Scientific Foundation SPIROSKI, Skopje, Republic of Macedonia
Open Access Macedonian Journal of Medical Sciences. 2023 Jan 02; 11(A):70-75.
https://doi.org/10.3889/oamjms.2023.11340
eISSN: 1857-9655
Category: *A - Basic Sciences*
Section: *Biochemistry*

Since 2002

OPEN ACCESS

Determination of Factors Affecting Vitamin D Levels in Women at Risk Using Classification and Regression Tree Analysis

Antoaneta Yordanova[1]*, Anna Tolekova[2], Pavlina Teneva[2], Ivelina Dobreva[2], Katya Kichukova[2]

[1]*Department of Social Medicine, Healthcare Management and Disaster Medicine, Trakia University, Stara Zagora, Bulgaria;* [2]*Department of Health Care, Medical College, Trakia University, Stara Zagora, Bulgaria*

Edited by: Slavica Hristomanova-Mitkovska
Citation: Yordanova A, Tolekova A, Teneva P, Dobreva I, Kichukova K. Determination of Factors Affecting Vitamin D Levels in Women at Risk Using Classification and Regression Tree Analysis. Open Access Maced J Med Sci. 2023 Jan 02; 11(A):70-75. https://doi.org/10.3889/oamjms.2023.11340
Keywords: Vitamin D; Premenopausal and menopausal women; Indoor working women health; Classification and regression trees; Data mining
***Correspondence:** Antoaneta Yordanova, Department of Social Medicine, Healthcare Management and Disaster Medicine, Trakia University, 11 Ameyska Str, Stara Zagora 6000, Bulgaria.
E-mail: antoaneta.yordanova@trakia-uni.bg
Received: 01-Dec-2022
Revised: 18-Dec-2022
Accepted: 19-Dec-2022

Funding: The research and publication of this article were funded by Trakia University, Stara Zagora, Bulgaria (NIP-MK 2021. The first author expresses gratitude for the project grant MU21-FMI-015 of NPD at University of Plovdiv Paisii Hilendarski, financed by the Bulgarian Ministry of Education and Science.
Competing Interests: The authors have declared that no competing interests exist

Abstract

BACKGROUND: The influence of the factors on Vitamin D as a health indicator in premenopausal and menopausal women is a significant subject to be investigated.

AIM: The study uses the potential of classification and regression trees (CART) as a data mining method for medical type samples.

METHODS: The data set is built by records of 84 indoor working women at the age of 45 to 67 years from five Bulgarian companies. The data are obtained through laboratory tests of serum concentrations of 25-OH-Vitamin D and a questionnaire, created for the study. Statistical data processing is made by descriptive statistics and the CART method.

RESULTS: The results show Vitamin D deficiency in 69% of the studied women at risk from Stara Zagora. For the target variable – Vitamin D (the quantity of 25-OH-Vitamin D), a regression CART tree was built. The calculated percentages of normalized importance for each independent variable reveal that the most important predictors, affecting Vitamin D, are body mass index (100%), alcohol (84.2%), education (70.3%), coffee (70.2%), Ca_Vit D (69.8%), and sports frequency (42.4%), while the other variables have much less importance.

CONCLUSION: The application of the CART method makes it possible to study the distribution and importance of the factors influencing the state of vitamin D. The presence of such a high percentage of women at risk requires a comprehensive approach, including educational programs and strict application of guidelines for vitamin D supplementation to prevent the effects of hypovitaminosis.

Introduction

Many factors affect the level of 25-OH-vitamin D, which is needed for normal bone density building. The study of Amrein *et al.* [1] shows that there is a worldwide deficiency of 25-OH-vitamin D. Some authors have investigated the relationship between its levels and human/women health status and have confirmed that the insufficient levels of 25-OH-vitamin D are relevant to many diseases such as osteoporosis, cognitive decline, depression, cardiovascular disease, hypertension, type 2 diabetes, and cancer [2], [3], [4].

The systematic review of Sowah *et al.* [5] suggests that suboptimal vitamin D levels depend on occupation and it is a major determinant. This way indoor workers have vitamin D deficiency more often comparing to the other occupations.

