Available online at www.sciencedirect.com

ScienceDirect

Journal of Nutritional Biochemistry 142 (2025) 109919

Journal of Nutritional Biochemistry

RESEARCH PAPER

Dietary and lifestyle determinants of vitamin D status in the UK Biobank Cohort study for predictive modeling

Ángela Alcalá-Santiago[a,b,c,*], Belén García-Villanova[a], María Dolores Ruíz-López[a,c], Ángel Gil[b,c,d,e], Miguel Rodriguez-Barranco[b,f,g], Maria José Sánchez[b,f,g], Esther Molina-Montes[a,b,c,f]

[a] Department of Nutrition and Food Science, Faculty of Pharmacy, University of Granada, Granada, Spain
[b] Instituto de Investigación Biosanitaria ibs. GRANADA, Granada, Spain
[c] Institute of Nutrition and Food Technology (INYTA) 'José Mataix', Biomedical Research Centre, University of Granada, Avenida del Conocimiento s/n, Granada, Spain
[d] Department of Biochemistry and Molecular Biology II, Faculty of Pharmacy, University of Granada, Granada, Spain
[e] CIBER de Obesidad y Nutrición (CIBEROBN), Madrid, Spain
[f] CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[g] Escuela Andaluza de Salud Pública (EASP), Granada, Spain

Received 27 May 2024; received in revised form 31 October 2024; accepted 8 April 2025

Abstract

Vitamin D (VD) is involved in a wide variety of physiological processes. The high prevalence of VD deficiency in the population requires stronger preventive measures. The aim was to characterize the dietary and lifestyle determinants of VD levels in blood and of VD deficiency to further develop predictive models of these two outcomes. A total of 63,759 participants from the UK Biobank study with available data on dietary intake of VD, assessed via 24-hour recalls, and with measurements of serum 25(OH)D levels were included. Linear and logistic regression models were applied to identify factors associated with VD levels and VD deficiency outcomes, and to evaluate the influence of covariates on the association between VD in serum and VD in the diet. Predictive models for both VD outcomes were constructed using classical regression models and machine learning methods based on penalized likelihood methods. Approximately 10% of the participants had VD deficiency (VD < 25 nmol/L), and 38.9% were at risk of VD inadequacy (VD 25–49 nmol/L). The dietary intake of VD was significantly lower in the VD deficient group. This latter group showed lower engagement in physical activity (22.1%) compared to the non-deficient group (13.4%; $P<.001$). Also, overweight and obesity (vs normal weight) were related to a greater likelihood of VD deficiency (OR=1.18 and 1.96, respectively). A similar odds of VD deficiency was observed for abdominal obesity (OR=1.83). A weaker association was observed between dietary VD intake, based on participant reports, and VD levels. With regard to sunlight exposure, darker skin tones (OR dark vs fair skin=3.11), season (OR winter vs autumn=3.76) and less outdoor time activities (OR per 1 h increase=0.96) were also related to VD deficiency. Predictive models for both classical regression and machine learning, showed good accuracy (AUC=0.8–0.9 for VD deficiency). In conclusion, while a rich diet in VD boosts its levels, sun exposure plays a more significant role particularly in populations from the UK or Northern Europe. A predictive model including key determinants could effectively assess VD deficiency.

Keywords: Vitamin D status; Vitamin D deficiency; Diet; Lifestyle; Cohort; Predictive model.

1. Introduction

Vitamin D (VD), including cholecalciferol and ergocalciferol, is a fat-soluble vitamin involved in a wide variety of physiological processes and has multiple functions; it is the most important bone mineralization factor because it regulates the levels of calcium and phosphorus in the bone matrix [1,2]. Its regulatory impact extends to multiple aspects, including not only calcium homeostasis but also the endocrine system, the proliferation of skin keratinocytes, and notably, the modulation of the immune system, related to T

Abbreviations: AI, adequate Intakes; AIC, Akaike Information Criterion; AUC, Area under the curve; BMI, body mass index; CVD, cardiovascular disease; CIs, confidence intervals; EFSA, European Food Safety Authority; ENET, Elastic-Net; LASSO, least absolute shrinkage and selection operator; ORs, Odds Ratios; RMSE, root mean square error; R2, R squared; SE, standard error; UKB, UK Biobank; UVB, ultraviolet B radiation; VD, Vitamin D; 24-HR, 24-hour recalls; 25(OH)D, 25-hydroxyvitamin D.

* Corresponding author at: Angela Alcalá-Santiago, Department of Nutrition and Food Science, Faculty of Pharmacy. University of Granada, Granada 18071, Spain. Tel.: 958027450.
E-mail address: angela.alcala@ugr.es (Á. Alcalá-Santiago).

https://doi.org/10.1016/j.jnutbio.2025.109919

cells, monocytes and dendritic cells [3–5]. VD is also a crucial regulator of protein expression due to its ability to bind to various genes, influencing mRNA synthesis [3–5].

The primary source of VD in humans is sunlight exposure, together with dietary VD. It is synthesized in the skin after exposure to sunlight (ultraviolet B radiation, UVB). VD can also be obtained from the diet, which becomes essential when sunlight exposure is limited, as is the case during winter months at northern latitudes or when exposure to UVB is restricted [6]. Dietary intake is commonly derived from natural sources (such as fish and eggs), fortified food products (including dairies and cereals) or supplements [2]. Indeed, VD supplementation and fortified foods have gained significant importance in recent years, given the widespread deficiency of VD observed in the general population. In fact, observational data indicate that approximately 40% of the European population may experience VD deficiency or risk of inadequacy, with 13% facing deficiency [7]. Overall, levels of less than 20-30 ng/mL 25-hydroxycholecalciferol in the serum, also termed as 25-hydroxyvitamin D (25(OH)D), a metabolite of VD, are considered to indicate VD deficiency [8,9]. Recommended dietary intakes of VD provided by the European Food Safety Authority (EFSA) in terms of the Adequate Intakes (AI) refer to 15 μg/d for adults of both sexes [10]. However, there is limited evidence to affirm that sufficient intake levels of this vitamin exerts beneficial health effects beyond bone health [11], or whether this intake allows us to overcome VD deficiency [8]. Despite the lack of conclusive data, in recent years, numerous investigations have suggested a potential association between adequate VD levels and positive outcomes in conditions such as cancer [12], cardiovascular disease (CVD) [13], diabetes mellitus, and obesity [14], among other chronic diseases. However, the evidence remains mixed, and further research is needed to better understand the role of VD in these conditions. Additionally, diminished VD levels in blood have been associated with a heightened incidence of comorbidities in clinical contexts [11]. Hence, inconsistent or insufficient VD intake comibined with limited UVB exposure could lead to persistent VD deficiency, increasing the risk of developing certain diseases or exacerbating their severity, as previously highlighted.

The most common method for assessing VD status in individuals involves serum 25(OH)D levels. However, unnecessary blood tests are neither beneficial nor cost-effective, and their financial implications must be considered [15]. Implementing a screening strategy to anticipate serum VD levels and identify individuals at risk of VD deficiency could be highly valuable. Likewise, forecasting VD levels based on the subject's characteristics could be key to preventing unnecessary analytical determinations. Predictive models allow classification and quantitative assessments based on patterns identified in the data [2]. Classical statistical methods, such as stepwise regression, face several challenges in the selection and estimation of covariate effects [16]. Machine learning approaches for classification and prediction have become a new paradigm in predictive modeling. These methods use training sets to identify patterns and test sets to validate predictions [17,18]. Among these methods, penalized likelihood methods such as least absolute shrinkage and selection operator (LASSO) and Elastic-Net (ENET) are approaches that avoid problems related to the stability of the estimated parameters compared to classical methods [19,20].

Previous research has revealed associations between VD levels and various sociodemographic and lifestyle factors, including dietary VD intake, supplements use, sun exposure, age, adiposity, physical activity, season, and others [21,22]. Even race and socioeconomic status have emerged as potential determinants of VD status in populations from the UK [23]. Additionally, some studies have developed predictive models for VD deficiency in adults, with the majority of these studies focusing on populations from the USA, Netherlands, Australia, and Spain [2,24–26].

The aim of this study was to characterize the dietary and lifestyle factors associated with serum 25(OH)D levels, allowing us to explore in depth the determinants related to VD deficiency, using a large population sample from the UK Biobank (UKB) study. Based on this first assessment, we also aimed to apply predictive modeling approaches to predict serum 25(OH)D levels and, therefore, VD status and VD deficiency.

2. Methods

2.1. Study design and population

This study is based on data from the UKB cohort study. This initiative is designed to research both genetic and nongenetic factors influencing diseases in middle-aged and aged individuals. Its primary objective involves a meticulous evaluation of various exposures, accompanied by a thorough follow-up and characterization of diverse health-related outcomes [27].

The study included nearly half a million participants aged 40 to 69 years who were recruited from the UK during their baseline assessments, which were conducted between 2006 and 2010. The assessment involved 22 centers across the UK to encompass a diverse range of settings and of exposures. Various physical metrics, such as height, weight, and blood pressure, along with biological samples including blood, urine, and saliva, were collected [28]. Furthermore, additional phenotyping assessments across all participants or large subsets have been performed [27].

2.2. Data access and selection of participants

The data received approval from the UKB under reference number 76564. Thus, the data used was authorized by the data owner and adheres to ethical and legal standards. The information extraction was carried out between the months of January and July 2023. The inclusion criteria for this study were as follows: (i) complete blood VD level determinations and (ii) complete dietary information obtained from 24-hour recalls (24-HR). A total of 63,759 participants met the eligibility criteria for the study (Supplementary Fig. 1).

