



Observational Study

Vitamin d deficiency and metabolic syndrome: The joint effect on cardiovascular and all-cause mortality in the United States adults

Longjian Liu, Saishi Cui, Stella L Volpe, Nathalie S May, Deeptha Sukumar, Rose Ann DiMaria-Ghalili, Howard J Eisen

Specialty type: Nutrition and dietetics

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Peng D, China; Wolde MA, Ethiopia

Received: August 30, 2021

Peer-review started: August 30, 2021

First decision: April 7, 2022

Revised: April 25, 2022

Accepted: June 17, 2022

Article in press: June 17, 2022

Published online: July 26, 2022



Longjian Liu, Saishi Cui, Department of Epidemiology and Biostatistics, Dornsife School of Public Health, Drexel University, Philadelphia, PA 19104, United States

Stella L Volpe, Department of Human Nutrition, Foods, and Exercise, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, United States

Nathalie S May, Department of Medicine, College of Medicine, Drexel University, Philadelphia, PA 19102, United States

Deeptha Sukumar, Department of Nutrition Sciences, College of Nursing and Health Professions, Drexel University, Philadelphia, PA 19102, United States

Rose Ann DiMaria-Ghalili, Department of Graduate Nursing, College of Nursing and Health Professions, Drexel University, Philadelphia, PA 19102, United States

Howard J Eisen, Division of Cardiology, Heart and Vascular Institute, Pennsylvania State University, Hershey, PA 17033, United States

Corresponding author: Longjian Liu, MD, MSc, PhD, Doctor, Department of Epidemiology and Biostatistics, Dornsife School of Public Health, Drexel University, 3215 Market Street, Philadelphia, PA 19104, United States. ll85@drexel.edu

Abstract

BACKGROUND

The long-term impact of vitamin D deficiency and metabolic syndrome (MetS) on cardiovascular disease (CVD) and all-cause mortality are still a matter of debate.

AIM

To test the hypotheses that lower serum 25 hydroxyvitamin D [25(OH)D] concentrations (a marker of vitamin D level) and MetS have a long-term impact on the risk of CVD and all-cause mortality, and individuals with vitamin D deficiency can be identified by multiple factors.

METHODS

A sample of 9094 adults, 20 to 90 years of age, who participated in the Third National Health and Nutrition Examination Survey (NHANES III, 1988 to 1994) were followed through December 2015 was analyzed. The associations of serum

25(OH)D concentrations and MetS with CVD and all-cause mortality were analyzed longitudinally using Cox regression models. Classification and regression tree (CART) for machine learning was applied to classify individuals with vitamin D deficiency.

RESULTS

Of 9094 participants, 30% had serum 25(OH)D concentrations < 20 ng/mL (defined as vitamin D deficiency), 39% had serum 25(OH)D concentrations between 20 to 29 ng/mL (insufficiency), and 31% had serum 25(OH)D concentrations ≥30 ng/mL (sufficiency). Prevalence of MetS was 28.4%. During a mean of 18 years follow-up, vitamin D deficiency and MetS were significantly associated with increased risk of CVD and all-cause mortality. Subjects with both vitamin D deficiency and MetS had the highest risk of CVD mortality (HR = 1.77, 95%CI: 1.22-2.58) and all-cause mortality (HR = 1.62, 95%CI: 1.26-2.09), followed by those with both vitamin D insufficiency and MetS for CVD mortality (HR = 1.59, 95%CI: 1.12-2.24), and all-cause mortality (HR = 1.41, 95%CI: 1.08-1.85). Meanwhile, vitamin D sufficiency significantly decreased the risk of CVD and all-cause mortality for those who even had MetS. Among the total study sample, CART analysis suggests that being non-Hispanic Black, having lower serum folate level, and being female were the first three predictors for those with serum 25(OH)D deficiency.

CONCLUSION

Vitamin D deficiency and MetS were significantly associated with increased risk of CVD and all-cause mortality. There was a significant joint effect of vitamin D deficiency and MetS on the risk of mortality. Findings of the CART analysis may be useful to identify individuals positioned to benefit from interventions to reduce the risk of CVD and all-cause mortality.

Key Words: Joint effect; Serum 25 hydroxyvitamin D concentration; Metabolic syndrome; Cardiovascular and all-cause mortality; Cox model and machine learning

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: To investigate the long-term effect of vitamin D deficiency and metabolic syndrome on the risk of cardiovascular disease and all-cause mortality using a nationally representative sample. Standard measurements of the study exposures, co-variables and outcomes are processed. Multivariate Cox's proportional hazards regression analysis was used to prospectively test the associations between the exposures and outcomes. Classification and regression tree for machine learning was applied to classify subjects with higher risk of lower serum vitamin D concentrations.

Citation: Liu L, Cui S, Volpe SL, May NS, Sukumar D, DiMaria-Ghalili RA, Eisen HJ. Vitamin d deficiency and metabolic syndrome: The joint effect on cardiovascular and all-cause mortality in the United States adults. *World J Cardiol* 2022; 14(7): 411-426

URL: <https://www.wjgnet.com/1949-8462/full/v14/i7/411.htm>

DOI: <https://dx.doi.org/10.4330/wjc.v14.i7.411>

INTRODUCTION

The classical function of vitamin D is to increase the intestinal absorption of calcium for proper bone mineralization. Vitamin D can be obtained by the diet or synthesized on the skin with exposure to ultraviolet-B from the sun. The best method to determine vitamin D status is through measurement of serum 25 hydroxyvitamin D [25(OH)D] concentrations[1,2]. Recent evidence has demonstrated that individuals with vitamin D deficiency are more likely to have cardiovascular disease (CVD), and are associated with CVD risk factors, including metabolic syndrome (MetS)[3-7]. However, most studies examined the cross-sectional association of serum 25(OH)D concentrations and MetS with the prevalence of CVD. Few studies, with a longitudinal study design, estimated the causal association between vitamin D deficiency and the risk of CVD. Furthermore, the findings of the previous studies are inconsistent. For example, the Women's Health Initiative calcium-vitamin D trial, a randomized clinical trial, observed a weak association between vitamin D supplementation and CVD risk reduction among postmenopausal women[8,9]. Meanwhile, the characteristics of the population with vitamin D deficiency varied by different studies[2]. In our early report in 2012, we tested the medium-term effect of serum vitamin D deficiency and risk of heart failure and premature mortality[4]. In the present study, we aimed to examine the long-term effect of vitamin D and MetS on the risk of CVD and all-cause

mortality using data from the third National Health and Nutrition Examination Survey (NHANES III). The NHANES III started in 1988 to 1994 and is an ongoing follow-up study for the participants' vital status through the national Linked Mortality Files (LMFs). The NHANES III provides us a unique opportunity to prospectively examine the association between baseline serum 25(OH)D concentrations and risk of disease-specific and all-cause mortality among community-dwelling residents in the United States[10,11]. In this study, we aimed to test two hypotheses: (1) Serum 25(OH)D deficiency and MetS are significantly associated with risk of CVD and all-cause mortality; and (2) serum 25(OH)D deficiency varies by age, sex, race/ethnicity, poverty and key measures of serum biomarkers.

MATERIALS AND METHODS

The NHANES III is a nationwide survey conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) in the United States[10,12]. The baseline survey was conducted from October 1988 through October 1994 in two phases (phase 1: 1988-1991, and phase 2: 1991-1994) gathering information representing the health and nutritional status of the noninstitutionalized civilian United States residents ages 2 mo and older. Participants were recruited based on a nationwide probability sample across four regions of the USA (Northeast, Midwest, South, and West). The study consisted of (1) household interviews on sociodemographic factors, health behaviors, and history of medical conditions through standard survey instruments; (2) physical examinations in mobile examination centers (MEC); and (3) laboratory measures from blood and urine samples. Blood samples were obtained at the MEC for measurements of serum and urinary biomarkers. To minimize hemolysis of blood samples in the measures of serum lipid profiles and glucose concentrations, participants were asked to fast for 12 hours before a morning examination, or 6 hours before an afternoon or evening examination. All phlebotomists were certified and trained in standardized laboratory procedures. Detailed information on the NHANES III has been provided elsewhere[10,12].