The problem becomes especially important in the context of the working conditions, along with the increased urbanization in the past century. The examined women who work indoors are defined as at risk for developing osteoporosis due to hormonal changes during premenopause and menopause [6]. Eating habits, diet, working conditions, and physical activity have been studied as factors that might affect the amount of 25-OH-vitamin D (Vitamin D) and hence – bone health [7], [8], [9].

The present work aims to study the influence of 15 factors on the levels of the health indicator 25-OH-vitamin D in premenopausal and menopausal women, which includes:

- Building a model, using the powerful data mining technique of classification and regression tree (CART) for the dependent variable vitamin D and 15 factors, related to the workplace, physical activity, eating habits, and lifestyle
- Obtaining a classification of the independent factors by importance concerning their influence on the measured values of 25-OH-vitamin D. The introduction should be succinct, with no subheadings. Limited figures may be included only if they are truly introductory and contain no new results.

Materials and Methods

Data collection and structure, data processing

The data set is composed of records of 84 working women aged 45–67 years from 5 companies, divided into two groups (Firm) – Faculties and Factories.

The study was conducted in December 2019 in the city of Stara Zagora, Bulgaria.

The data are obtained through laboratory tests for serum vitamin D concentrations and a questionnaire designed for the study. The women, in the study, had an increased risk of osteoporosis due to their sedentary indoor work and age over 42. In factories, these are seamstresses and operators of different machines, whose work is characterized by the same type of activities performed manually in a sitting position. The group includes 56 women aged 45–67 years. The group of faculties includes 28 women who are from the teaching and administrative staff. Their work does not involve intense physical activity and is mostly sedentary.

The laboratory blood tests determined the 25-OH-vitamin D levels, which were measured by the Cobas e 411 devices, using the electrochemiluminescent immunoassay method. Measurements of the height and weight of the examined women were performed, which were divided into five age groups.

The questionnaire includes several groups of questions related to the physical activity both at work and outside working hours – physical activity at home, sports (frequency and intensity), how to get to and from work; eating problems – the type of diet, frequency of eating, and overeating. In the last group, the questions were related to the lifestyle in terms of the frequency of drinking coffee, alcohol, and smoking. A group of questions directly related to the vitamin D levels, such as exposure to sunlight and intake of Ca and vitamin D supplements, is included in the study.

Statistical models

The data are processed using descriptive statistics and the CART method, implemented in the SPSS package, version 25. CART is a non-parametric modeling algorithm proposed in Brejman *et al.* [10]. The statistical analysis is based on models with a dependent variable vitamin D (continuous type) and 15 independent variables (age, education, physical activity, frequency of exercise, diet, coffee, alcohol, smoking, exposure to sunlight, company, back distortion, overeating, activity, body mass index [BMI], and Ca_Vit D). They reflect the components of the eating habits and lifestyle.

When examining environmental and lifestyle factors may affect vitamin D, it is most often preferred to apply descriptive statistics, ANOVA [7], and logistic regression [8]. The CART method is one of the visually non-parametric methods that provide an easier interpretation of the results of the statistical analysis.

The application of this method to the regression tree does not require assumptions about the normal distribution of the variables as it works by forming homogeneous subgroups for the processed data. However, the CART model is affected by multicollinearity, deviations, and missing values [10].

In the past years, the use of the CART method has been popular in various fields, but no publications have been found using it to determine the factors, having an impact on Vitamin D. The present study proposes a model created by the CART method to determine the significant factors, influencing the Vitamin D levels in women at risk from the Stara Zagora region in Bulgaria.

Results

Descriptive statistics

In terms of age, the cases are divided into five groups: (45–49), (50–54), (55–59), (60–64), and (65–69), for which the mean values of the vitamin D levels are, respectively, 19.70 ng/mL; 20.29 ng/mL; 16.61 ng/mL; 16.70 ng/mL; and 17.07 ng/mL. The highest level is in the age group (50-54). In Figure 1, the levels are illustrated by a Boxplot diagram.