2.3. Assessment of VD status

Biochemical assays were conducted on blood samples obtained during the initial assessment. Briefly, serum samples were collected in a silica clot accelerator tube and stored at −80°C. These samples were subsequently processed in a central laboratory using an automated dispensing system. The serum 25(OH)D concentration was determined through a chemiluminescence immunoassay (DiaSorin LIAISON XL, Italy), certified by the VD Standardization-Certification Program of the Centers for Disease Control and Prevention. To ensure analytical precision, quali-ty control samples at various concentrations were analyzed, and the VD testing assay was validated through the RIQAS Immunoassay Specialty I EQA program (Randox Laboratories), an external quality assurance scheme [2,29].

2.4. Assessment of dietary VD

A 24-HR dietary questionnaire was emailed to participants with known work email addresses (~320,000 participants). They were administered four times between February 2011 and April 2012, in order to capture seasonal dietary variations and provide an average measure of habitual consumption. In addition, an online questionnaire, validated by the Cancer Epidemiology Unit for the UKB,

that included detailed questions on the intake of approximately 200 common foods and drinks over the past 24 h was administered. Participants indicated food amounts or used standard servings. The questionnaire also included inquiries about meal timings and locations [30].

2.5. Selection of covariates

The covariates were selected based on their relevance to 25 (OH)D serum levels in the body. The risk of nutritional insufficiency, related to the probability of not reaching the adequate intake of essential nutrients, is defined as intakes below 15 μg for VD. These data included age, sex, time spent outdoors in winter and summer, VD intake from the diet (24-HR), and supplement use. Additionally, covariates such as smoking, body mass index (BMI), waist circumference, physical activity, alcohol consumption and the deprivation index [31], which are proxies for socioeconomic status, were considered. All covariates were self-reported at baseline. The season was determined according to the month of blood sampling and categorized as summer (June-August), autumn (September-November), winter (December-February) or spring (March-May). As for the assessment center, the north-to-south location was set to one of three latitudes ($<51^0$, 52-53^0, and 54-55^0). Age was divided into three groups ($<$50, 50-60, $>$60), and BMI was determined at baseline using height and weight measurements (calculated as body weight in kilograms divided by height squared in meters) and categorized into three groups: normal weight ($<$25), overweight ($<$30), and obesity ($\geq$30) [32]. To ascertain the presence or absence of abdominal obesity, waist circumference was taken into account. For women, abdominal obesity was categorized as $>$88 cm, while for men, abdominal obesity was considered when the values were $>$102 cm [33]. In relation to physical activity, we considered both METs (vigorous, moderate, or walking activity, measured in minutes per week) and physical activity levels derived from the International Physical Activity Questionnaire IPAQ [34]. Physical activity was thereby categorized into three levels: low, moderate, and high. Alcohol consumption was assessed based on consumption status and frequency. Consumption status was classified into three levels: never, previous, and current. The frequency of alcohol consumption was categorized into five levels (daily, 3/4 per week, 1/2 per week, 1 to 3 per month, special occasion and never). The skin type and ease of skin tanning were considered to account for melanogenic differences and ethnic skin types. In particular, skin type was categorized into five levels (very fair, fair, light olive, dark olive, brown and black), while the ease of skin tanning was categorized into four levels (very tanned, moderately tanned, mildly tanned, and never tanned).

2.6. Statistical analysis

All analyses were performed using R v 4.3.0 [35]. For continuous variables, descriptive statistics were based on medians along with the 25th (p25) and 75th (p75) percentiles since all variables followed a nonnormal distribution (Shapiro-Wilks test, $P<.05$). For categorical variables, the relative frequencies and percentages were calculated. The Chi-square test and the Kruskall-Wallis test (nonparametric) were applied to evaluate differences between groups within categorical and continuous variables, respectively. The initial significance level was set at 0.05; the p-value adjusted for pairwise comparisons (Type I error) by Bonferroni (0.05/n tests) was set at 0.001.

Linear and logistic regression models were fitted to examine the associations of dietary and lifestyle variables with VD status (continuous outcome variable) and with VD deficiency (binary outcome variable, yes vs no), respectively. For analyses involving serum 25(OH)D levels as a continuous outcome measure, we applied natural logarithmic transformation to account for its skewed distribution in the linear regression model. All models were adjusted for age, sex, and center. Subsequently, models were individually adjusted for obesity (overall and abdominal), skin type, season, smoking status, physical activity, and other relevant covariates. In these linear regression models, the results were conveyed in terms of the beta coefficient and corresponding confidence intervals (CIs), the standard error (SE), and the R squared (R2). With respect to the logistic regression model, by which the association between VD deficiency and the aforementioned covariates was examined, the results are presented as Odds Ratios (ORs), along with the upper and lower CIs.

To assess the association between serum 25(OH)D levels and the dietary VD intake (adjusted for energy intake), we applied linear regression models in which log-transformed serum 25(OH)D was considered as the outcome variable and dietary VD intake was the main predictor variable. For the latter, we considered log2 transformed VD to represent a doubling of intake. The models were adjusted for age, sex and center, and subsequently adjusted for the other variables including physical activity, time spent outdoors in winter and summer, BMI, waist circumference, skin type and color, season of the year, and alcohol and tobacco consumption. This comprehensive adjustment aimed to discern the multifaceted influence of these factors on the association between blood 25(OH)D levels and dietary intake. To account for multiple testing issues, adjusted p-values obtained by Bonferroni correction were deemed significant. In addition, the strength of the association between serum 25(OH)D and dietary VD was explored via Spearman correlation analyses.

We evaluated effect modification on these associations by variables such as sex, age, BMI and season, by introducing an interaction term defined as VD*covariate in the model. Interaction was deemed significant for P-values$<.05$ (Wald test). In addition, subgroup analyses by these variables were carried out to evaluate potential effect modification further.

For the prediction models, we ran both classical and machine learning approaches. Overall, the selected variables were ranked according to variable importance. The prediction models were built on a dataset with complete observations (N=48,913) and 21 variables, with 70% of the participants (N=34,239) used for the training set and the remaining 30% (N=14,457) for the test set.

In the first case, we applied multiple linear and logistic regression to select the predictors in the model. The absence of collinearity was verified in the models. Herewith, the model used to predict either VD levels (continuous outcome) or VD deficiency (binary outcome) was determined from a set of predictors (x1, x2, x3, ...) using the stepAIC function (both stepwise selection procedures) in R. The performance of the models was evaluated via the Akaike Information Criterion (AIC). For internal validation purposes, we randomly divided the data into a training set (70% for creating a predictive model) and a test set (30% for evaluating the model). The best classification (VD deficiency, yes vs no) model was defined as the model that had the lowest classification error rate or highest area under the curve (AUC) and accuracy in the test data ("pROC" package in R). To forecast VD levels as the outcome, the model was defined in terms of the root mean square error (RMSE) by calculating the average difference between the observed and predicted values and the R2, with a lower RMSE and higher R2 indicating a better model.

In the second case, several feature selection methods based on penalized likelihood methods (Ridge Regression, least absolute shrinkage and selection operator-LASSO or ENET) were applied using the "caret" and "glmnet" R packages. These methods adjust for collinearity through a penalization approach [17–20]. To select the

model parameters (alpha and lambda), the data were again randomly split into a test (30% of the observations) and training set (70% of the observations) in a 100 times repeated 10-fold cross-validation. The estimated performance was obtained from the average of all repetitions. The accuracy and kappa metrics were used to choose the parameters and best predictive model for VD deficiency ("MLeval" package in R), whereas the RMSE and R2 were considered for VD levels.

The ENET method showed the highest performance metrics for predicting VD deficiency (accuracy=0.91) and VD levels in blood (RMSE=18.5). The estimated tuning parameters were: alpha=0.889 and lambda=0.00753 for VD deficiency, and alpha=1 and lambda=0.01 for VD levels.

2.7. Ethics

The study received ethical approval from the Granada Research Ethics Committee (CEIm/CEI Provincial de Granada) to utilize the UKB data (reference: 6/21). Additionally, approval for ethical considerations concerning the UKB was obtained from the National Information Governance Board for Health and Social Care and the North West Multicentre Research Ethics Committee (11/NW/0382). This analysis was conducted in accordance with UKB application 76564.

3. Results

3.1. Characteristics of the study population

Among the 63,759 participants considered in this analysis (Table 1), 6,215 (9.75%) exhibited VD deficiency, and 24,806 (38.9%) were at risk of VD inadequacy (Table S1). The mean 25(OH)D concentration in the entire cohort was 50.7 nmol/L; the median 25(OH)D concentration in participants with VD deficiency was 20.2 nmol/L, while in participants without VD deficiency, it was 53.5 nmol/L. Notably, the dietary intake of VD was significantly lower in the VD-deficient group. Around 80% of the participant were at risk of nutritional inadequacy for VD (<10 μg/d). Participants with VD deficiency were younger (median age=54.0 years) than nondeficient participants (median age=58.0 years) (p<.001). Women seemed to be more prone to VD deficiency than men were (P=.009). Indeed, women had lower 25(OH)D serum concentrations (Table S2). There were also significant differences according to socioeconomic status with regard to VD deficiency, with VD deficient participants having a greater deprivation index than their counterparts.

Concerning physical activity, the deficient group exhibited lower engagement in physical activity (22.1%) than did the nondeficient group (13.4%) (p<.001). The level of moderate physical activity was significantly lower in the VD-deficient group (280 min/wk) than in the nondeficient group (480 min/wk) (p<.001). Similarly, walking activity was lower in the deficient group (528 min/wk) than in the nondeficient group (693 min/wk) (p<.001).