Follow-up data for participants' survival status or disease-specific and all-cause mortality were extracted from the NHANES III Linked Mortality File (LMF). The LMF was linked by the NCHS working with the State's Offices of Vital Statistics to link NHANES III data with death certificate records using the National Death Index (NDI)[13]. The NDI is a central computerized index of death record information on files. This standard and technique-supported linkage process provide an opportunity to conduct longitudinal analyses between measures at baseline and outcomes through the follow-up period. In the present study, we used data from the recently available follow-up of the LMF for the NHANES III from baseline 1988-1994 through December 31, 2015. The NHANES III survey was approved by the CDC NCHS institutional review board. Data obtained from the NCHS for the present study were de-identified for participants' private information security. The study, using the deidentified NHANES III and its LMF to examine risk factors for mortality has been approved by the Drexel University Office of Institutional Review Board (#2105008546). Of 18799 participants, 23 without follow-up and 2869 without serum 25(OH)D measures were excluded. Of the rest 15907, we further excluded 6251 who had fasting blood sample < 8 hours (to assess triglycerides and glucose concentrations), and 562 who had no measurements of five MetS components. That, the final study sample size was 9094 for those who had serum 25(OH)D concentrations measured, had valid follow-up records for their vital status, and had full measurements of MetS components [waist circumference, systolic blood pressure (SBP), diastolic blood pressure (DBP), and high-density lipoprotein cholesterol (HDL-C), triglyceride (TG), and fasting glucose concentrations]. Figure 1A shows the sample size flowchart.

Exposures

Measurements of serum 25(OH)D concentrations were performed as part of the nutritional biomarkers of the NHANES III (1988 to 1994)[10]. Serum 25(OH)D concentrations were measured at the National Center for Environmental Health, CDC, Atlanta GA, using the DiaSorin RIA kit (Stillwater MN). In the study, we analyzed serum 25(OH)D concentrations as a categorical variable in the Cox regression models to examine hazard ratios of serum 25(OH)D concentrations for risk of mortality. We applied the commonly used cut-off points to categorize serum 25(OH)D levels into three groups: (1) Vitamin D deficiency was defined as serum 25(OH)D concentrations < 20 ng/mL (< 50 nmol/L); (2) Vitamin D insufficiency was defined as serum 25(OH)D concentration 20 to 29 ng/mL (50 to 74 nmol/L); and (3) Vitamin D sufficiency was defined as serum 25(OH)D concentrations ≥ 30 ng/mL (≥ 75 nmol/L)[1,4,14,15]

On MetS, five components are included. These are waist circumference (WC), which was measured using a standard flexible and tension-regulated tape measure. Systolic and diastolic blood pressure (SBP and DBP, respectively), were measured with an automated blood pressure monitor after five minutes of rest, with the last two of three readings, averaged and recorded. Blood samples collected at least 8 hours fasting were used to assess TG, and fasting glucose concentrations. To define MetS, we applied the cut-off criteria modified from the American Heart Association (AHA), the American Diabetes Association (ADA), and the Adult Treatment Panel III (ATP III)[16]. Individuals with MetS were defined by the presence of 3 of five MetS components: (1) Large WC (a marker of abdominal adiposity): WC > 102 cm

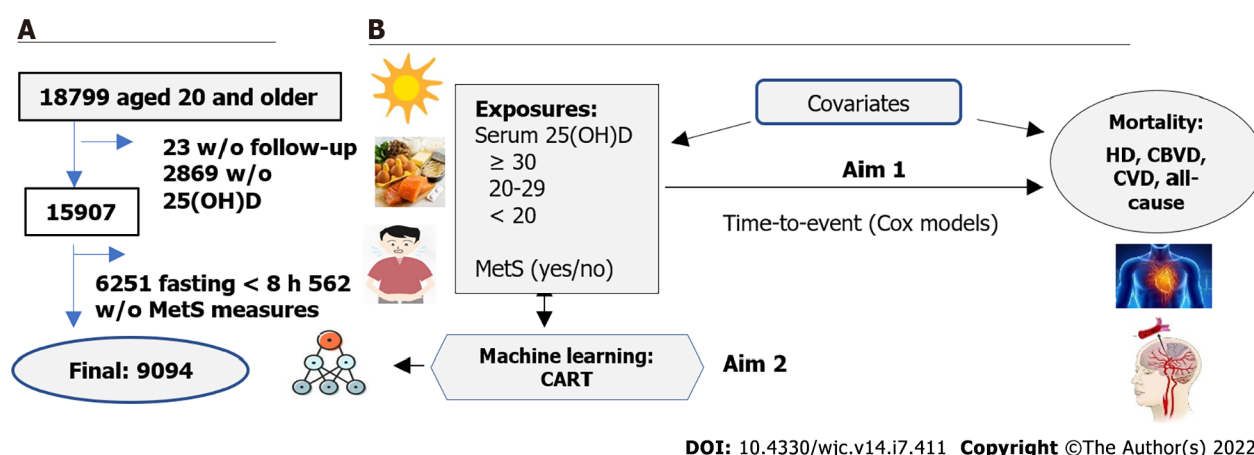


Figure 1 Participant flowchart (A) and the analysis conceptual model (B) in the study. Serum 25(OH)D: Serum 25 hydroxyvitamin D (a marker of vitamin D level in blood); MetS: Metabolic syndrome; CART: Classification and regression tree; HD: Heart disease; CBVD: Cerebrovascular disease; CVD: Cardiovascular disease.

in males and > 88 cm in females; (2) Elevated BP: SBP \geq 130 or DBP \geq 85 mmHg or under antihypertensive medication use; (3) Elevated TG: TG \geq 150 mg/dL; (4) Elevated glucose: fasting glucose \geq 100 mg/dL or use of glucose-lowering medications; and (5) Low HDL-C: HDL-C < 40 mg/dL in males and HDL-C < 50 mg/dL in females.

Outcomes

The study outcomes included mortality from (1) Heart disease (HD) (*International Classification of Disease, Tenth Revision, ICD-10* codes: I00-I09, I11, I13, I20-I51); (2) Cerebrovascular diseases (CBVD, ICD-10 codes: I60-I69); (3) CVD (included either HD and CBVD); and (4) All-cause mortality. Follow-up time (years) was calculated from the baseline interview to the end of follow-up (December 31, 2015), or until the date of death if a participant died before the end of follow-up, whichever occurred first.

Covariates

We included the following factors as covariates in the analysis. These variables are: age (years), sex (males, and females), race/ethnicity, educational attainments, smoking status, alcohol consumption, physical activity, and chronic conditions. Race/ethnicity are classified based on participants' self-reports using standard survey questionnaire. Three groups are recorded: non-Hispanic White, non-Hispanic Black (*i.e.*, African American), and the other groups (including Hispanic or Latino, Asian, native Hawaiian or other Pacific Islander). Educational attainments are classified into three groups (less than high school, high school, and colleges or higher levels), smoking status, alcohol consumption, physical activity, and chronic condition. Smoking was categorized as "never", "former", and "current smokers". Alcohol consumption was defined by a question, "Have you had at least 12 drinks in the last 12 mo?" (*i.e.*, at least once a month, "yes" or "no"). Physical activity was categorized as more activity, about the same, or less activity from a survey question: "How does the amount of activity that you reported for the past month compare with your physical activity for the past 12 mo?" Chronic conditions were self-reported diagnoses on questions: "Have you been told by a doctor that you have the following disease(s): hypertension, diabetes mellitus, coronary heart disease, heart failure or stroke?" ("yes" or "no").

Statistical analysis

We conducted a serial analysis to examine the two specific aims (Figure 1B). In the first set of analyses, we described the characteristics of the study participants across three groups of serum 25(OH)D concentrations (*i.e.*, serum 25(OH)D < 20, 20 to 29, and \geq 30 ng/mL, respectively). Differences in MetS rates, sex, race/ethnicity, education attainment, smoking status, alcohol consumption, and history of chronic conditions at baseline were tested using the Chi-squared test. Differences in continuous variables were tested using analysis of variance. In the second-set analyses, we estimated age-sex-adjusted mortality rates (per 10000 person-years) from HD, CBVD, CVD, and all-cause mortality by serum 25(OH)D concentrations and MetS using Poisson regression models in SAS GENMOD procedure. In the third-set analyses, we estimated the hazard ratio and its 95% confidence interval of serum 25(OH)D concentrations and MetS for HD, CBVD, CVD, and all-cause mortalities using time-to-event Cox proportional hazards regression (Cox) models. In the analyses, follow-up time (years) and mortality (yes or no) were the dependent variables, and serum 25(OH)D concentrations (reference to 25(OH)D \geq 30 ng/mL) and MetS (yes *vs* no) were the independent variables. To test Cox proportional hazard assumption, we applied plots of log(-log) survival curves and Schoenfeld residuals methods. To control potential

confounders, we performed two multivariate-adjusted models. Model 1 adjusted for age (years), sex (males and females), race/ethnicity [non-Hispanic White (NHW), non-Hispanic Black (NHB), and the others], education (< high school, high school, and > high school), and living in regions of the United States (Northeast, Midwest, South, and West). We included the adjustment for regions for control potential difference in the exposures to sunshine across the nation because sunshine is one of the possible resources for a human to get vitamin D. Model 2 further adjusted for covariates including smoking (never, former, and current smokers), alcohol consumption (yes or no), physical activity (more activity, about the same, or less activity) and baseline prevalence of CVD (yes or no) because these lifestyle and baseline health conditions were associated with the study exposures and outcomes. We tested the interaction effects of serum 25(OH)D and MetS on mortality risk as well. In the fourth-set analyses, because we did not observe a significant interaction effect of serum 25(OH)D and MetS on the risk of CVD and all-cause mortality, we further estimated their joint effect on the risk of HD, CBVD, CVD, and all-cause mortality.

