine by Partial Automation. *Clinical chemistry* **1963**, *12*, 408-416.
15. Benotti, J.; Benotti, N.; Pino, S.; Gardyna, H. Determination of total iodine in urine, stool, diets, and tissue. *Clinical chemistry* **1965**, *11*, 932-936.
16. Chavers, B.M.; Simonson, J.; Michael, A.F. A solid phase fluorescent immunoassay for the measurement of human urinary albumin. *Kidney international* **1984**, *25*, 576-578.
17. Pruszkowska, E.; Carnrick, G.R.; Slavin, W. Direct determination of cadmium in urine with use of a stabilized temperature platform furnace and Zeeman background correction. *Clinical chemistry* **1983**, *29*, 477-480.
18. Center for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey. Availabe online: http://www.cdc.gov/nchs/nhanes.htm. (accessed on October 25).
19. Shipchandler, M.T.; Moore, E.G. Rapid, fully automated measurement of plasma homocyst(e)ine with the Abbott IMx analyzer. *Clinical chemistry* **1995**, *41*, 991-994.
20. Boushey, C.J.; Beresford, S.A.; Omenn, G.S.; Motulsky, A.G. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. *Jama* **1995**, *274*, 1049-1057, doi:10.1001/jama.1995.03530130055028.
21. Ueland, P.M.; Refsum, H.; Stabler, S.P.; Malinow, M.R.; Andersson, A.; Allen, R.H. Total homocysteine in plasma or serum: methods and clinical applications. *Clinical chemistry* **1993**, *39*, 1764-1779.
22. (CDC), C.f.D.C.a.P. Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Surveys (NHANES 2005-06): Description of Laboratory Methodology: Vitamin B-12. Availabe online:

http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/B12_D.htm#Description_of_Laboratory_Methodology (accessed on February 15).
23. (CDC), C.f.D.C.a.P. National Health and Nutrition Examination Surveys (NHANES 2005-06): Description of Laboratory Methodology: Folate. Availabe online: ttp://www.cdc.gov/nchs/nhanes/nhanes2005-2006/FOLATE_D.htm#Description_of_Laboratory_Methodology (accessed on February 15).
24. Xia, Y.; McGuffey, J.E.; Bhattacharyya, S.; Sellergren, B.; Yilmaz, E.; Wang, L.; Bernert, J.T. Analysis of the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol in urine by extraction on a molecularly imprinted polymer column and liquid chromatography/atmospheric pressure ionization tandem mass spectrometry. *Analytical chemistry* **2005**, *77*, 7639-7645, doi:10.1021/ac058027u.
25. Diagnostics, Q. Vitamin B-12 (cobalamin) and folate panel. Availabe online: https://testdirectory.questdiagnostics.com/test/test-detail/7065/vitamin-b12-cobalamin-and-folate-panel-serum?cc=MASTER (accessed on October 21st).
26. Diagnostics, Q. Hemogram. Availabe online: https://www.questdiagnostics.com/testcenter/BUOrderInfo.action?tc=7008&labCode=DAL (accessed on May 13sth).
27. techs, O.l.c.e.f.c.l.a.m. Red Blood Cell Distribution Width (RDW): Definition and Calculation. **2019**.
28. Blackwell, E.; de Leon, C.F.; Miller, G.E. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. *Psychosom Med* **2006**, *68*, 870-878, doi:01.psy.0000239144.91689.ca [pii]

10.1097/01.psy.0000239144.91689.ca.