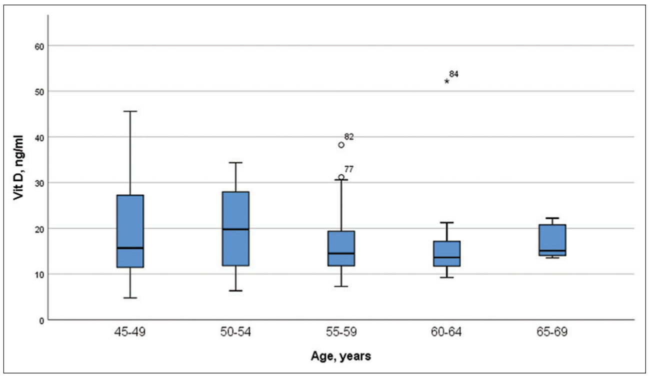

Figure 1: Boxplot diagram of vitamin D levels by age groups

Two extreme values were measured: one in the group (45–49) with a value of 45.60 ng/mL and one in the group (60–64) – 52.19 ng/mL. These two individuals were followed up and found to have a special vitamin D compensation regimen. In the first case, due to a vertebral fracture and proven osteoporosis, Prolia injection therapy is being performed twice a year, accompanied by calcium supplements, as well as vitamin D twice a week. The second case is a result of visits to the solarium once a week (6–8 min) throughout the year.

The following coding of 25-OH-vitamin values was applied: severe deficiency <10.0 ng/mL;

insufficiency ≥10 and <20 ng/mL, sufficiency ≥20 ng/mL, and optimal levels ≥40 and ≤ 80 ng/ml. The group with an insufficient level of vitamin D stands out as a percentage – 59.5%. In total, the first two groups cover a high percentage of cases – 69% of the women surveyed had insufficient vitamin D levels or hypovitaminosis. This result can be explained by some leading factors such as working conditions and lifestyle.

In the optimal group, we have only two cases, described above (osteoporosis therapy and regular visits to the solarium).

A Kruskal–Wallis H test showed that there was a statistically significant difference in vitamin D between the different education, $\chi^2(5) = 13.2$, $p = 0.022$, with a mean rank vitamin D of 24.67 for primary degree, 42.91 for secondary degree, 32.47 for professional degree, 31.83 for bachelor, 50.29 for master, and 74.50 for a higher degree.

Creating a model

The CART method was applied to create a model of the influence of 15 factors on the values of vitamin D for the studied group of women at risk.

The following specifications have been selected for model settings: minimum cases in parent node 5, minimum cases in child node 2, cross-validation – 10, and maximum tree depth 5. The resulting tree has 19 nodes, ten of which are terminal nodes.

In the initial stage, all 84 observations were located in the root node 0 of the tree with an average as a predicted value of 8.852 ng/mL for vitamin D. These cases are divided concerning the first variable firm into two groups: Factories and Faculties.

In the left node 1, 56 cases were classified with an average for vitamin D of 15.611 ng/mL. In the right node 2, 20 cases are classified, for which the expected (average) value is 22,953 ng/mL. For each dividing variable in the decision tree, the degree of importance with regard to improving its contribution to the model is calculated. The next separation procedure continues from node 1. A more important variable is Ca_VitD. The data in the two child nodes 3 and 4 are calculated and presented as in the first stage of the algorithm. By repeating this procedure, we obtain a decision tree, as shown in Figure 2.

The regression tree can be considered as a series of rules, defined by separation inequalities. A full description of the rules for the resulting terminal nodes of the tree is given in Table 1.

The decision tree can be used to classify and predict new cases by following the rules. For example, if a factories employee does not take Ca_VitD supplements, it can be expected that the Vitamin D value will be close to 14.5858 ng/mL (node 4). Moreover, if they take Ca_VitD from time to time and BMI ≤ 24.19, the predicted (average) value for vitamin D is 32.7367 ng/mL. The results in the other terminal nodes are interpreted similarly.

Interest evokes the grouping of the two cases with the highest values for the dependent variable in the terminal node 10. They are from faculties with a normal weight BMI ≤ 22.57 and consuming alcohol every day. Alcohol is a grouping factor in the last division.

In these two cases, the individuals apply therapies. When building a model without the alcohol factor, these two cases are grouped by another common feature – the back distortion. When specifying the factors included in the study with an emphasis on the therapies for proven osteoporosis, this will be the actual factor of grouping and division at this level in the CART model.

The constructed model has a coefficient of determination $R^2 = 0.714$. The model describes 71% of the changes in Vitamin D. The distribution of model errors is close to normal – the Shapiro–Wilk test = 0.089 > 0.05.