With regard to other lifestyle factors, the smoking prevalence was greater in the deficient group (12.3%) than in the the nondeficient group (7.99%) (P<.001). Alcohol consumption patterns also differed, with a lower proportion of current drinkers in the VD-deficient group (86.9%) than in the nondeficient group (93.3%) (P<.001). VD deficiency was also associated with a greater prevalence of obesity (32.5%) and abdominal obesity (39.6%), while the nondeficient counterpart exhibited lower rates of both obesity measures (obesity: 21.7%, abdominal obesity: 27.7%) (P<.001). Additionally, the median concentration of 25(OH)D differed according to overall obesity status (Table S3), and abdominal obesity subgroups (Table S4).

Focusing on sun exposure and its impact on VD levels, individuals with VD deficiency reported spending less time outdoors both in winter and in summer (P<.001). The distribution of skin types varied significantly between the two groups (P<.001). The nondeficient group tended to have more fair skin (67.6%) than the deficient group (56.5%). VD supplementation practices were also more common in the deficient group than in the non-deficient group (P<.001). By season at which blood was drawn, it was observed that VD deficiency was more prevalent in winter than in summer (P<.001). In addition, VD levels were shown to vary by season (Table S5). Moreover, VD deficiency was to be more frequent in centers of northern latitudes.

3.2. Dietary and lifestyle variables associated with VD deficiency and VD status

The dietary and lifestyle variables associated with VD deficiency and VD status are shown in Table 2. Significant associations (at p value level<0.001) between VD deficiency/status and various variables were detected via multivariate models adjusted for age, sex and center. These estimates were close to those obtained in univariate analyses (Table 1). Dietary intake of VD was positively related to blood levels of VD (β per 1μg VD/2000 kcal=0.004) but inversely related to VD deficiency (OR=0.94). An increase of one unit in BMI was associated with a 6% greater likelihood of VD deficiency. Similarly, each additional centimeter in waist circumference was associated with a 2.7% increase in the odds of VD deficiency. Likewise, with respect to normal weight, overweight increased the odds by 18%, while obesity did so by 96%. Abdominal obesity in nonobese subjects also increased the risk of VD deficiency (OR per 1 unit increase=1.83. An increase of 1 unit in body mass index (BMI) was linked to a decrease of 0.017nmol/L in VD levels. Also, per 1-unit increase in waist circumference VD levels decreased on average by 0.007nmol/L.

Regarding physical activity, moderate and high levels (vs low levels) had protective effects against VD deficiency, reducing the odds by 35% and 59%, respectively.

Individuals with darker skin (vs. very fair skin) had greater odds of having VD deficiency (black skin: OR=3.12), and those who never tanned (vs. very tanned) also had an increased risk of VD deficiency (OR=1.5). Additionally, the duration of outdoor activities in both winter and summer was inversely associated with vitamin D deficiency (OR per 1-h increase in outdoor activities=0.96 and 0.93, respectively). Each additional hour spent outdoors in winter was associated with an increase of 0.010 nmol/L in the serum 25(OH)D levels. Concerning seasons, in summer, there was a significant decrease of 37.7% in the odds of this deficiency (OR: 0.62) compared with autumn; in contrast, in winter, there was a 3.8 fold increase in the odds of VD deficiency.

Furthermore, VD deficiency was determined by various lifestyle factors. There was an 11.2% decrease in the odds of VD deficiency among former smokers compared to nonsmokers and a 44.5% increase in the odds of VD deficiency among current smokers. Overall, former smokers had on average 0.019 lower levels of serum 25(OH)D levels than nonsmokers did; in current smokers, the levels decreased by 0.08 units. Regarding alcohol consumption, former drinkers had a 39.4% lower odds of VD deficiency than nondrinkers did (increased VD levels: β=0.12); moreover, current drinkers had a 62% reduction of this odds (increased VD levels: β=0.21). In addition, drinking only on special occasions was associated with an increased likelihood of VD deficiency compared to daily alcohol consumption (OR: 1.86). Importantly, those who had never consumed alcohol showed 2.3 times higher odds of having VD deficiency (increased VD levels: β=0.17).

Table 1
Characteristics of the study population (n=63,759), overall and by VD status deficiency.

	Overall		VD deficiency (nmol/L)		
	Median (p25,p75) N (%)	N=63,759* N	No; N=57,544 Median (p25,p75) N (%)	Yes; N=6,215 Median (p25,p75) N (%)	OR (95% CI)
Serum 25(OH)D (nmol/L)	50.7 (36.1;65.5)	63,759	53.5 (40.6;67.3)	20.2 (16.9;22.8)	NA
VD μg/2,000 kcal	1.64 (0.81;3.09)	63,759	1.67 (0.83;3.15)	1.36 (0.64;2.53)	0.94 (0.93;0.95)
VD μg in tertiles		63,759			
T1: (0.3,0.8)	21,267 (33.4%)		18,721 (32.5%)	2,546 (41.0%)	Ref.
T2: (1.3,2.0)	21,328 (33.5%)		19,298 (33.5%)	2,030 (32.7%)	0.77 (0.73;0.82)
T3: (3.2,9.1)	21,163 (33.2%)		19,525 (33.9%)	1,639 (26.4%)	0.62 (0.58;0.66)
VD supplementation		26,023			
No	25,598 (40.1%)		24,292 (42.2%)	1,306 (21.0%)	Ref.
Yes	425 (0.67%)		414 (0.72%)	11 (0.18%)	0.50 (0.26;0.87)
Age (years)	57.0 (49.0;63.0)	63,759	58.0 (50.0;63.0)	54.0 (47.0;61.0)	0.97 (0.96;0.97)
Sex:		63,759			
Men	29,036 (45.5%)		26,303 (45.7%)	2,733 (44.0%)	Ref.
Women	34,723 (54.5%)		31241 (54.3%)	3,482 (56.0%)	1.07 (1.02;1.13)
Centre		63,759			
$\leq 51^0$	32,092 (50.3%)		28,894 (50.2%)	3,198 (51.5%)	Ref.
$52–53^0$	27,921 (43.8%)		25,123 (43.7%)	2,798 (45.0%)	1.01 (0.95;1.06)
$54–55^0$	3,746 (5.88%)		3,527 (6.13%)	219 (3.52%)	0.56 (0.49;0.64)
Multiple deprivation index	12.7 (7.58;21.6)	62,231	12.3 (7.46;20.9)	16.8 (9.65;27.1)	0.548 (0.545; 0.550)*
Smoking status		63,579			
Never	35,947 (56.4%)		32,376 (56.3%)	3,571 (57.5%)	Ref.
Previous	22,270 (34.9%)		20,416 (35.5%)	1,854 (29.8%)	0.82 (0.78;0.87)
Current	5,362 (8.41%)		4,599 (7.99%)	763 (12.3%)	1.50 (1.38;1.64)
Alcohol status		63,691			
Never	2,397 (3.76%)		1,899 (3.30%)	498 (8.01%)	Ref.
Previous	2,221 (3.48%)		1,918 (3.33%)	303 (4.88%)	0.60 (0.52;0.70)
Current	59,073 (92.7%)		53,674 (93.3%)	5,399 (86.9%)	0.38 (0.35;0.43)
Alcohol frequency		63,706			
Daily	13,906 (21.8%)		12,783 (22.2%)	1,123 (18.1%)	Ref.
3/4 per week	14,806 (23.2%)		13,746 (23.9%)	1,060 (17.1%)	0.88 (0.80;0.96)
1/2 per week	15,963 (25.0%)		14,603 (25.4%)	1,360 (21.9%)	1.06 (0.98;1.15)
1 to 3 per month	7,312 (11.5%)		6,493 (11.3%)	819 (13.2%)	1.44 (1.31;1.58)
Special occasion	7,086 (11.1%)		6,049 (10.5%)	1,037 (16.7%)	1.95 (1.78;2.13)
Never	4,633 (7.27%)		3,830 (6.66%)	803 (12.9%)	2.39 (2.16;2.63)
IPAQ physical activity		53,420			
Low	9,105 (14.3%)		7,730 (13.4%)	1,375 (22.1%)	Ref.
Moderate	22,099 (34.7%)		19,832 (34.5%)	2,267 (36.5%)	0.64 (0.60;0.69)
High	22,216 (34.8%)		20,712 (36.0%)	1,504 (24.2%)	0.41 (0.38;0.44)
Moderate activity (METs) (min/wk)	480 (160;1,200)	53,420	480 (160;1,260)	280 (40.0;840)	0.565 (0.562; 0.567)†
Vigorous activity (METs) (min/week)	240 (0.00;960)	53,420	240 (0.00;960)	0.00 (0.00;480)	0.563 (0.561; 0.565)†
Walking activity (METs) (min/wk)	693 (330;1,386)	53,420	693 (330;1,386)	528 (248;1,040)	0.563 (0.561; 0.566)†
Time outdoors in winter (h/d)	1.00 (1.00;2.00)	60,793	1.00 (1.00;2.00)	1.00 (0.00;2.00)	0.88 (0.87;0.90)
Time outdoors in summer (h/d)	3.00 (2.00;5.00)	60,788	3.00 (2.00;5.00)	2.00 (2.00;4.00)	0.86 (0.85;0.87)
BMI (kg/m^2)	26.5 (23.9;29.6)	63,609	26.4 (23.8;29.4)	27.5 (24.4;31.4)	1.06 (1.05;1.06)
Waist circumference (cm)	89.0 (80.0;98.9)	63,696	89.0 (79.0;98.0)	92.0 (82.0;102)	1.02 (1.02;1.02)
Overall obesity:		63,609			
Normal weight	22,724 (35.6%)		20,900 (36.3%)	1,824 (29.3%)	Ref.
Over weight	26,407 (41.4%)		24,077 (41.8%)	2,330 (37.5%)	1.11 (1.04;1.18)
obesity	14,478 (22.7%)		12,461 (21.7%)	2,017 (32.5%)	1.85 (1.73;1.98)
Abdominal obesity:		63,696			
No	45,290 (71.0%)		41,550 (72.2%)	3,740 (60.2%)	Ref.
Yes	18,406 (28.9%)		15,947 (27.7%)	2,459 (39.6%)	1.71 (1.62;1.81)