In the fifth-set analyses, we applied classification and regression trees (CART), a predictive algorithm in machine learning techniques, to test the characteristics of participants with different serum 25(OH)D concentrations to use classification tree analysis to identify individuals who are at higher risk of vitamin D deficiency. CART analysis is a standard decision tree analysis in machine learning introduced by Breiman *et al*[17] and Bzdok *et al*[18]. CART analysis does not request data with a certain mathematical distribution (*i.e.*, distribution-free). Compared to Principal Component Analysis and Factor Analysis, CART analysis more easily handles noisy data, such as outliers from multiple variables, because CART's splitting algorithms can isolate outliers into a separate node[17]. In CART analysis, data are partitioned along the predictor axes into subsets with homogeneous values of the outcome variable. A decision tree is created, which can be used to make predictions from new observations[18]. The selection of important predictors can be assessed by mean decrease in Gini (MDG) index. MDG index is an average of a variable's total decrease in node impurity, weighted by the proportion of samples reaching that node in each decision tree in the random forest. A higher MDG index indicates higher variable importance. To valid the classification process, we used two sub-datasets: (1) 20% of the total participants were randomly selected serving as the validation dataset; and (2) The remaining 80% served as the training dataset to construct the predicting models. In the analysis, serum 25(OH)D concentrations in two and three categorical groups as the dependent variable and the other 46 variables as the independent variables (Supplementary Table 1). Serum 25(OH)D concentrations were categorized as: (1) Serum 25(OH)D ≥ 20 vs < 20 ng/mL; (2) Serum 25(OH)D ≥ 30 , vs 20 to 29, vs < 20 ng/mL; and (3) Serum 25(OH)D ≥ 30 vs < 30 ng/mL.

Statistical analyses were performed using SAS 9.14 (SAS Institute, Cary, North Carolina). SAS analyses for complex sample surveys were used by taking into consideration the NHANES III's complex, multistage, probability sampling design. CART analyses were performed using R programming[19], and "rpart" package[20]. A 2-sided *P* value < 0.05 was considered statistically significant in all data analyses.

Finally, we conducted a sensitivity analysis by the exclusion of those who died within the first year of follow-up to avoid a potential over-estimate of the association of serum 25(OH)D concentrations and MetS with mortality risk because patients with serious health conditions might die within the first year of the follow-up.

RESULTS

Participants' characteristics

Of 9094 participants, 30% had serum 25(OH)D concentrations less than 20 ng/mL (defined as vitamin D deficiency), 39% had serum 25(OH)D concentrations in the 20 to 29 ng/mL range (insufficiency), and 31% had serum 25(OH)D concentrations ≥ 30 ng/mL (sufficiency). The total prevalence of MetS was 28.4%. Table 1 shows that across the three groups of serum 25(OH)D concentrations, the highest prevalence of MetS (35.3%) was observed among participants with serum 25(OH)D < 20 ng/mL, followed by 29.2% and 20.7%, respectively, among those with serum 25(OH)D concentrations in the 20 to 29 ng/mL range and those with serum 25(OH)D concentrations ≥ 30 ng/mL (testing for rates differences, *P* < 0.001). The mean ages were 45.8, 44.7, and 40.3 years older among those with serum 25(OH)D concentrations of < 20 ng/mL, 20 to 29 ng/mL, and ≥ 30 ng/mL, respectively (*P* < 0.001).

Participants with the lowest serum 25(OH)D concentrations had the highest mean waist circumference (WC), body mass index (BMI), SBP and DBP, serum TG, and glucose concentrations, but slightly lower HDL-C concentrations compared to those with serum 25(OH)D ≥ 30 ng/mL. Among those with serum 25(OH)D concentrations < 20 ng/mL, 61.7% were females. The proportions of race/ethnicity, education attachments, and alcohol consumption by serum 25(OH)D concentrations were significantly different (*P* < 0.001). Meanwhile, participants with serum 25(OH)D concentration < 20 ng/mL had a significantly higher prevalence of hypertension (36.5%), diabetes mellitus (9.3%), and CVD (6.3%) compared to those with serum 25(OH)D 20-29 ng/mL and those with serum 25(OH)D ≥ 30 ng/mL, respectively (*P* < 0.01).

Table 1 Characteristics of participants at baseline

	Serum 25(OH)D concentrations									P value
	< 20 ng/mL			20 to 29 ng/mL			≥ 30 ng/mL			
	No.	mean, %	(SE)	No.	mean, %	(SE)	No.	mean, %	(SE)	
MetS, %										< 0.001
No	2640	64.7		2230	70.8		1276	79.3		
Yes	1394	35.3		1089	29.2		465	20.7		
Age, yr	4034	45.8	(0.6)	3319	44.7	(0.6)	1741	40.3	(0.6)	< 0.001
Waist circumference, cm	4034	94.2	(0.4)	3319	92.2	(0.3)	1741	88.3	(0.4)	< 0.001
Body mass index, kg/m ²	4033	27.8	(0.2)	3318	26.5	(0.1)	1739	25.0	(0.1)	< 0.001
Systolic BP, mmHg	4034	123.4	(0.6)	3319	122.0	(0.6)	1741	119.0	(0.6)	<0.001
Diastolic BP, mmHg	4034	74.7	(0.3)	3317	74.1	(0.3)	1741	72.9	(0.4)	< 0.001
Triglycerides, mg/dL	4034	138.9	(4.1)	3319	136.4	(2.6)	1741	126.5	(3.8)	< 0.001
HDL-C, mg/dL	4034	51.0	(0.5)	3319	49.3	(0.4)	1741	51.5	(0.6)	< 0.001
Glucose, mg/ dL	4034	100.7	(0.7)	3319	98.4	(0.5)	1741	94.7	(0.4)	< 0.001
Categorical variable, %										
Sex, male, %	1561	38.3		178	51.9		975	55.4		< 0.001
Race/ethnicity, %										< 0.001
NHW	894	55.1		1557	79.7		1213	92.0		
NHB	1788	25.8		575	6.0		139	1.9		
Others	1352	19.1		1187	14.3		389	6.1		
Region										< 0.001
Northeast	490	15.8		429	20.6		303	26.7		
Midwest	651	18.9		709	24.5		434	28.1		
South	1933	40.3		1299	33.2		685	29.1		
West	960	25.0		882	21.6		319	16.1		
Education level										0.001
< High school	930	13.1		818	11.8		355	8.2		
High school	1971	49.2		1532	45.8		788	46.1		
> High school	1113	37.7		949	42.4		592	45.6		
Smoking										0.07
Never	2562	62.1		2230	63.9		1130	62.5		
Former	193	5.5		191	6.0		105	4.0		
Current	1279	32.4		898	30.2		506	33.5		
Alcohol consumption, yes	1704	44.8		1592	55.9		910	63.7		< 0.001
Exercise, yes										0.13
More activity	559	14.1		475	16.3		287	18.4		
About the same	2433	58.3		1998	56.2		1005	52.7		
Less activity	1040	27.6		841	27.5		445	28.9		
Health conditions										
Hypertension, yes	1489	36.5		1129	29.6		565	25.5		< 0.001

Diabetes mellitus	465	9.3	312	6.7	105	3.3	< 0.001
CVD	278	6.3	252	5.3	124	3.4	0.001

Weighted mean and percentage are estimated because NHANES III applied a complex, multistage, probability sampling design. SE: Standard error; MetS: Metabolic syndrome; BP: Blood pressure; HDL-C: High-density lipoprotein cholesterol; NHW: Non-Hispanic White; NHB: Non-Hispanic Black; Others: Including all the other race/ethnicity groups; CVD: Cardiovascular disease included those with coronary heart disease, heart failure, or stroke.

Mortality rate by serum 25(OH)D concentrations and MetS status

Table 2 shows that during a mean of 18 years follow-up, participants with serum 25(OH)D concentrations < 20 ng/mL had significantly higher age-sex-adjusted mortality from HD (17.4 per 10000 person-years), CVD (21.3 per 10000 person-years), and all-cause mortality (103.9 per 10000 person-years) compared to those with serum 25(OH)D concentrations \geq 30 ng/mL. Participants with MetS had significantly higher age-sex-adjusted mortality from HD (16.7 per 10000 person-years), CVD (22.0 per 10000 person-years), and all-cause mortality (102.4 per 10000 person-years) compared to those without MetS.