The description of the model includes 15 factors. The contribution of the first 7 is presented in Table 2. The most important feature is the BMI (considered 100%), and the rest are normalized to it. The second factor influencing vitamin D in the sample is alcohol with about 84.2%, followed by education – 70.3% and coffee drinking – 70.2%.

Discussion

A representative study of vitamin D levels conducted in 2012 for Bulgaria obtained overall

Table 1: Descriptions of the rules and predicted values for the ten terminal nodes of the classification and regression tree model

Terminal node	Number of cases	Rules	Mean/predicted value
Node 10	2	Firm = {Faculties}, BMI ≤ 22.57, alcohol = {Every day}	48.89
Node 7	3	Firm = {Factories}, Ca_Vitamin D = {From time to time}, BMI ≤ 24.19	32.74
Node 18	3	Firm = {Faculties}, BMI > 22.57. eating extra food = {At least once a day, no, never}, BMI ≤ 30.01, age > 52	32.45
Node 9	3	Firm = {Faculties}, BMI ≤ 22.57, alcoho1 = {1–3 times/week, rarely, I do not use}	31.33
Node 17	2	Firm = {Faculties}. BMI > 22.57, eating extra food = {At least once a day, no, never}, BMI ≤ 30.01, age ≤ 52	26.65
Node 14	3	Firm={Faculties}, BMI > 22.57, eating extra food = {At least once a day, no, never}, BMI > 30.01	19.45
Node 15	11	Firm = {Faculties}, BMI > 22.57, eating extra food = {Yes, 2–3 times a week}, age ≤ 60.5	17.83
Node 8	5	Firm = {Factories}, Ca_Vitamin D = {Time to time}, BMI > 24.19	15.17
Node 4	48	Firm = {Factories}, Ca_Vitamin D = {No, regularly}	14.59
Node 16	4	Firm = {Faculties}, BMI > 22.57, eating extra food = {Yes, 2–3 times a week}, age > 60.5	11.44

CART: Classification and regression tree, BMI: Body mass index.