(*continued on next page*)

Table 1 (*continued*)

	Overall		VD deficiency (nmol/L)		
	Median (p25,p75) N (%)	*N*=63,759* N	No; *N*=57,544 Median (p25,p75) N (%)	Yes; *N*=6,215 Median (p25,p75) N (%)	OR (95% CI)
Skin type		63,040			
Very fair	5,035 (7.90%)		4,410 (7.66%)	625 (10.1%)	Ref.
Fair	42,410 (66.5%)		38,901 (67.6%)	3,509 (56.5%)	0.64 (0.58;0.70)
Light olive	11,833 (18.6%)		10,945 (19.0%)	888 (14.3%)	0.57 (0.51;0.64)
Dark olive	1,091 (1.71%)		953 (1.66%)	138 (2.22%)	1.02 (0.84;1.24)
Brown	2,126 (3.33%)		1,350 (2.35%)	776 (12.5%)	4.05 (3.59;4.58)
Black	545 (0.85%)		369 (0.64%)	176 (2.83%)	3.37 (2.76;4.10)
Ease of skin tanning		62,301			
Very tanned	13,588 (21.3%)		12,408 (21.6%)	1,180 (19.0%)	Ref.
Moderately tanned	25,178 (39.5%)		23,156 (40.2%)	2,022 (32.5%)	0.92 (0.85;0.99)
Midly tanned	13,185 (20.7%)		11,753 (20.4%)	1,432 (23.0%)	1.28 (1.18;1.39)
Never tanned	10,350 (16.2%)		9,060 (15.7%)	1,290 (20.8%)	1.50 (1.38;1.63)
Season		63,759			
Autumn	19,367 (30.4%)		18,127 (31.5%)	1,240 (20.0%)	Ref.
Spring	15,616 (24.5%)		2,739 (44.1%)	2,739 (44.1%)	3.11 (2.90;3.34)
Summer	21,884 (34.3%)		21,016 (36.5%)	868 (14.0%)	0.60 (0.55;0.66)
Winter	6,892 (10.8%)		5,524 (9.60%)	1,368 (22.0%)	3.62 (3.33;3.93)

[1]Numbers do not sum up due to missingness. Data for continuous variables are expressed in median (p25-p75), and data for categorical variables are presented as frequencies and percentages. VD deficiency was established for values less than 25 nmol/L 25(OH)D. The Chi square test for categorical variables and the Kruskall-Wallis test for continuous variables were used to assess differences according to VD deficiency groups. Differences according to VD deficiency were statistically significant: $P<.001$ for all covariates and for sex ($P=.009$). VD intake was adjusted for energy intake, considering a 2000 kcal diet.
†Calculated by standard deviation increase. NA, not applicable since VD levels were used as cutoff points to define VD deficiency. T, Tertile.

Finally, the variable that explained most of the variance of serum 25(OH)D levels was the season (R2=0.106), followed by skin type and obesity. Dietary VD intake, which was determined based on a subjective questionnaire, explained only 1.5% of the variance of VD levels.

3.3. The association between dietary VD and VD status, and the effect of other covariates on this association

There was no correlation between the serum 25(OH)D concentrations and the energy-adjusted dietary intake of VD (rho=0.05). Likewise, dietary VD intake from the 24 h preceding blood sampling showed a weak correlation with serum 25(OH)D levels ($r=0.07$). The correlations remained similar when correlations were explored among individuals with and without VD deficiency (rho=0.03 in both groups), suggesting a weak relationship between VD status and dietary VD. Table 3 shows the associations between serum levels of 25(OH)D and dietary intake of VD, and the impacts of other covariates have on this association. With multivariate adjusted regression models adjusted for age, sex and center (Model 1), we observed a significant, albeit weak association between both VD measures (β for doubling intakes of energy-adjusted dietary VD=0.03). The explained variance of this model was 0.015. After adjusting for anthropometric variables, the association between serum 25(OH)D and dietary VD became weaker ($\beta=0.01$) and the explained variance increased up to 0.046. Thus, anthropometric variables such as BMI and/or waist circumference strongly influence to a high extent the association between serum and dietary VD, although the association remains statistically significant. When physical activity-related variables were considered in the model, we also observed a notable change in the estimates ($\beta=0.01$) and an increase in the explained variance (R2=0.02-0.04). For other sunlight exposure-related variables, when skin type was added, ease of skin tanning, or time spent outdoors were included in Model 1, the association between the serum 25(OH)D and dietary VD also remained statistically significant albeit with more modest effect sizes ($\beta=0.01$). Collectively, these variables also contributed to a higher explained variance in the serum 25(OH)D levels (R2=0.02-0.06). Importantly, season had the greatest influence on the association between serum 25(OH)D and dietary VD (R2=0.121). Concerning smoking and alcohol-related variables, statistically significant associations remained, despite the effect sizes being small. The combined impact of smoking and alcohol consumption on the association between serum and dietary VD was also notable. Thus, an association between serum and dietary VD was detected, with several variables affecting this association.

There was no evidence of effect modification by any covariate ($P>.05$) on the association between VD in blood and dietary intake of VD (data not shown). Subgroup analyses by sex and other variables (obesity and season) also did not reveal any evidence for effect modification (data not shown).

3.4. Predictive models to estimate VD deficiency and VD status

3.4.1. Using Classical Regression Models in Predictive Modeling

Table 4 shows the predictive models obtained to forecast VD deficiency and the estimates of the selected variables. The predictive model of VD deficiency (binary model) with the lowest AIC included the following variables: sex, age, center, IPAQ, vigorous METS, smoking status (current), alcohol consumption, skin type (fair and brown) and ease of skin tanning, season, waist circumference, BMI (overweight), time spent outdoors in winter and summer, energy-adjusted dietary VD, and deprivation index. This model showed the lowest classification error rate (0.91). The predictive ability of this model in terms of the AUC was 0.8, suggesting that the ability of the model to distinguish between individuals with and without VD deficiency is high.

Table 2
Association between dietary and lifestyle variables and VD deficiency or VD status among UKB participants.

	VD deficiency (nmol/L)			VD levels nmol/L			
	OR	(95% CI)	*P*-value	Beta (β)	SE	R2	*P*-value
Energy-adjusted dietary VD per 1 unit (μg)	0.94	(0.930;0.950)	<.001	0.004	0.0003	0.015	<.001
Anthropometric variables							
BMI (kg/m2) per 1 unit	1.060	(1.055;1.065)	<.001	−0.017	0.000	0.042	<.001
Waist circumference (cm) per 1 unit	1.027	(1.027;1.029)	<.001	−0.007	0.000	0.049	<.001
Overall obesity:						0.041	
Normal weight	1. 00	(Ref)		1. 00	(Ref)		
Overweight	1.180	(1.106;1.260)	<.001	−0.064	0.004		<.001
obesity	1.963	(1.834;2.101)	<.001	−0.195	0.004		<.001
Abdominal obesity:						0.039	
No	1. 00	(Ref)		1. 00	(Ref)		
Yes	1.829	(1.731;1.931)	<.001	−0.163	0.003		<.001
Sunlight exposure variables							
IPAQ physical activity						0.038	
Low	1. 00	(Ref)		1. 00	(Ref)		
Moderate	0.651	(0.605;0.700)	<.001	0.093	0.005		<.001
High	0.410	(0.379;0.443)	<.001	0.198	0.005		<.001
Skin type						0.046	
Very Fair	1. 00	(Ref)		1. 00	(Ref)		
Fair	0.657	(0.600:0.720)	<.001	0.105	0.006		<.001
Light Olive	0.572	(0.513;0.638)	<.001	0.149	0.007		<.001
Dark Olive	1.012	(0.827;1.230)	.9065	0.084	0.014		<.001
Brown	3.841	(3.399;4.344)	<.001	−0.342	0.011		<.001
Black	3.108	(2.543;3.789)	<.001	−0.324	0.019		<.001
Time outdoors in winter (hour/day)	0.960	(0.955;0.965)	<.001	0.010	0.000	0.022	<.001
Time outdoors in summer (hour/day)	0.933	(0.928;0.939)	<.001	0.019	0.000	0.033	<.001
Ease of skin tanning						0.027	
Very Tanned	1. 00	(Ref)		1. 00	(Ref)		
Moderately Tanned	0.924	(0.857;0.996)	.0391	−0.011	0.004		.0172
Mildly Tanned	1.247	(1.149;1.354)	<.001	−0.092	0.005		<.001
Never Taned	1.499	(1.378;1.631)	<.001	−0.137	0.005		<.001
Season						0.106	
Autumn	1. 00	(Ref)		1. 00	(Ref)		
Spring	3.206	(2.986;3.443)	<.001	−0.252	0.004		<.001
Summer	0.623	(0.570;0.681)	<.001	0.063	0.004		<.001
Winter	3.759	(3.459;4.085)	<.001	−0.303	0.005		<.001
Lifestyle variables							
Smoking status						0.015	
Never	1. 00	(Ref)		1. 00	(Ref)		
Previous	0.888	(0.837;0.943)	<.001	0.019	0.003		<.001
Current	1.445	(1.327;1.572)	<.001	−0.081	0.006		<.001
Alcohol status						0.019	
Never	1. 00	(Ref)		1. 00	(Ref)		
Previous	0.606	(0.517;0.708)	<.001	0.121	0.013		<.001
Current	0.380	(0.343;0.422)	<.001	0.209	0.009		<.001
Alcohol frequency						0.027	
Daily	1. 00	(Ref)		1. 00	(Ref)		
3/4 per week	0.829	(0.760;0.906)	<.001	0.022	0.005		<.001
1/2 per week	0.977	(0.899;1.062)	.5841	0.0015	0.005		.775
1 to 3 per month	1.311	(1.191;1.443)	<.001	−0.0623	0.006		<.001
Special occasion	1.861	(1.698;2.039)	<.001	−0.135	0.006		<.001
Never	2.302	(2.085;2.540)	<.001	−0.170	0.007		<.001