Multivariate-adjusted hazard ratios of serum 25(OH)D and MetS for mortality

Table 3 shows that, after adjusting for age, sex, race/ethnicity, and region (Model 1), and additional adjusting for health behaviors and baseline CVD (Model 2), participants with serum 25(OH)D concentrations < 20 ng/mL had a significantly higher risk of HD, CVD, and all-cause mortality compared to those with serum 25(OH)D concentrations \geq 30 ng/mL. The corresponding hazard ratios [HRs and 95% confidence intervals (CI)] in Model 2 were 1.43 (1.07-1.91, $P = 0.017$) for HD, 1.36 (1.04-1.78, $P = 0.028$) for CVD, and 1.32 (1.10-1.58, $P = 0.004$) for all-cause mortality, respectively.

Participants with MetS had a significantly higher risk of CVD and all-cause mortality compared to those without MetS in Models 1 and 2. The corresponding HRs (95% CI) in Model 2 were 1.29 (1.05-1.59, $P = 0.017$) for CVD, and 1.20 (1.05-1.36, $P = 0.01$) for all-cause mortality.

No significant association of serum 25(OH)D concentrations and MetS with the risk of CVD mortality, and no significant interaction effects of serum 25(OH)D concentrations and MetS on the risk of HD, CVD, and all-cause mortality were observed ($P > 0.05$).

Joint effects of serum 25(OH)D concentrations and MetS on mortality risk

We further estimated the joint effect of those who had lower serum 25(OH)D concentrations and MetS on mortality risk. **Figure 2** depicts that individuals with serum 25(OH)D concentrations < 20 ng/mL and MetS had the highest risk of CVD (HR = 1.77, 95% CI: 1.22-2.58, $P < 0.01$, **Figure 2A**) and all-cause mortality (HR = 1.62, 95% CI: 1.26-2.09, $P < 0.01$, **Figure 2B**), followed by those with serum 25(OH)D concentrations 20 to 29 ng/mL and with MetS [1.59 (1.12-2.24, $P = 0.01$) for CVD, and 1.41 (1.08-1.85, $P = 0.01$) for all-cause mortality] as compared to those with serum 25(OH)D \geq 30 ng/mL and without MetS (the reference group). **Figure 2** also shows that individuals with MetS, but with serum 25(OH)D concentrations \geq 30 ng/mL, did not have a significantly higher risk of CVD (HR = 1.44, 95% CI: 0.96-2.22, $P = 0.008$) and all-cause mortality (HR = 1.30, 95% CI: 0.95-1.78, $P = 0.09$) compared to the reference group (shown as green bars in **Figure 2**), indicating a higher protective effect of serum 25(OH)D concentrations \geq 30 ng/mL on CVD and all-cause mortality compared to MetS.

Identifying the main predictors for those with lower serum 25(OH)D concentrations

We used CART to identify the top 10 important predictors from 46 variables (**Supplementary Table 1**) for participants with serum 25(OH)D at different concentrations and by sex. **Table 4** shows the overall results. **Figure 3** depicts an example of CART analysis for serum 25(OH)D concentrations \geq 20 ng/mL *vs* < 20 ng/mL among the total sample (S1 of **Table 4**). Racegp (race/ethnicity) is grouped as racegp = 1 for non-Hispanic White (NHW), 2 for non-Hispanic Black (NHB), and 3 for the other racial/ethnic groups. **Figure 3A** indicates that the classification tree starts from node 1 (also called the root) and ended in node 5 (*i.e.*, the “leaves”). Under each oval-shaped box, the values on the left-listed indicate the percentage of individuals with serum 25(OH)D concentrations \geq 20 ng/mL [such as under node 1, 55% of individuals had serum 25(OH)D concentrations \geq 20 ng/mL]. It should be noted that if the left-listed value is more than 0.5 (*i.e.*, > 50%), the oval-shaped box will be labeled as 25(OH)D \geq 20 ng/mL, otherwise labeled as 25(OH)D < 20 ng/mL (such as 5a and 5b of node 5, *i.e.*, 0.29 and 0.38 are less than 0.50). Under each box, the value on the right-listed indicates the percentage of the participants included in the node (such as 100% in node 1). **Figure 3A** depicts that among the randomly selected training sample, 55% of them had serum 25(OH)D concentrations \geq 20 ng/mL. In the analysis, race/ethnicity was first selected by the CART algorithm as a predictor. Of the sample, 28% (the value showing on the right-listed under box 5a of node 5) were NHB (*i.e.*, racegp = 2). Among NHB, 29% (on the left-listed value) had serum 25(OH)D concentrations \geq 20 ng/mL (note: the oval-shaped box is labeled as 25(OH)D < 20 ng/mL because 0.29 is less than 0.5). In other words, 71% of NHB had serum 25(OH)D concentrations < 20 ng/mL. Of the total training sample, 72% were NHW and other racial/ethnic groups (*i.e.*, racegp \neq 2), and 66% of them had

Table 2 Age-sex-adjusted mortality from heart disease, cerebrovascular disease, cardiovascular disease, and all-cause mortality

	Observed number		Weighted mortalityrate per 10000 PY		P value
	No. at risk	Case	Rate	(95%CI)	
Heart Disease					
25(OH)D, ng/mL					
≥ 30	1741	130	11.7	(8.2-16.6)	Ref.
20-29	3319	272	13.2	(10.3-16.9)	0.43
< 20	4034	292	17.4	(13.1-23.0)	0.012
CBVD					
25(OH)D, ng/mL					
≥ 30	1741	34	3.6	(2.0-6.6)	Ref.
20-29	3319	83	3.6	(2.5-5.1)	0.97
< 20	4034	88	4.5	(2.5-7.9)	0.53
CVD					
25(OH)D, ng/mL					
≥ 30	1741	164	15.3	(11.4-20.5)	Ref.
20-29	3319	355	16.6	(13.3-20.7)	0.56
< 20	4034	382	21.3	(16.6-27.3)	0.023
All-cause					
25(OH)D, ng/mL					
≥ 30	1741	545	75.9	(64.9-88.8)	Ref.
20-29	3319	1141	87.5	(79.4-96.3)	0.066
< 20	4034	1290	103.9	(93.2-115.8)	< 0.001
MetS status					
Heart Disease					
MetS, no	6146	327	12.5	(9.7-16.2)	Ref.
MetS, yes	2948	369	16.7	(12.9-21.7)	0.002
CBVD					
MetS, no	6146	97	3.2	(2.2-4.7)	Ref.
MetS, yes	2948	108	5.1	(2.9-8.9)	0.059
CVD					
MetS, no	6146	424	15.9	(13.0-19.5)	Ref.
MetS, yes	2948	477	22.0	(16.9-28.6)	< 0.001
All-cause					
MetS, no	6146	1490	82.5	(75.3-90.5)	Ref.
MetS, yes	2948	1486	102.4	(91.0-115.3)	< 0.001

PY: Person-years; CBVD: Cerebrovascular disease; CVD: Cardiovascular disease; 25(OH)D: Serum 25 hydroxyvitamin D concentration; MetS: Metabolic syndrome.

serum 25(OH)D concentrations ≥ 20 ng/mL (these values showing under the box of node 2). Of the 72% sample (*i.e.*, racegp $\neq 2$), 32% were the other race/ethnicity (*i.e.*, racegp = 3), and 54% of them had serum 25(OH)D concentrations ≥ 20 ng/mL. The rest 40% (racegp $\neq 2$ and racegp $\neq 3$) were NHW, and 75% of NHW had serum 25(OH)D concentrations ≥ 20 ng/mL (Box 5e of node 5). Thirty-two percent (32%) of the total sample were other racial/ethnic groups (racegp = 3, node 3), 16% of the total sample were other racial/ethnic females (racegp = 3 and sex = 2, node 4), and of them, 43% had serum 25(OH)D concen-

Table 3 Hazard ratios of serum 25 hydroxyvitamin D concentration and metabolic syndrome for heart disease, cerebrovascular disease, cardiovascular disease, and all-cause mortality