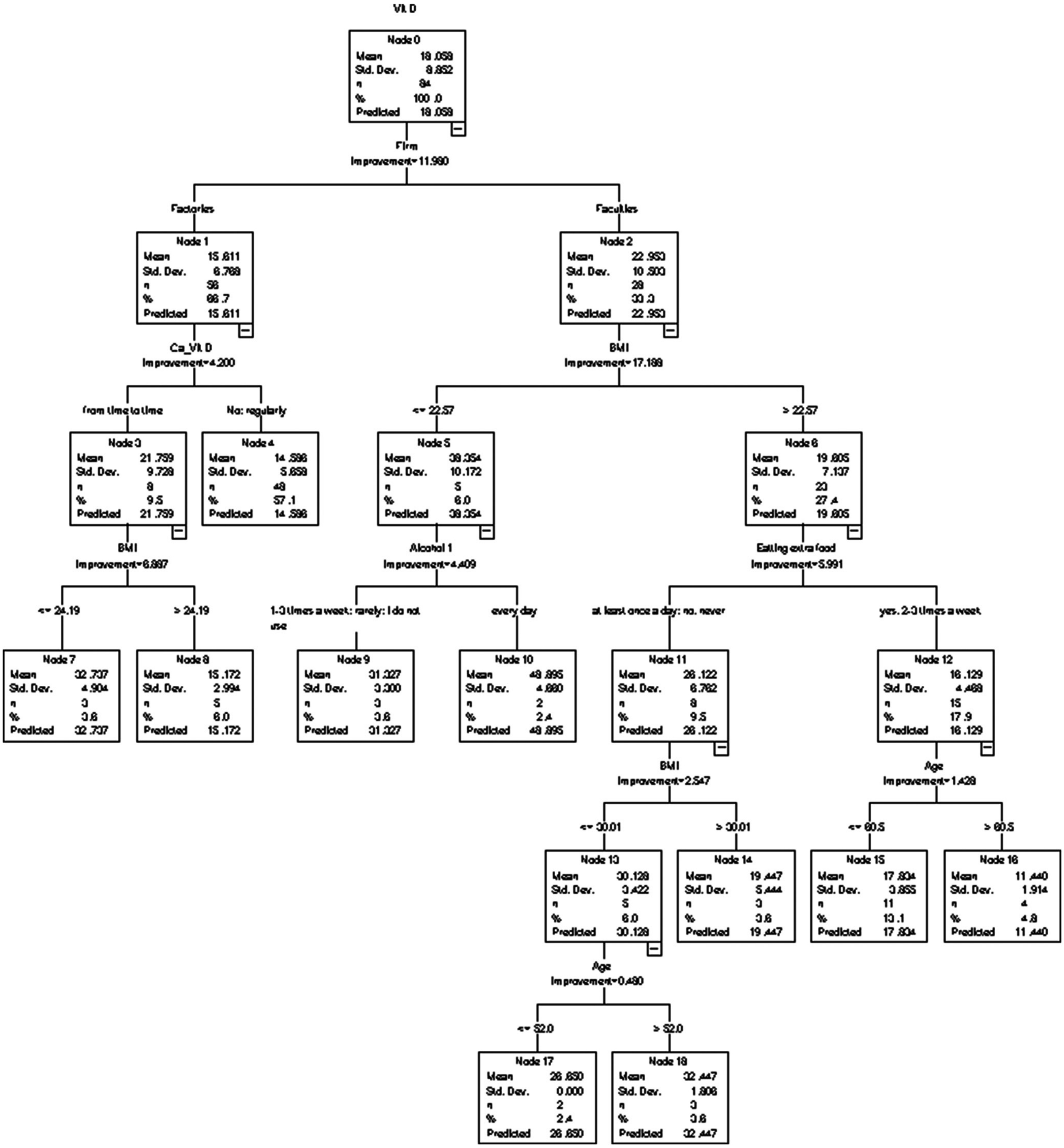

Figure 2: Classification and regression trees tree built for the target variable vitamin D and 15 predictors

Table 2: Independent variable importance

Independent variable	Importance	Normalized importance (%)
BMI	30.114	100.0
Alcohol	25.348	84.2
Education	21.171	70.3
Coffee	21.142	70.2
Ca_Vitamin D	21.030	69.8
Sports frequency	12.776	42.4
Firm	11.980	39.8

BMI: Body mass index.

averages, percentages, and applied a variance analysis to identify statistical differences between the sex and age groups for men and women [11]. Comparing the levels of the mean values and the medians of the two studies for the relative groups, the conclusion is that for the present study, the levels are within the established limits. The mean values obtained in the 25-OH-vitamin D target study were in the range (16–20) ng/mL, and the median was between (13 and 19) ng/mL, while, in the mentioned study, the mean values were about 37 nmol/L (14.8 ng/mL) and the median was 35 nmol/L (14 ng/mL).

Our studies confirm the role of increased BMI as a major factor, determining vitamin D homeostasis. The importance of overweight as a factor determining the homeostasis of vitamin D has been shown in many publications [4], [12], [13], and an inverse relationship between the BMI and the plasma vitamin D concentration has even been reported [14]. The increased amount of subcutaneous adipose tissue, which leads to increased BMI, leads to a disturbance of the vitamin D concentration. The adipose tissue absorbs vitamin D and serves as a reservoir, releasing vitamin D when needed.

The data for the effect of alcohol on vitamin D homeostasis are very contradictory. According to Tardelli *et al.* [15], it was found that, in 30.7%, there is a positive relationship between alcohol consumption and the vitamin D concentration, in 36.7%, it is negative, but other researchers do not find similar relationships. According to Ogunsakin *et al.* [16], chronic exposure to alcohol has the potential to reduce the levels of pulmonary vitamin D and results in subsequent downregulation of cathelicidin/LL-77.

In the present study, the two extreme cases (45.60 and 52.19 ng/mL) showed daily alcohol consumption. This does not make a clear positive relationship between alcohol consumption and vitamin D levels and only indicates that with normal body weight, the applied therapies (drugs for proven osteoporosis and weekly exposure to ultraviolet rays) compensate for the possible negative effects of alcohol.

Caffeine is a factor of 4th place importance in our study. The research of Rapuri *et al.* [17] results in conclusions that caffeine decreases vitamin D receptor protein expression and 1,25(OH)2D3 stimulated alkaline phosphatase activity in human osteoblast cells.

The relationship between physical activity parameters and vitamin D is a subject of some studies. The review of Koundourakis *et al.* [18] presents the links between vitamin D levels and the ability to physical activities and exercise [19]. Reports the significant relationship between 25(OH)D levels and physical exercise in elderly women for maintaining independence in daily activities.