Logistic (VD deficiency) and linear (log-transformed serum levels of VD) regressions were conducted to assess the effect of various variables (main predictors) on the outcome VD variables in multivariate models adjusted for age, sex and center. Regression models were conducted, for VD intake variables (energy-adjusted dietary VD), anthropometric measures (BMI and waist circumference), sun exposure variables (IPAQ, walking activity, moderate activity and vigorous activity (METs), skin type, ease of skin tanning, time outdoors in winter, time outdoors in summer and season), and for lifestyle factors (smoking, alcohol status and alcohol frequency). R2=*R*-squared.

Table 3
Influence of energy-adjusted dietary VD (log2 transformed) on serum 25(OH)D levels, considering the impact of other variables on the associations among UKB cohort participants.

	VD levels nmol/L			
	Beta (β)	SE	(95%CI)	R2
Model 1				
Age, sex, center	0.03	0.002	(0.026;0.033)	0.015
Model 1 + anthropometric variables				
BMI (kg/m2) per 1 unit	0.01	0.001	(0.009;0.012)	0.046
Waist circumference (cm) per 1 unit	0.01	0.001	(0.009;0.011)	0.052
BMI+WC	0.03	0.002	(0.026;0.033)	0.054
Model 1 + sunlight exposure variables				
IPAQ	0.009	0.001	(0.008;0.011)	0.040
Walking activity (METs) (min/week)	0.009	0.001	(0.008;0.011)	0.021
Moderate activity (METs) (min/week)	0.009	0.001	(0.008;0.011)	0.024
Vigorous activity (METs) (min/week)	0.009	0.001	(0.008;0.011)	0.030
Skin type	0.008	0.001	(0.007;0.010)	0.059
Ease of skin tanning	0.01	0.001	(0.009;0.011)	0.018
Time outdoors in winter (hours/day)	0.01	0.001	(0.008;0.011)	0.024
Time outdoors in summer (hours/day)	0.01	0.001	(0.008;0.011)	0.035
Season	0.009	0.001	(0.008;0.011)	0.121
Model 1 + lifestyle variables				
Smoke	0.01	0.001	(0.008;0.011)	0.017
Alcohol status	0.009	0.001	(0.008;0.011)	0.022
Alcohol frequency	0.009	0.001	(0.008;0.011)	0.032
Smoke+Alcohol	0.03	0.002	(0.026;0.033)	0.024

Linear regressions were performed to assess associations between log-transformed serum 25(OH)D levels and log2-transformed energy-adjusted dietary VD intake to account for the impact of doubling intakes of VD on VD status, considering the influence of other variables on this association. In Model 1, the association between dietary VD (the main predictor variable) and serum 25(OH)D levels, adjusted for age, sex and center, was assessed. Model 1 was further adjusted for anthropometric variables (BMI and/or waist circumference), for sun exposure variables (IPAQ, walking activity, Moderate activity and Vigorous activity (METs), skin type, ease of skin tanning, time out-doors in winter, time outdoors in summer and season) and lifestyle factors (smoking status, alcohol status and alcohol frequency). R2=*R*-squared.

With regard to the predictive models of serum 25(OH)D levels (linear model), the selected variables and beta coefficients are shown in Table 4. The model with the lowest AIC and RMSE included the variables described above for the VD deficiency model. The observed outcome values in the test data and the outcome values predicted by the model in the training set were similar (data not shown). The R2 value confirmed that the model could accurately predict the serum 25(OH)D levels.

3.4.2. Applying machine learning approaches in predictive modeling

Table 4 also shows the results of the estimates and beta coefficients of the predictive models derived by machine learning, for VD deficiency and VD status, respectively. With respect to the binary model, the ENET-based model yielded a higher accuracy (0.91) than did the other methods (LASSO and Ridge Regression: 0.90). The concordance of the predictions (kappa) in the training and test sets was also greater in the ENET model (data not shown). The selected variables of this model were: age, IPAQ (high), vigorous METS score, smoking status (current), alcohol consumption (current), skin type (brown and black) and ease of skin tanning, season, waist circumference, BMI, time spent outdoors in summer, energy-adjusted dietary VD, and deprivation index. Brown skin, winter and spring were among the most important variables (Fig. 1A). This model showed an even higher value of the AUC for the prediction of VD deficiency prediction (AUC=0.89).

According to the linear model, the RMSE values were lower for the ENET-based model (mean = 18.487), than fo the other methods (LASSO and Ridge Regression: mean=18.489). This model also exhibited the highest R2 values (data not shown). Thus, the ENET model also emerged as the most accurate predictor of serum 25(OH)D levels. The variables selected were: sex, age, IPAQ score, all kinds of physical activity in METs, smoking and alcohol status (current), skin type and ease of skin tanning, season, waist circumference, BMI (obesity), time spent outdoors in winter and summer, energy-adjusted dietary VD, and deprivation index. The variables skin type (brown and black) and season (spring and winter) had the highest scores in the model (Figure 1B). The predicted mean of serum 25(OH)D in the training and test set was close (52.2 vs 52.7 nmol/L, respectively), suggesting that the modelś performance was good.

4. Discussion

This study examined the determinants of VD deficiency and VD status, to shed light on variables associated with low levels of this vitamin and to identify populations at high-risk of VD deficiency, a condition that has been associated with unfavorable health outcomes. The results of this study are thus highly valuable for screening for VD deficiency in the population. Indeed, our study showed that overweight and obese individuals, those who are less physically active and elderly subjects, are more prone to low serum 25(OH)D levels. Dietary VD is not the main determinant of VD status, whereas variables related to sunlight exposure strongly contribute to high VD levels in this North-European population. However, our study also showed, that both dietary and serum 25(OH)D are positively related with each other, regardless of age, sex, center, and even BMI or other variables. This study also aimed to develop a predictive model for VD status and VD deficiency from 21 distinct variables sourced from the UKB study. Various approaches including multiple linear regression, logistic regression, and ma-

Table 4
Beta coefficients of variables included in the prediction model obtained by classical models and machine learning models (ENET regression) for VD deficiency and VD blood levels.

	Classical regression models		Machine learning models	
	Binary model Estimate (VD deficiency)	*Linear model* β coefficients (VD blood)	*Binary model* Estimate (VD deficiency)	*Linear model* β coefficients (VD blood)
Intercept	−3.73	36.24	−2.20	66.2
Sex (Women)	0.134	−0.650	.	−21.8
Age	−0.0162		−0.0121	0.0707
IPAQ (Moderate)	−0.161	0.279	.	1.32
IPAQ (High)	−0.371	1.08	−0.171	3.72
METs (moderate)		−0.000109	.	−0.000430
METs (vigorous)	−0.000141	0.000215	−0.00000562	0.000950
METs (walikng)		−0.000159	.	−0.000138
Smoke (Previous)	.	.	.	0.285
Smoke (Current)	0.567	−0.763	0.254	−3.47
Alcohol freq (3/4 per week)	−0.153	.		
Alcohol freq (1/2 per week)	.	.		
Alcohol freq (1 to 3 per month)	.	−0.474		
Alcohol freq (Special occasion)	0.246	−0.420		
Alcohol freq (Never)	0.398	NA	.	2.62
Alcohol status (Previous)		1.13		
Alcohol_status (Current)		1.18	−0.249	3.89
Skin Fair	−0.1949	.	.	2.02
Skin light olive	.	.	.	1.67
Skin dark olive	0.297	.	.	−0.0691
Skin brown	0.888	−1.78	1.57	−14.8
Skin black	.	.	0.668	−13.4
Time outdoors winter	0.0400	−0.101	.	−0.333
Time outdoors summer	−0.146	0.316	−0.101	1.41
Ease skin (Moderately tanned)	0.176	−0.498	.	−1.90
Ease skin (Midly tanned)	0.403	−1.05	0.0153	−4.63
Ease skin (Never tanned)	0.5232	−1.23	0.125	−5.25
BMI		.	0.0135	−0.0822
Waist circumference	0.0186	−0.0531	0.00188	−0.219
BMI (overweight)	−0.169	.	.	.
BMI (obesity)	.		0.116	−1.64
Abdominal obese (yes)			0.163	−0.0125
Age group (50–60 y)	.	.	.	0.0000525
Age group (>60 y)	.	0.447	−0.0659	1.46
Season (spring)	1.22	−1.84	0.968	−10.5
Season (summer)	−0.618	0.715	−0.467	3.74
Season (winter)	1.36	−2.42	1.12	−12.5
VitD adjusted by E (R24h) μg/2,000 kcal	−0.0696	0.0505	−0.102	1.05
Multiple deprivation index	0.0139	−0.0130	0.0130	−0.121
Centre (52-53°)		0.372		
Centre (54-55°)		0.886		

Beta-coefficients are of different signs in both models because the binary model predicts VD deficiency, that is, levels less than 25 nmol/L, meanwhile, the linear model predicts VD levels. The OR of the binary model can be derived from the exponentiated beta coefficient. For a given predictor variable, the coefficient (β) can be interpreted as the average effect on y (VD deficiency or VD blood levels in this case) of a one-unit increase in the predictor, holding all other predictors fixed. The prediction models were built on a dataset with complete observations (N=48,913) and 21 variables, with 70% of the participants (N=34,239) used for the training set and the remaining 30% (N=14,457) for the test set.

chine learning techniques were applied. The classical and machine learning prediction models were robust and able to predict both VD outcomes.