	Model 1			Model 2		
	HR	(95%CI)	P value	HR	(95%CI)	P value
Ref to 25(OH)D ≥ 30 ng/dL						
25(OH)D for HD						
20-29	1.13	(0.82-1.55)	0.45	1.13	(0.83-1.52)	0.43
< 20	1.54	(1.14-2.08)	0.005	1.43	(1.07-1.91)	0.017
25(OH)D for CBVD						
20-29	0.97	(0.61-1.53)	0.88	0.98	(0.62-1.56)	0.94
< 20	1.15	(0.58-2.26)	0.69	1.14	(0.58-2.25)	0.69
25(OH)D for CVD						
20-29	1.09	(0.82-1.46)	0.54	1.10	(0.82-1.46)	0.52
< 20	1.44	(1.09-1.91)	0.011	1.36	(1.04-1.78)	0.028
25(OH)D for all-cause mortality						
20-29	1.16	(0.99-1.35)	0.06	1.16	(0.99-1.36)	0.07
< 20	1.38	(1.16-1.64)	0.001	1.32	(1.10-1.58)	0.004
Ref to Non-MetS						
MetS for HD	1.31	(1.09-1.58)	0.005	1.22	(1.00-1.49)	0.053
MetS for CBVD	1.58	(0.94-2.66)	0.083	1.60	(0.96-2.65)	0.069
MetS for CVD	1.37	(1.12-1.66)	0.002	1.29	(1.05-1.59)	0.017
MetS for all-cause mortality	1.22	(1.08-1.37)	0.002	1.20	(1.05-1.36)	0.008
Interaction of						
25(OH)D \times MetS on						
HD	0.91	(0.72-1.17)	0.46	0.91	(0.72-1.15)	0.41
CBVD	0.80	(0.39-1.66)	0.54	0.79	(0.38-1.62)	0.50
CVD	0.89	(0.71-1.12)	0.31	0.88	(0.70-1.11)	0.28
All-cause	0.98	(0.84-1.14)	0.78	0.96	(0.82-1.13)	0.63

Model 1 adjusted for age, sex, race/ethnicity, education, and regions; Model 2 adjusted covariate in Model 1 plus smoking status, physical activity, alcohol consumption, and baseline CVD. HR: Hazard ratio; 25(OH)D: Serum 25 Hydroxyvitamin D concentration; HD: Heart disease; CBVD: Cerebrovascular disease; CVD: Cardiovascular disease; MetS: Metabolic syndrome.

trations ≥ 20 ng/mL (note: the oval-shaped box is labeled as 25(OH)D < 20 ng/mL because 0.43 is less than 0.5). The rest 16% of the total sample were other racial/ethnic males (racegp = 3 and sex = 1). Of the males, 65% had serum 25(OH)D concentrations ≥ 20 ng/mL (Box 5d of node 5). Among the females (node 4), the CART algorithm further identified serum folate concentrations (FOP, a marker of folate intake) as an important factor to continue the classification. Sixteen percent (16%) of the total sample were other racial/ethnic females, 14% of them had serum folate concentrations (FOP) < 8.9 ng/mL, and 38% of them had serum 25(OH)D concentrations ≥ 20 ng/mL. Overall, node 3 it suggests that sex and serum folate level play a role in the further classification among the other racial/ethnic groups. The classified rates for those with serum 25(OH)D concentrations ≥ 20 range from 62% [*i.e.*, $(1-0.38 = 0.62)$, Box 5b of node 5] to 75% (box 5e of node 5). **Figure 3B** depicts the top 10 important factors that predicted those with serum 25(OH)D concentrations ≥ 20 *vs* < 20 ng/mL. These ranks are assigned based on the values of their MDG indexes, a larger bubbler size indicating a stronger predictor.

Table 4 shows the final selected top 10 predictors. Race/ethnicity was consistently selected as one of the important predictors in sample 1 (S1, the total sample) to sample 6 (in females). In S1, the top 10 important predictors for classifying those with serum 25(OH)D concentrations ≥ 20 *vs* < 20 ng/mL were: race, serum folate level, sex, region, red blood cell folate (RBP, a marker of tissue stores of folate), poverty, educational level, glomerular filtration rate (GFR), serum uric acid (UAP), and alcohol

Table 4 Top 10 important predictors for classifying serum 25 Hydroxyvitamin D levels in the total and sex-specific samples

	S1: In both sex		S2: In males		S3: In females		S4: In both sex		S5: In males		S6: In females	
	≥ 20 vs < 20		≥ 20 vs < 20		≥ 20 vs < 20		≥ 30 vs < 30		≥ 30 vs < 30		≥ 30 vs < 30	
Rank	Predictor	MDG	Predictor	MDG	Predictor	MDG	Predictor	MDG	Predictor	MDG	Predictor	MDG
1	Race	519.5	Race	229.7	Race	267.7	Race	200.0	Race	95.2	Race	103.9
2	Folate	90.4	RBP	19.6	Folate	101.3	Age	97.5	WC	34.0	Age	79.4
3	Sex	57.3	Age	14.8	Age	77.5	Folate	65.1	BMI	29.0	GFR	50.6
4	Region	34.3	Glucose	14.2	GFR	73.7	GFR	54.8	AGE	28.9	Folate	49.6
5	RBP	34.0	GFR	13.1	RBP	50.7	BMI	48.8	Region	24.8	RBP	34.1
6	Poverty	33.3	Poverty	12.8	Poverty	50.1	WC	42.2	GFR	22.7	WC	32.0
7	Edu	30.3	Region	11.9	WC	22.9	Poverty	41.4	Insulin	18.9	SBP	26.9
8	GFR	29.6	WC	11.1	VEP	21.1	RBP	39.8	C1P	18.8	BMI	23.2
9	UAP	27.7	Folate	10.9	SBP	16.1	Glucose	22.0	RBP	18.6	VEP	22.0
10	Alcoh	21.6	Edu	9.8	A1C	14.6	DairyHEI	20.8	Poverty	17.4	Poverty	18.8
Accuracy, %	71.1		70.8		68.0		81.5		79.3		81.8	
Precision, %	68.4		68.2		68.2		83.0		82.4		85.6	
Recall, %	63.0		47.5		68.5		97.0		93.5		94.0	
F1 score, %	65.6		56.0		68.3		89.5		87.6		89.6	

S1-S3: Samples 1 to 3. S1 to S3 for the classification of serum 25(OH)D ≥ 20 vs <20 ng/mL. Of them, S1: In the total sample. S2: In males. S3: In females. S4-S6: Samples 4 to 6. S4 to S6 for the classification of serum 25(OH)D ≥ 30 vs <30 ng/mL. Of them, S4: in the total sample. S5: In males. S6: In females. MDG: Mean decrease in Gini index. A higher MDG index indicates higher variable importance; RBP: Red blood cell folate; Edu: Educational level; GFR: Glomerular filtration rate; UAP: Serum uric acid; Alcoh: Alcohol consumption; WC: Waist circumference; BMI: Body mass index; VEP: Serum concentrations of vitamin E; SBP: Systolic blood pressure; A1C: Hemoglobin A1c; DairyHEI: Dairy healthy eating index; C1P: Serum C-peptide.

consumption. Table 4 also gives the prediction accuracy rate (the proportion of the correct predictions for both true positives and true negatives), precision rate (the proportion of the correct predictions for the true positives among the total positive predications), the recall rate (the proportion of the true positive predictions among the total true (observed) cases, and F1 score (a weighted average of the precision and recall rates, where an F1 score reaches its best value at 1 and worst at 0). For example, in S1, the average accuracy rate to classify individuals with serum 25(OH)D concentrations ≥ 20 vs < 20 ng/mL was 71.1%, the precision rate was 68.4%, the recall rate was 63%, and the F1 score was 65.6%. These rates vary with sex-stratified samples and by the cutoffs of serum 25(OH)D concentrations.



Figure 2 Joint effect of serum 25 hydroxyvitamin D and metabolic syndrome on the risk of cardiovascular disease (A) and all-cause mortality (B). Serum 25(OH)D: Serum 25-hydroxyvitamin D; MetS: Metabolic syndrome; CVD: Cardiovascular disease; HR: Hazard ratio.

In the CART analyses, race/ethnicity, sex, age, and serum folate level were included as the important predictors in most prediction models (S1-S6). **Figure 4** further depicts these associations that non-Hispanic Black (NHB) had significantly lower mean serum 25(OH)D concentrations (16 ng/mL in females, and 18.5 ng/mL in males, **Figure 4A**) than NHW and the other race/ethnicity groups in both males and females. Overall, a decreased trend of serum 25(OH)D concentrations by age was observed in NHW (**Figure 4B**). However, among female NHB and males in the other racial/ethnic groups, serum 25(OH)D concentrations decreased after age 65 years older (**Figure 4B**). An increasing trend of mean 25(OH)D concentration was observed among females in the other racial/ethnic groups (**Figure 4B**). **Figure 4C** indicates an overall increasing trend of serum 25(OH)D concentrations with the increase in serum folate levels in both males and females and NHW, NHB, and the other racial/ethnic groups, except for male NHB with a decreased trend of serum 25(OH)D concentrations at around 10 ng/mL of serum folate concentrations, and with a decreased trend for male NHW and males in the other racial/ethnic groups at around 15 ng/mL of serum folate concentrations.

Finally, results from the sensitivity analysis, by the exclusion of those who died within the first year of follow-up from the longitudinal analysis, do not change the overall results. Therefore, we present the results without this exclusion in the report.