Our study provides relevant findings on the relationship between sun exposure and 25(OH)D levels. The results showed that individuals with VD deficiency spent less time outdoors in both winter and summer, emphasizing the potential influence of seasonal sun exposure on serum 25(OH)D levels. The study also revealed a significant decrease in physical activity in those with VD deficiency compared to those in the nondeficient group. This finding suggests a potential link between VD deficiency and reduced engagement in moderate physical activity, as observed in prior studies in which VD deficiency was shown to be influenced by both sun exposure and physical activity [36,37]. In fact, as has been previously reported, engaging in physical activity, especially outdoors, can confer benefits to the population by reducing the prevalence of VD deficiency [37]. Additionally, a significant association has been observed regarding obesity, specifically with both vis-

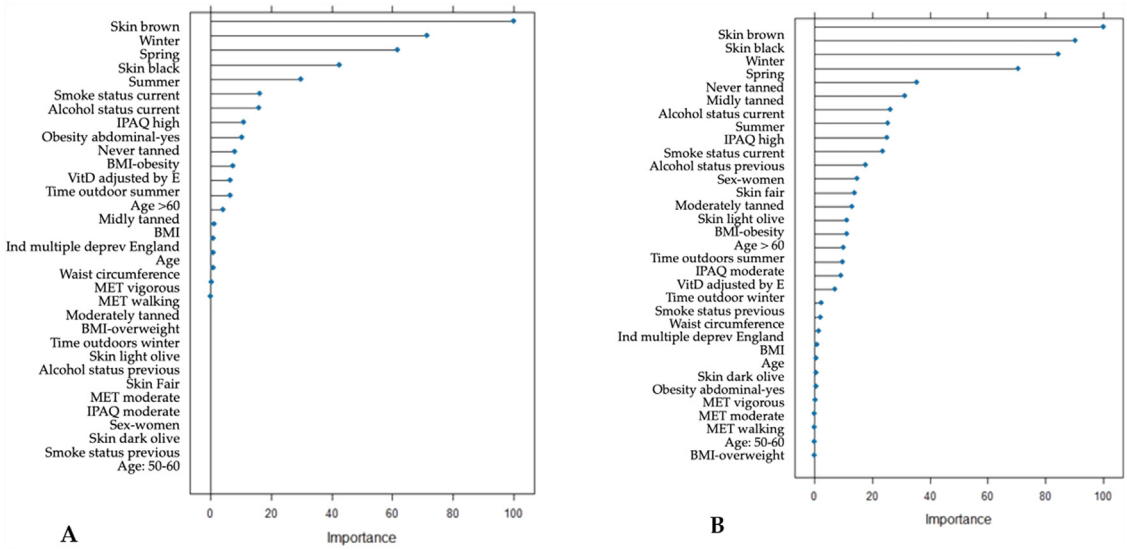

Fig. 1. Variable Importance plots retrieved from ENET for VD deficiency (alpha = 0.889, lambda = 0.00753) (A) and VD levels in blood (alpha=1, lambda=0.01) (B). The variable importance plot is scaled to 100 units. A 10-fold cross-validation procedure with 100 repetitions was carried out to select the variables and to fit the predictive model with appropriate alpha and lambda values on a training (observations=70%; *N*=34,239) and test sets (observations=30%; *N*=14,457). For the ENET model only those variables that are key in the predictive model were selected. The LASSO model provided similar results, though with higher RMSE values (data not shown). The Ridge Regression approach selected all the variables and obtained higher RMSE values. The variable importance defined by both models was similar.

ceral and abdominal obesity and VD. Previous observational studies have reported a negative association between serum 25(OH)D levels and the presence of obesity or overweight [38–41]. One potential explanation for this result is the lipophilic nature of VD, which leads to its accumulation in adipose tissue and lowers the serum concentration of the vitamin. However, other studies, mostly those that were cross-sectional and had limited sample sizes, have not shown any association between 25(OH)D levels and obesity in adults [42] or adolescents [43]. Hence, additional studies addressing this topic are warranted to provide knowledge for monitoring serum 25(OH)D levels in overweight or obese individuals. Another explanation could be the fact that individuals with obesity tend to be less active, resulting in reduced sun exposure and, consequently, lower serum levels of 25(OH)D. This inactivity increases susceptibility to VD deficiency [44,45]. As mentioned above, engaging in physical activity, especially outdoors, may boost VD levels in these individuals.

Regarding skin type, tanning and season, our study suggests a link between VD deficiency and factors such as fair skin prevalence and seasonal variation. For instance, the results support that darker skin tones are associated with lower 25(OH)D levels. In fact, previous studies have proposed that heightened skin pigmentation might contribute to the ineffective synthesis of VD in the skin. Melanin pigment, the chief determinant of skin color can hinder the photoproduction of VD by acting as a natural filter for solar UVB radiation within the epidermis [46,47]. Similar findings have been reported in other studies [48], indicating that sun tanning substantially contributes to serum 25(OH)D concentrations during the summer months, irrespective of race or ethnicity. Thus, these findings underscore the impact of skin tanning on 25(OH)D levels and seasonal variation.

In relation to the connection between serum 25(OH)D levels and dietary VD intake, we observed a noteworthy, albeit modest correlation between the two measures of VD. This correlation was influenced by anthropometric variables, physical activity, and the season of the year, although it was statistically significant. Diet is recognized as a key factor influencing 25(OH)D levels, especially in individuals with limited outdoor activity. It has been consistently shown that dietary intake of VD affects serum levels of 25(OH)D [49,50], and that VD supplementation contributes to increasing levels of 25(OH)D [51,52]. While diet is acknowledged as an important factor affecting VD levels, some studies suggest the need for a more comprehensive understanding of the mechanisms underlying the interaction between dietary VD and macronutrients [45]. For instance, it has been reported that the dietary composition of a meal can alter the bioavailability of VD, and postprandial events seem to influence VD [53]. These factors might account for the observed weak association in our study. Main food sources of VD are blue fish and oil, egg yolk, mushrooms, amongst others. A previous study conducted within the UKB cohort that evaluated the association of foods rich in VD with 25(OH)D levels in serum, reported that regular consumption of fatty fish was associated with lower odds of VD deficiency, this food being the main dieary determinant of VD status [9]. It is also important to consider that fortified foods are significant sources of VD in the diet. Although fortification was not considered in this study, it would be interesting to include it in future research, as the general population, especially in the UK, consumes fortified foods such as fat spreads and breakfast cereals to achieve the recommended levels of VD [54]. Importantly, a relatively high proportion of the study population had risk of dietary VD inadequacy. However, not all exhibited VD deficient levels in serum, emphasizing that VD status is influenced by variables other than dietary VD.

Finally, it is worth noting that alcohol and tobacco consumption influence serum 25(OH)D levels and, consequently, on the presence or absence of VD deficiency. In our study, individuals with VD deficiency exhibited a slightly higher prevalence of smoking and a lower proportion of current alcohol drinkers than did those without deficiency. The multivariate regression analyses also demonstrated the impact of tobacco and alcohol consumption patterns on the likelihood of VD deficiency. Other previous studies have reported similar findings with regard to alcohol [25]. In particular, individuals who did not consume alcohol in the past 12 months were more likely to have VD deficiency [22]. It has been suggested that this association might be linked to the fact that alcohol consumption is often associated with increased social interaction,

leading to greater sunlight exposure. As a result, individuals who consume alcohol may be less prone to having insufficient levels of serum 25(OH)D [22]. Regarding tobacco, it has been seen in other studies that circulating VD levels tend to be even lower in smokers [55,56], likely due to tobacco's role in disrupting the VD endocrine system. Smoking has been shown to interfere with the activation and metabolism of VD in the liver and kidneys, key organs responsible for converting VD into its active form. Additionally, smoking can lead to a reduced intake of VD through dietary changes, as it can alter taste perception and appetite, causing smokers to consume less VD-rich foods. The toxic compounds in tobacco may also increase inflammation and oxidative stress, further depleting VD levels by increasing the body's demands for this nutrient [56,57].