DISCUSSION

The main findings from this study indicate that: (1) Individuals with serum 25(OH)D deficiency and MetS had a significantly higher risk of CVD and all-cause mortality; (2) There was a significant joint effect of serum 25(OH)D concentrations and MetS on the mortality risk; and (3) Among both males and females (sample 1 of **Table 4**), the top 10 predictors were race/ethnicity, serum folate concentrations, sex, region, serum retinol-binding protein, poverty, educational level, glomerular filtration rate (GFR), serum uric acid, and alcohol consumption.

The impact of vitamin D deficiency on the risk of a variety of chronic diseases, including diabetes mellitus, multiple sclerosis, cancer, and CVD, has been reported in recent decades but is still a matter of debate[21-23]. Because findings from previous studies are inconsistent[9,22,24]. For example, in meta-analyses of eight prospective cohort studies, in which 2624 participants died of CVD during follow-up, individuals in the lowest quantile of serum 25 (OH)D concentrations had an increased risk of CVD and all-cause mortality[24]. However, this finding may be confounded by the heterogeneity of data from different study designs, adjustment for different covariates in each study, several studies were possibly

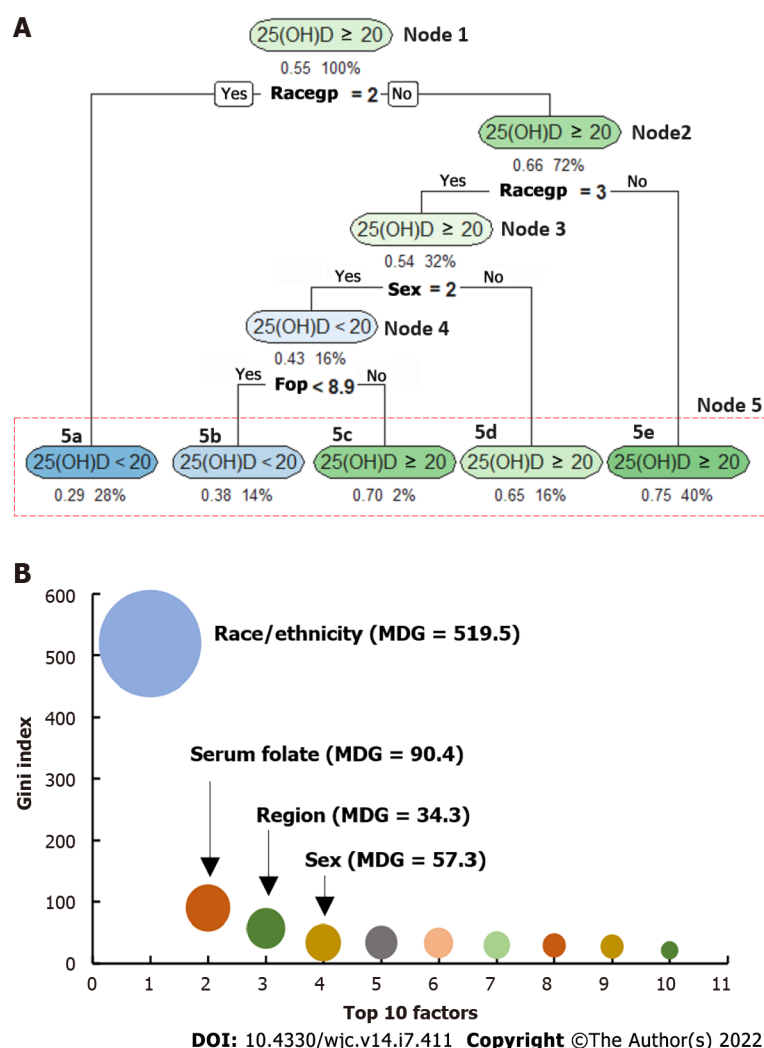


Figure 3 Example of classification and regression trees with serum 25 hydroxyvitamin D ≥ 20 vs < 20 ng/dL in both sexes. A: Decision tree. Node 1: is the "root" node. Nodes 2 to 5 are the "leaves". Node 5 is also called as the terminal node, including five sub-nodes (5a – 5e). Each "leaf" is a classification labeled for the output variable {serum 25-hydroxyvitamin D [25(OH)D]}, showing in each oval-shaped box. Under each box, the values listed on right indicates the % of total sample size, and the value listed on the left indicates the % of those with serum 25(OH)D concentration ≥ 20 ng/mL. If the value on the left-listed is > 0.5 then the corresponding oval-shaped box is labeled as 25(OH) ≥ 20 ng/mL, otherwise labeled as < 20 ng/mL; B: An example of top 10 important variables presented by their mean decrease in Gini (MDG) index, a higher MDG indicates a stronger predictor for the output variable. 25(OH)D: 25-hydroxyvitamin D; RACEGP: Race group, coded as racegp = 1 for non-Hispanic White, racegp = 2 for non-Hispanic Black, and racegp = 3 for all the other race groups; Sex = 1 for male, and sex = 2 for female; FOP: Serum folate level; MDG: Mean decrease in Gini index.

confounded by other lifestyle-related factors, such as body weight and physical activity. Studies have shown that serum 25(OH)D concentrations are significantly associated with obesity and physical activity[5,25], as well as MetS is significantly associated with other CVD risks[2]. In our analysis, we adjusted key covariates and further examined the joint effect of vitamin D deficiency and MetS on the risk of CVD and all-cause mortality, which also demonstrated that individuals with vitamin D sufficiency [serum 25(OH)D concentration ≥ 30 ng/mL] had a significant protective effect on CVD and all-cause mortality reduction even among those with positive MetS.

The mechanism by which vitamin D deficiency and MetS lead to an increased risk of CVD and all-cause mortality has not been fully elucidated. Several potential mechanisms have been hypothesized. First, the heart and the vasculature possess vitamin D receptors as well as the 1α -hydroxylase enzyme to activate serum 25(OH)D to 1,25-dihydroxyvitamin D ($1,25(\text{OH})_2\text{D}$), and thus are important target tissues for vitamin D. Second, several vitamin D effects on the electrophysiology, contractility, and structure of the heart suggest that vitamin D deficiency might be a causal factor for myocardial diseases [22,26]. Third, significant associations of vitamin D deficiency with obesity, hypertension, insulin resistance, and dyslipidemia (*i.e.*, the major components of MetS) have been reported by several researchers[5,24]. Although the complex association of vitamin D and MetS with risk of CVD requests more studies, findings of the joint effect of vitamin D deficiency and MetS, as well as a strong protective effect of vitamin D on CVD and all-cause mortality risk highlight the importance to control vitamin D deficiency and MetS for risk reduction of CVD and all-cause mortality.