In summary, results of our study suggest significant correlations between lifestyle factors and both serum levels and dietary VD intake.These variables were considered for the predictive modelling using classical and machine learning approaches. In both models, variables such as skin type (especially brown and black), season, current smoking status and alcohol consumption, IPAQ, obesity status (visceral and abdominal), skin tanning, and time spent outdoors were consistently chosen. Previous studies have also constructed predictive models related to 25(OH)D serum levels [2,24] and VD deficiency [25,26,58,59], but none of them used a combination of classical and machine learning approaches or considered as many variables as our study, nor did they encompass predictive models for both serum 25(OH)D levels and VD deficiency. For instance, in a Spanish study on predictive models of serum VD levels involving 200 participants, it was reported that sex, age, physical activity and BMI, as well as other sunlight exposure variables, composed a model able to predict the VD status and discriminate between VD deficiency and non-deficiency with an AUC of 0.8 [2]. Another study used using the classical approach and incorporating similar variables obtained low R2 values, but was able validate the model in independent samples [24]. In another study that used machine learning-based predictive models among 5,106 participants, as in our study, the model baed on the ENET regression reached the highest AUC [58]. In the study by Sohl et al. [59], where the classical approach was used, two levels at 50 and 30 nmol/L were considered. This study achieved similar AUC values for both VD cutpoints, and yielded an AUC similar to ours. These AUC values have also been reported by others [26]. Finally, the study by Narang et al. [25] also used a classical logistic regression predictive model and obtained slightly lower AUC values (approximately 0.7). However, they emphasized the significant role of non-European ethnicity as a crucial risk factor for VD deficiency. In our study, ethnicity did not appear to be a decisive determinant of VD deficiency. Nonetheless, skin type and tanning ease were considered to account for melanogenic differences and ethnic skin types.

Compared to these studies, our study included more variables and a larger population, leading to superior AUC, RMSE, and R2 results. Finally, although most of these studies were not externally validated, our model incorporates similar variables and estimates.

4.1. Strengths and limitations

Our study is based in a large sample study population, for which comprehensive sociodemographic and lifestyle data are available. This enabled us to explore indepth the determinants related to VD deficiency. Another strength of our study is the consideration of dietary VD derived from 24-HR. Various predictive models and subsequent comparative analyses were used, to illustrate the effectiveness of both classical approaches and machine learning models for VD prediction. Although 25(OH)D blood sampling remains the gold standard for assessing VD deficiency in the population, this model could be useful for identifying people with this condition.

There are also several limitations to consider. The first one is that the blood sample and the 24-HR were not done on the same day. However, the dietary intake of VD relied on several 24-HR. A limitation of this study is the indirect and subjective nature of estimating daily VD intake from food, which relies on self-reported dietary data and may not fully capture the actual consumption or variations in fortified food products. Another limitation to consider is the low representativeness of the UKB study population with regard to the general population [7]. However, the study is valid for assessing exposure-outcome relationships. In line with this limitation, since our study included adults, the results may have limited applicability to specific groups such as children, characterized by distinct metabolic and physiological processes. Additionally, our study lacks external validation for predictive models, making it more suitable for epidemiological and research contexts than for clinical settings, unless externally validated in other studies. Nevertheless, predictive models using machine learning approaches and cross-validation procedures, as applied in this study, reduce overfitting issues and improve model performance [14,15]. In addition, we used a single 25(OH)D serum measurement, which does not necessarily resemble VD status in the long term. It is also important to note the potential for bias in the self-reported questionnaire answers, which could lead to participant misclassification due to underreporting or overreporting of sunlight exposure and lifestyle variables. Furthermore, sun protection measures, which could affect the synthesis of VD in the skin, were not considered. This variable, however, did not emerge as relevant in our study. In addition we did not take into consideration VD supplementation, but few participants ($n = 425$) reported the use of multivitamin supplements. It is important to highlight that VD food fortification was accounted for in the nutritional assessment.

5. Conclusions

Several variables related to sunlight exposure and lifestyle variables, including dietary VD intake, and anthropometrics indices among others, are determinants of serum 25(OH)D levels and VD deficiency. While high dietary VD contributes to increased VD levels, there are various factors, mostly referred to sun exposure, that impact VD levels to a greater extent. The predictive model based on machine learning approaches and comprising the main determinants of VD status could be valuable for assessing VD deficiency, aiding in the assessment of the risk of diseases associated with this condition.

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Ángela Alcalá-Santiago: Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Belén García-Villanova:** Writing – review & editing, Supervision. **María Dolores Ruíz-López:** Writing – review &

editing. **Ángel Gil:** Writing – review & editing. **Miguel Rodriguez-Barranco:** Writing – review & editing. **Maria José Sánchez:** Writing – review & editing. **Esther Molina-Montes:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Acknowledgments

This paper and the results presented constitute part of Ángela Alcalá Santiago's Doctoral Thesis performed in the Nutrition and Food Science Doctorate Program of the University of Granada. She was supported by a Research Fellowship from the Government of Spain (FPU22/02915). Funding for open access charge: Universidad de Granada / CBUA.

Funding

This research was funded by Project PECOVID-0200-2020, funded by Consejería de Salud y Consumo de la Junta de Andalucía and cofunded by the European Regional Development Fund (ERDF-FEDER).

Data availability

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ukbiobank.ac.uk/enable-your-research/about-our-data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jnutbio.2025.109919.

References

[1] Alcalá-Santiago Á, Rodríguez-Barranco M, Rava M, Jiménez-Sousa MÁ, Gil Á, Sánchez MJ, et al. Vitamin D deficiency and COVID-19: a biological database study on pathways and gene-disease associations. Int J Mol Sci 2022;23:14256. doi:10.3390/ijms232214256.
[2] Valer-Martinez A, Sayon-Orea C, Martínez Hernandez JA, De la Fuente-Arrillaga C, Pérez de Rojas J, Barcones F, et al. Forecasting levels of serum 25-hydroxyvitamin D based on dietary intake, lifestyle and personal determinants in a sample of Southern Europeans. Br J Nutr 2023;130:1814–22. doi:10.1017/S0007114523000946.
[3] Szodoray P, Nakken B, Gaal J, Jonsson R, Szegedi A, Zold E, et al. The complex role of vitamin D in autoimmune diseases. Scand J Immunol 2008;68:261–9. doi:10.1111/j.1365-3083.2008.02127.x.
[4] Dipasquale V, Lo Presti G, Milani GP, Corsello A, Agostoni C, Romano C. Vitamin D in prevention of autoimmune diseases. Front Biosci-Landmark 2022;27:288. doi:10.31083/j.fbl2710288.
[5] Ao T, Kikuta J, Ishii M. The effects of vitamin D on immune system and inflammatory diseases. Biomolecules 2021;11:1624. doi:10.3390/biom11111624.
[6] Tanna NK, Karki M, Webber I, Alaa A, El-Costa A, Blair M. Knowledge, attitudes, and practices associated with vitamin D supplementation: a cross-sectional online community survey of adults in the UK. PLoS One 2023;18:e0281172. doi:10.1371/journal.pone.0281172.
[7] Amrein K, Scherkl M, Hoffmann M, Neuwersch-Sommeregger S, Köstenberger M, Tmava Berisha A, et al. Vitamin D deficiency 2.0: an update on the current status worldwide. Eur J Clin Nutr 2020;74:1498–513. doi:10.1038/s41430-020-0558-y.
[8] Holick MF, Chen TC. Vitamin D deficiency: a worldwide problem with health consequences. Am J Clin Nutr 2008;87 1080S–1086S. doi:10.1093/ajcn/87.4.1080S.
[9] Sutherland JP, Zhou A, Leach MJ, Hyppönen E. Differences and determinants of vitamin D deficiency among UK biobank participants: a cross-ethnic and socioeconomic study. Clin Nutr 2021;40:3436–47. doi:10.1016/j.clnu.2020.11.019.
[10] EFSA Panel on Dietetic Products N and A (NDA). Dietary reference values for vitamin D. EFSA Journal 2016;14:e04547. doi:10.2903/j.efsa.2016.4547.
[11] Newberry SJ, Chung M, Shekelle PG, Suttorp Booth M, Liu JL, Ruelaz Maher A, et al. Vitamin D and calcium: a systematic review of health outcomes (Update). Evid Rep Technol Assess (Full Rep) 2014;217:1–929. doi:10.23970/AHRQEPCERTA217.
[12] Zhang R, Zhang Y, Liu Z, Pei Y, Xu P, Chong W, et al. Association between Vitamin D supplementation and cancer mortality: a systematic review and meta-analysis. Cancers (Basel) 2022;14:3717. doi:10.3390/cancers14153717.
[13] Pei Y-Y, Zhang Y, Peng X-C, Liu Z-R, Xu P, Fang F. Association of Vitamin D supplementation with cardiovascular events: a systematic review and meta-analysis. Nutrients 2022;14:3158. doi:10.3390/nu14153158.
[14] Dewansingh P, Reckman GAR, Mijlius CF, Krijnen WP, van der Schans CP, Jager-Wittenaar H, et al. Protein, calcium, vitamin D intake and 25(OH)D status in normal weight, overweight, and obese older adults: a systematic review and meta-analysis. Front Nutr 2021;8. doi:10.3389/fnut.2021.718658.
[15] Shahangian S, Alspach TD, Astles JR, Yesupriya A, Dettwyler WK. Trends in laboratory test volumes for Medicare Part B reimbursements, 2000–2010. Arch Pathol Lab Med 2014;138:189–203. doi:10.5858/arpa.2013-0149-OA.
[16] O'Neill M, Burke K. Variable selection using a smooth information criterion for distributional regression models. Stat Comput 2023;33:71. doi:10.1007/s11222-023-10204-8.
[17] Kosciolek T, Jones DT. Accurate contact predictions using covariation techniques and machine learning. Proteins 2016;84:145–51. doi:10.1002/prot.24863.
[18] Sibieude E, Khandelwal A, Hesthaven JS, Girard P, Terranova N. Fast screening of covariates in population models empowered by machine learning. J Pharmacokinet Pharmacodyn 2021;48:597–609. doi:10.1007/s10928-021-09757-w.
[19] Zhao Q, Small DS, Ertefaie A. Selective inference for effect modification via the lasso. J R Stat Soc Series B Stat Methodol 2022;84:382–413. doi:10.1111/rssb.12483.
[20] Teppola P, Taavitsainen V-M. Parsimonious and robust multivariate calibration with rational function least Absolute shrinkage and selection Operator and rational function elastic net. Anal Chim Acta 2013;768:57–68. doi:10.1016/j.aca.2013.01.005.
[21] Manios Y, Moschonis G, Lambrinou CP, Mavrogianni C, Tsirigoti L, Hoeller U, et al. Associations of vitamin D status with dietary intakes and physical activity levels among adults from seven European countries: the Food4Me study. Eur J Nutr 2018;57:1357–68. doi:10.1007/s00394-017-1415-1.
[22] Lips P, van Schoor NM, de Jongh RT. Diet, sun, and lifestyle as determinants of vitamin D status. Ann N Y Acad Sci 2014;1317:92–8. doi:10.1111/nyas.12443.
[23] UK Biobank. Background n.d. https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us.
[24] Bertrand KA, Giovannucci E, Liu Y, Malspeis S, Eliassen AH, Wu K, et al. Determinants of plasma 25-hydroxyvitamin D and development of prediction models in three US cohorts. Br J Nutr 2012;108:1889–96. doi:10.1017/S0007114511007409.
[25] Narang RK, Gamble GG, Khaw K-T, Camargo CA, Sluyter JD, Scragg RKR, et al. A prediction tool for vitamin D deficiency in New Zealand adults. Arch Osteoporos 2020;15:172. doi:10.1007/s11657-020-00844-y.
[26] Tran B, Armstrong BK, McGeechan K, Ebeling PR, English DR, Kimlin MG, et al. Predicting vitamin D deficiency in older Australian adults. Clin Endocrinol (Oxf) 2013;79:631–40. doi:10.1111/cen.12203.
[27] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779. doi:10.1371/journal.pmed.1001779.
[28] Liu W, Zhuang Z, Wang W, Huang T, Liu Z. An improved genome-wide polygenic score model for predicting the risk of type 2 diabetes. Front Genet 2021;12. doi:10.3389/fgene.2021.632385.
[29] Lin L-Y, Smeeth L, Langan S, Warren-Gash C. Distribution of vitamin D status in the UK: a cross-sectional analysis of UK Biobank. BMJ Open 2021;11:e038503. doi:10.1136/bmjopen-2020-038503.
[30] UK Biobank. 24-hour dietary recall questionnaire. UK Biobank; 2012. Stockport, UK. http://www.ukbiobank.ac.uk/ accessed: October-December, 2023.
[31] UK Biobank Index of multiple deprivation scores. Stockport, UK: UK Biobank; 2018.
[32] Nuttall FQ. Body mass index. Nutr Today 2015;50:117–28. doi:10.1097/NT.0000000000000092.
[33] Kis OS, Buch A, Eldor R, Moran DS. Hand grip strength relative to waist circumference as a means to identify men and women possessing intact mobility in a cohort of older adults with type 2 diabetes. Biomedicines 2023;11:352. doi:10.3390/biomedicines11020352.
[34] Cleland C, Ferguson S, Ellis G, Hunter RF. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. BMC Med Res Methodol 2018;18:176. doi:10.1186/s12874-018-0642-3.
[35] R Core Team R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
[36] Al-Othman A, Al-Musharaf S, Al-Daghri NM, Krishnaswamy S, Yusuf DS, Alkharfy KM, et al. Effect of physical activity and sun exposure on vitamin D status of Saudi children and adolescents. BMC Pediatr 2012;12:92. doi:10.1186/1471-2431-12-92.
[37] Araujo-Moura K, Nascimento-Ferreira MV, Schaan B, Bloch K, de Carvalho K, Cureau F, et al. Serum vitamin D levels mediate the association between physical activity and blood pressure in adolescents. J Phys Act Health 2024:1–8. doi:10.1123/jpah.2022-0532.
[38] Lee P, Greenfield JR, Seibel MJ, Eisman JA, Center JR. Adequacy of vitamin D replacement in severe deficiency is dependent on body mass index. Am J Med 2009;122:1056–60. doi:10.1016/j.amjmed.2009.06.008.

[39] Bischof MG, Heinze G, Vierhapper H. Vitamin D status and its relation to age and body mass index. Horm Res Paediatr 2006;66:211–15. doi:10.1159/000094932.
[40] Abiri B, Valizadeh M, Ramezani Ahmadi A, Amini S, Nikoohemmat M, Abbaspour F, et al. Association of vitamin D levels with anthropometric and adiposity indicators across all age groups: a systematic review of epidemiologic studies. Endocr Connect 2023. doi:10.1530/EC-23-0394.
[41] Lu X, Sun Z. Association between vitamin levels and obesity in the national health and nutrition examination surveys 2017 to 2018. J Dev Orig Health Dis 2024:1–8. doi:10.1017/S2040174423000466.
[42] Alharbi AF, Ali M, Alrefaey YI, Filimban HA, Alhwaity AS, Alamoudi A, et al. Association between serum vitamin D levels and body mass index status: a cross-sectional study at King Khalid Hospital, Jeddah, Saudi Arabia, from 2019 to 2020. Cureus 2023. doi:10.7759/cureus.46927.
[43] Cárdenas V, Serrano C, Amezquita MV. Déficit de vitamina D en adolescentes: ¿existe diferencia según el grado de obesidad? Andes Pediatr 2023;94:339. doi:10.32641/andespediatr.v94i3.4395.
[44] Carter SJ, Baranauskas MN, Fly AD. Considerations for obesity, vitamin D, and physical activity amid the COVID-19 pandemic. Obesity 2020;28:1176–7. doi:10.1002/oby.22838.
[45] Jakicic JM, Davis KK. Obesity and physical activity. Psychiatr Clin North Am 2011;34:829–40. doi:10.1016/j.psc.2011.08.009.
[46] Hanel A, Carlberg C. Skin colour and vitamin D: an update. Exp Dermatol 2020;29:864–75. doi:10.1111/exd.14142.
[47] Khalid AT, Moore CG, Hall C, Olabopo F, Rozario NL, Holick MF, et al. Utility of sun-reactive skin typing and melanin index for discerning vitamin D deficiency. Pediatr Res 2017;82:444–51. doi:10.1038/pr.2017.114.
[48] Sawicki CM, Van Rompay MI, Au LE, Gordon CM, Sacheck JM. Sun-exposed skin color is associated with changes in serum 25-hydroxyvitamin D in racially/ethnically diverse children. J Nutr 2016;146:751–7. doi:10.3945/jn.115.222505.
[49] Spiro A, Buttriss JL. Vitamin D: an overview of vitamin D status and intake in Europe. Nutr Bull 2014;39:322–50. doi:10.1111/nbu.12108.
[50] Calvo MS, Whiting SJ, Barton CN. Vitamin D intake: a global perspective of current status. J Nutr 2005;135:310–16. doi:10.1093/jn/135.2.310.
[51] Holt R, Jorsal MJ, Yahyavi SK, Qin S, Juul A, Jørgensen N, et al. High-dose cholecalciferol supplementation to obese infertile men is sufficient to reach adequate vitamin D status. British Journal of Nutrition 2024;131:642–7. doi:10.1017/S0007114523002222.
[52] Saeidlou SN, Vahabzadeh D, Karimi F, Babaei F. Determining the vitamin D supplementation duration to reach an adequate or optimal vitamin D status and its effect on blood lipid profiles: a longitudinal study. J Health Popul Nutr 2024;43:81. doi:10.1186/s41043-024-00576-6.
[53] Sharifan P, Amiri Z, Moafian F, Mohamadi-Bajgiran M, Saffar Soflaei S, Darroudi S, et al. Investigating the association between dietary nutrient intake and vitamin D serum levels. Nutrire 2023;48:34. doi:10.1186/s41110-023-00218-7.
[54] Office for Health Improvement & Disparities GOVU Fortifying foods and drinks with vitamin D: summary. Office for Health Improvement & Disparities GOVU; 2024 https://www.gov.uk/government/publications/fortifying-food-and-drink-with-vitamin-D-a-sacn-rapid-review/fortifying-foods-and-drinks-with-vitamin-D-summary (accessed October 20, 2024).
[55] Brot C, Rye Jørgensen N, Helmer Sørensen O. The influence of smoking on vitamin D status and calcium metabolism. Eur J Clin Nutr 1999;53:920–6. doi:10.1038/sj.ejcn.1600870.
[56] Lin S, Li J, Zhang Y, Song X, Chen G, Pei L. Maternal passive smoking, vitamin D deficiency and risk of spontaneous abortion. Nutrients 2022;14:3674. doi:10.3390/nu14183674.
[57] Mousavi SE, Amini H, Heydarpour P, Amini Chermahini F, Godderis L. Air pollution, environmental chemicals, and smoking may trigger vitamin D deficiency: evidence and potential mechanisms. Environ Int 2019;122:67–90. doi:10.1016/j.envint.2018.11.052.
[58] Sluyter JD, Raita Y, Hasegawa K, Reid IR, Scragg R, Camargo CA. Prediction of vitamin D deficiency in older adults: the role of machine learning models. J Clin Endocrinol Metab 2022;107:2737–47. doi:10.1210/clinem/dgac432.
[59] Sohl E, Heymans MW, de Jongh RT, den Heijer M, Visser M, Merlijn T, et al. Prediction of vitamin D deficiency by simple patient characteristics. Am J Clin Nutr 2014;99:1089–95. doi:10.3945/ajcn.113.076430.