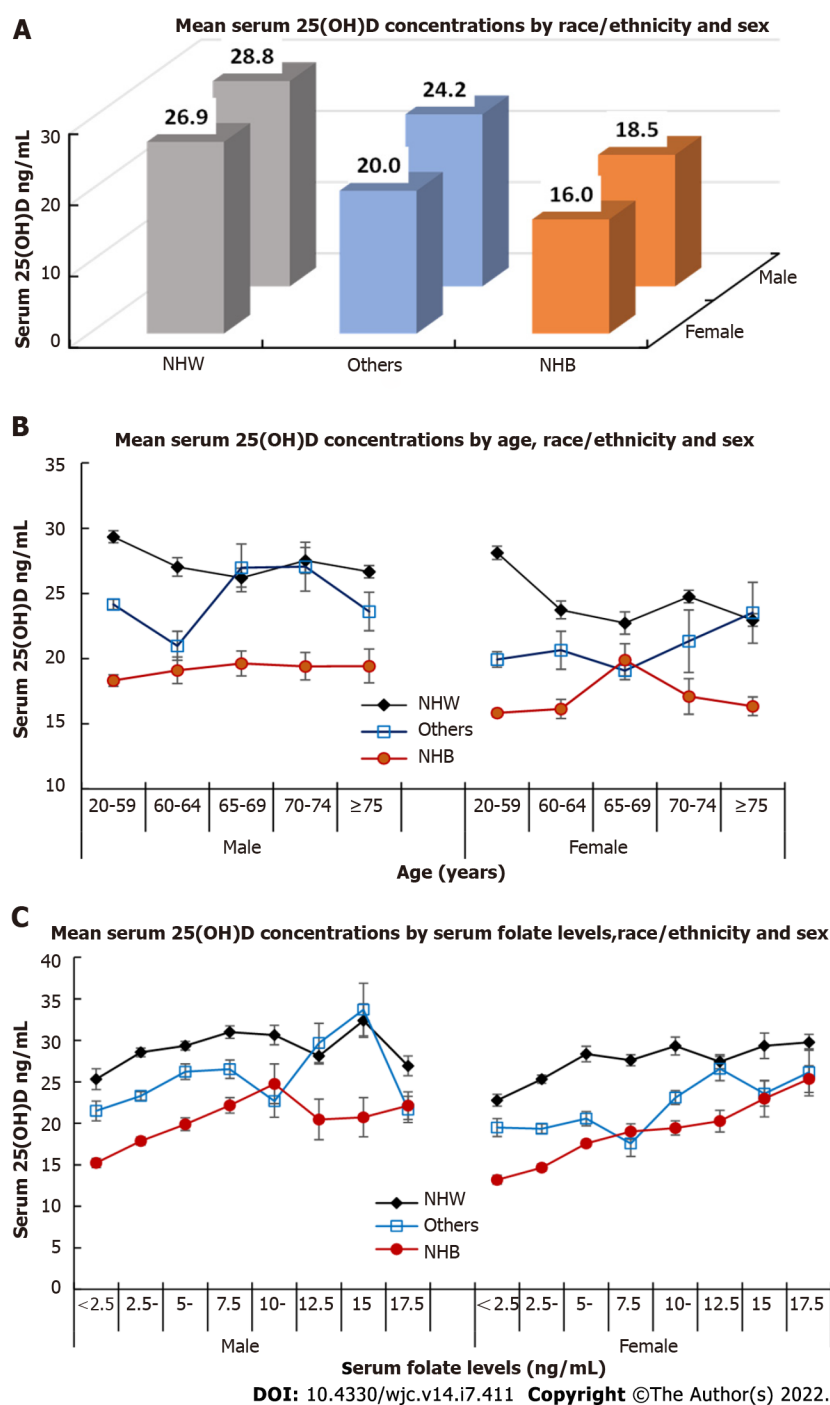


Figure 4 Mean Serum 25-hydroxyvitamin D concentrations (ng/mL) by race/ethnicity and sex (A), and by age (B), and serum folate levels (C). Serum 25(OH)D: Serum 25-hydroxyvitamin D; NHW: Non-Hispanic White; NHB: Non-Hispanic Black; Other: All other race/ethnicity.

In CART analysis, we observe that the top 10 important predictors varied by the study samples, and the cutoffs of the output variables [*i.e.*, serum 25(OH)D concentrations]. The main results indicate that race/ethnicity was one of the most important predictors. NHB had the lowest mean of serum 25(OH)D concentrations followed by the other racial/ethnic groups, and NHW. Differences in socioeconomic status and biological factors may explain the race/ethnic difference in mean serum 25(OH)D concentrations as well, including education, overweight and obesity, and prevalence of chronic conditions of hypertension, diabetes mellitus, and chronic kidney disease. Findings from the CART analysis suggest that to classify individuals with different serum 25(OH)D concentrations, sociodemographic factors (age, sex, poverty level, and regions) and biomarkers (BMI, waist circumference, glomerular filtration rate, and serum total folate, red blood folate, glucose, insulin, C-peptide, triglyceride, and serum vitamin E concentrations) were potential predictors. Of these biomarkers, serum folate and red blood cell folate concentrations had a relatively higher mean decrease in the GINI index (*i.e.*, a higher impact on the classification analysis) than several other biomarkers (Table 4). Folate is an essential micronu-

trient from diet or diet supplementation and is also synthesized locally by the intestinal microbiome [27]. A positive association between serum folate and serum 25(OH)D concentrations and a significant association between serum folate and muscle strength have been reported by other studies[28,29]. Although a detailed analysis of potential interactions of these biomarkers is beyond the scope of the present study, the role of serum folate with serum 25(OH)D in cardiovascular health and mortality warrants further study.

It should be noted that there are several limitations of the study when interpreting the main findings. First, the NHANES III does not include participants who were hospitalized at the time of the baseline survey, which may lead to an underestimation of the association between serum 25(OH)D concentrations and mortality risk. Because individuals with the severe disease might have been hospitalized or have died before the baseline survey. Second, the estimated associations of serum 25(OH)D concentrations and MetS with risk of CVD and all-cause mortality may be under- or over-estimated if the values of baseline serum 25(OH)D concentrations and MetS components changed during the follow-up. Third, the nonsignificant association of serum 25(OH)D concentrations and MetS with risk of cerebrovascular disease (CBVD) was possibly due to the small sample size of CBVD cases. Further studies are needed to focus on CBVD risk using a larger sample size. Last but not the least, although Machine Learning (ML) offers a useful tool to develop a prediction algorithm, findings from ML, including CART analysis offer an overall and useful data mining approach. For example, the results from Table 4 suggest that stratified samples are needed to improve the CART algorithms.

This study also has several strengths. First, the NHANES III, using a complex survey sampling approach, provides a nationally representative sample of residents living in the United States. Second, the survey design, process, and laboratory measures employed standard protocols that are supported and coordinated by the CDC. Third, the measures of CVD and all-cause mortality were highly reliable by using a standardized record process of the National Death Index system[13,30]. Fourth, we analyzed the association prospectively, which allowed us to test a potential causal association of baseline serum 25(OH)D concentrations and MetS with CVD and all-cause mortality risk. Last, to the best of our knowledge, the study is the first to examine the joint effect of vitamin D and MetS on the risk of CVD and all-cause mortality using a nationally representative sample and has the longest follow-up by using the NHANES III. Findings from the CART analysis add to new insights into classifying high-risk groups with lower vitamin D concentrations.

CONCLUSION

In conclusion, vitamin D deficiency and MetS were significantly associated with increased risk of CVD and all-cause mortality. There was a significant joint effect of vitamin D deficiency and MetS on the risk of mortality. Findings of the CART analysis may be useful to identify individuals positioned to benefit from interventions to reduce the risk of CVD and all-cause mortality.

ARTICLE HIGHLIGHTS

Research background

Studies that tested whether there is a significant joint effect of vitamin D intake and metabolic syndrome (MetS) on the risk of cardiovascular disease (CVD) and all-cause mortality are sparse.

Research motivation

The third National Health and Nutrition Examination Survey provides us an unique opportunity to test the research questions using a large-scale population sample size, with an average of 18 years follow-up. An integrated analysis approach of standard statistics methods and Machine Learning may provide new insights into the study field and add new evidence of clustering risk factors to control CVD and all-cause mortality.

Research objectives

To test the hypotheses that lower serum 25 hydroxyvitamin D [25(OH)D] concentrations (a marker of decreased vitamin D intake) and MetS have a long-term impact on the risk of CVD and all-cause mortality, and individuals with vitamin D deficiency can be detected by key covariates.

Research methods

A prospective analysis of 9094 adults who participated in the Third National Health and Nutrition Examination Survey in 1988 to 1994 and were followed for each participant's vital status by December 31, 2015, was conducted.

Research results

Findings from the study add new evidence to the body of research by highlighting the joint effects of vitamin D deficiency and MetS on the risk of CVD and all-cause mortality among the United States adults. A comprehensive intervention for both groups of risk factors are necessary to reduce the risk of CVD and all-cause mortality.

Research conclusions

There is a significant joint effect of vitamin D deficiency and MetS on the risk of CVD and all-cause mortality. The application of standard bio-statistics and Machine Learning techniques provides a new tool to test research hypothesis and provide new insights into health promotion for individuals who are at high risk of unhealthy exposures and risk of CVD and all-cause mortality.

Research perspectives

A high proportion of populations who have lower vitamin D levels and prevalent MetS poses a serious public health issue. Further studies are needed to examine the potential mechanisms by which that may cause vitamin D deficiency and MetS.

ACKNOWLEDGEMENTS

The study used data from the National Center for Health Statistics (NCHS). Findings and conclusions in this report are those of the authors and do not necessarily reflect the views or opinions of the NCHS.

FOOTNOTES

Author contributions: Liu L conceptualized the study and analysis designs, and performed the data analysis and drafted the manuscript; Cui S performed the machine learning analysis. Volpe SL, May NS, Sukumar D, DiMaria-Ghalili RA, Cui S, and Eisen H critically reviewed the study design and analysis methods, and carefully reviewed the results and edited the manuscript. All authors contributed to the study and approved the submission.

Institutional review board statement: The study was reviewed and approved by Drexel University Institutional Review Board (Approval No. 2105008546).

Conflict-of-interest statement: All authors report no relevant conflicts of interest for this article.

Data sharing statement: No additional data are available.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country/Territory of origin: United States

ORCID number: Longjian Liu 0000-0001-7956-7111; Saishi Cui 0000-0003-1108-920X; Stella L Volpe 0000-0002-8593-320X; Nathalie S May 0000-0002-6211-2564; Deeptha Sukumar 0000-0003-1150-4231; Rose Ann DiMaria-Ghalili 0000-0002-6146-7881; Howard J Eisen 0000-0003-1149-4344.

S-Editor: Ma YJ

L-Editor: A

P-Editor: Yu HG

REFERENCES

- 1 Holick MF. Vitamin D status: measurement, interpretation, and clinical application. *Ann Epidemiol* 2009; **19**: 73-78 [PMID: 18329892 DOI: 10.1016/j.annepidem.2007.12.001]
- 2 Romero DC, Manson J. Cardiology Patient Page: Vitamin D and Your Heart. *Circulation* 2015; **132**: e391-e392 [PMID: 26700012 DOI: 10.1161/CIRCULATIONAHA.115.019080]
- 3 Holick MF, Chen TC. Vitamin D deficiency: a worldwide problem with health consequences. *Am J Clin Nutr* 2008; **87**: 1080S-1086S [PMID: 18400738 DOI: 10.1093/ajcn/87.4.1080S]
- 4 Liu L, Chen M, Hankins SR, Nùñez AE, Watson RA, Weinstock PJ, Newschaffer CJ, Eisen HJ; Drexel Cardiovascular Health Collaborative Education, Research, and Evaluation Group. Serum 25-hydroxyvitamin D concentration and mortality

- from heart failure and cardiovascular disease, and premature mortality from all-cause in United States adults. *Am J Cardiol* 2012; **110**: 834-839 [PMID: [22658246](#) DOI: [10.1016/j.amjcard.2012.05.013](#)]
- 5 **Kabadi SM**, Lee BK, Liu L. Joint effects of obesity and vitamin D insufficiency on insulin resistance and type 2 diabetes: results from the NHANES 2001-2006. *Diabetes Care* 2012; **35**: 2048-2054 [PMID: [22751957](#) DOI: [10.2337/dc12-0235](#)]
- 6 **Cortese F**, Costantino MF, Luzi G, Di Marino S, Giordano P, Monitillo F. Vitamin D and cardiovascular disease risk. A literature overview. *Mol Biol Rep* 2022 [PMID: [35364717](#) DOI: [10.1007/s11033-022-07373-6](#)]
- 7 **D'Amelio P**. Vitamin D Deficiency and Risk of Metabolic Syndrome in Aging Men. *World J Mens Health* 2021; **39**: 291-301 [PMID: [33663024](#) DOI: [10.5534/wjmh.200189](#)]
- 8 **LaCroix AZ**, Kotchen J, Anderson G, Brzyski R, Cauley JA, Cummings SR, Gass M, Johnson KC, Ko M, Larson J, Manson JE, Stefanick ML, Wactawski-Wende J. Calcium plus vitamin D supplementation and mortality in postmenopausal women: the Women's Health Initiative calcium-vitamin D randomized controlled trial. *J Gerontol A Biol Sci Med Sci* 2009; **64**: 559-567 [PMID: [19221190](#) DOI: [10.1093/gerona/glp006](#)]
- 9 **Mursu J**, Robien K, Harnack LJ, Park K, Jacobs DR Jr. Dietary supplements and mortality rate in older women: the Iowa Women's Health Study. *Arch Intern Med* 2011; **171**: 1625-1633 [PMID: [21987192](#) DOI: [10.1001/archinternmed.2011.445](#)]
- 10 Plan and operation of the Third National Health and Nutrition Examination Survey, 1988-94. Series 1: programs and collection procedures. *Vital Health Stat 1* 1994; 1-407 [PMID: [7975354](#)]
- 11 **Herrick KA**, Storandt RJ, Afful J, Pfeiffer CM, Schleicher RL, Gahche JJ, Potischman N. Vitamin D status in the United States, 2011-2014. *Am J Clin Nutr* 2019; **110**: 150-157 [PMID: [31076739](#) DOI: [10.1093/ajcn/nqz037](#)]
- 12 **Centers for Disease Control and Prevention**. NHANES C-N. National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. Available from: <https://www.cdc.gov/nchs/nhanes/index.htm>
- 13 **Centers for Disease Control and Prevention**. National Center of Health Statistics. National Death Index. Vol 20082007. Available from: <https://www.cdc.gov/nchs/index.htm>
- 14 **Zittermann A**, Koerfer R. Vitamin D in the prevention and treatment of coronary heart disease. *Curr Opin Clin Nutr Metab Care* 2008; **11**: 752-757 [PMID: [18827580](#) DOI: [10.1097/MCO.0b013e328312c33f](#)]
- 15 **Ross AC**, Manson JE, Abrams SA, Aloia JF, Brannon PM, Clinton SK, Durazo-Arvizu RA, Gallagher JC, Gallo RL, Jones G, Kovacs CS, Mayne ST, Rosen CJ, Shapses SA. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. *J Clin Endocrinol Metab* 2011; **96**: 53-58 [PMID: [21118827](#) DOI: [10.1210/jc.2010-2704](#)]
- 16 **Grundey SM**, Brewer HB Jr, Cleeman JI, Smith SC Jr, Lenfant C; American Heart Association; National Heart, Lung, and Blood Institute. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation* 2004; **109**: 433-438 [PMID: [14744958](#) DOI: [10.1161/01.CIR.0000111245.75752.C6](#)]
- 17 **Breiman L**, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge; 2017
- 18 **Bzdok D**, Krzywinski M, Altman N. Points of Significance: Machine learning: a primer. *Nat Methods* 2017; **14**: 1119-1120 [PMID: [29664466](#) DOI: [10.1038/nmeth.4526](#)]
- 19 **R Core Team**. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. 2013
- 20 **Therneau T**, Atkinson B, Ripley B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-10. 2015
- 21 **Michos ED**, Blumenthal RS. Vitamin D supplementation and cardiovascular disease risk. *Circulation* 2007; **115**: 827-828 [PMID: [17309931](#) DOI: [10.1161/CIRCULATIONAHA.106.686238](#)]
- 22 **Fry CM**, Sanders TA. Vitamin D and risk of CVD: a review of the evidence. *Proc Nutr Soc* 2015; **74**: 245-257 [PMID: [25697289](#) DOI: [10.1017/S0029665115000014](#)]
- 23 **Ross AC**, Taylor CL, Yaktine AL, Del Valle HB. Institute of Medicine (US) Committee to Review Dietary Reference Intakes for Vitamin D and Calcium. Dietary reference intakes for calcium and vitamin D. Washington, DC: National Academies Press; 2011
- 24 **Schöttker B**, Jorde R, Peasey A, Thorand B, Jansen EH, Groot Ld, Streppel M, Gardiner J, Ordóñez-Mena JM, Perna L, Wilsgaard T, Rathmann W, Feskens E, Kampman E, Siganos G, Njølstad I, Mathiesen EB, Kubínová R, Pająk A, Topor-Madry R, Tamosiunas A, Hughes M, Kee F, Bobak M, Trichopoulou A, Boffetta P, Brenner H; Consortium on Health and Ageing: Network of Cohorts in Europe and the United States. Vitamin D and mortality: meta-analysis of individual participant data from a large consortium of cohort studies from Europe and the United States. *BMJ* 2014; **348**: g3656 [PMID: [24938302](#) DOI: [10.1136/bmj.g3656](#)]
- 25 **Kabadi SM**, Liu L, Auchincloss AH, Zakeri IF. Multivariate path analysis of serum 25-hydroxyvitamin D concentration, inflammation, and risk of type 2 diabetes mellitus. *Dis Markers* 2013; **35**: 187-193 [PMID: [24167365](#) DOI: [10.1155/2013/497256](#)]
- 26 **Cashman KD**. A review of vitamin D status and CVD. *Proc Nutr Soc* 2014; **73**: 65-72 [PMID: [24103845](#) DOI: [10.1017/S0029665113003595](#)]
- 27 **Eichholzer M**, Lüthy J, Gutzwiller F, Stähelin HB. The role of folate, antioxidant vitamins and other constituents in fruit and vegetables in the prevention of cardiovascular disease: the epidemiological evidence. *Int J Vitam Nutr Res* 2001; **71**: 5-17 [PMID: [11276922](#) DOI: [10.1024/0300-9831.71.1.5](#)]
- 28 **Rahman A**, Al-Taiar A, Shaban L, Al-Sabah R, Mojiminiyi O. Plasma 25-hydroxyvitamin D is positively associated with folate and vitamin B₁₂ levels in adolescents. *Nutr Res* 2020; **79**: 87-99 [PMID: [32653772](#) DOI: [10.1016/j.nutres.2020.06.003](#)]
- 29 **Wee AK**. Serum folate predicts muscle strength: a pilot cross-sectional study of the association between serum vitamin levels and muscle strength and gait measures in patients >65 years old with diabetes mellitus in a primary care setting. *Nutr J* 2016; **15**: 89 [PMID: [27756315](#) DOI: [10.1186/s12937-016-0208-3](#)]
- 30 **Blosnich JR**, Boyer TL. Concordance of Data About Sex From Electronic Health Records and the National Death Index: Implications for Transgender Populations. *Epidemiology* 2022; **33**: 383-385 [PMID: [35067566](#) DOI: [10.1097/EDE.0000000000001460](#)]