The predictor education in the present study results as the 3rd factor on importance. It reflects the school/university degrees indifference of Lee *et al.* [20], where the education has a meaning of offering knowledge about the problem with vitamin D, the needs of 30 min daily sunlight exposure and taking supplements. The level of education is an element of socioeconomic status as a risk factor for the level of vitamin D and health in some national studies. Its statistically significant influence is proved [21], [22], [23], [24], [25].

The data extraction methods allow to classify of the influencing factors and to determine their degree of importance.

Conclusion

For a sample of 84 observations of women at risk from 5 companies from the city of Stara Zagora, Bulgaria, divided into two groups of occupation, a classification tree was constructed for the studied dependence of vitamin D on 15 factors, describing their lifestyle and nutrition. The constructed model has a determination coefficient of $R^2 = 0.714$.

For this particular sample, a regression model of vitamin D-related predictors was obtained. The BMI, alcohol consumption, people's education, coffee drinking, supplementation, sports activities, and the conditions are the most important predictors.

The results show the presence of hypovitaminosis D in 69% of the surveyed women at risk from Stara Zagora and this leads to the next recommendations: acquisition of knowledge about the importance of vitamin D by the women at risk health, regular following of vitamin D the status, the implementation of adequate preventive, and medical therapies.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author on request without personal items.

References

1. 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(11):1498-513. https://doi.org/10.1038/s41430-020-0558-y

   PMid:31959942
2. Meeham M, Penckofer S. The role of vitamin D in the aging. J Aging Gerontol. 2014;2(2):60-71. https://doi.org/10.12974/2309-6128.2014.02.02.1

   PMid:25893188
3. Buttros D, Nahas-Neto J, Nahas E, Cangussu LM, Barral AB, Kawakami MS. Risk factors for osteoporosis in postmenopausal women from southeast Brazilian. Rev Bras Ginecol Obstet. 2011;33(6):295-302. https://doi.org/10.1590/s0100-72032011000600006

   PMid:21877019
4. Tsiaras WG, Weinstock MA. Factors influencing vitamin D status. Acta Derm Venereol. 2011;9(1):115-24. https://doi.org/10.2340/00015555-0980

PMid:21384086

5. Sowah D, Fan X, Dennett L, Hagtvedt R, Straube S. Vitamin D levels and deficiency with different occupations: A systematic review. BMC Public Health. 2017;17(1):519. https://doi.org/10.1186/s12889-017-4436-z

PMid:28637448

6. Divakar U, Sathish T, Soljak M, Bajpai R, Dunleavy G, Visvalingam N, *et al*. Prevalence of Vitamin D deficiency and its associated work-related factors among indoor workers in a multi-ethnic Southeast Asian country. Int J Environ Res Public Health. 2020;17(1):164. https://doi.org/10.3390/ijerph17010164

PMid:31881679

7. Goulet J, Provencher V, Piché ME, Lapointe A, John Weisnagel S, Nadeau A, *et al*. Relationship between eating behaviors and food and drink consumption in healthy postmenopausal women in a real-life context. Br J Nutr. 2008;100(4):910-7. https://doi.org/10.1017/S0007114508925459

PMid:18279556

8. Semra RD, Garrett E, Johnson BA, Guralnik JM, Fried LP. Vitamin D deficiency among older women with and without disability. Am J Clin Nutr. 2000;72(6):1529-34. https://doi.org/10.1093/ajcn/72.6.1529

PMid:11101482

9. Lowe NM, Mitra SR, Foster PC, Bhojani I, McCann J. Vitamin D status and markers of bone turnover in Caucasian and South Asian postmenopausal women living in the UK. Br J Nutr. 2010;103(12):1706-10. https://doi.org/10.1017/S0007114509993850

PMid:20102676

10. Brejman L, Frieman J, Olshen R, Stone CJ. Classification and Regression Trees. New York: Chapman & Hall; 1984.

11. Borissova A-M, Shinkov A, Vlahov J, Dakovska L, Todorov T, Svinarov D, *et al*. Frequency of deficiency, insufficiency and sufficiency of vitamin D in Bulgarian population (≥20-80 years old). Endocrinologia. 2012;17(3):122-34.

12. Vimaleswaran KS, Berr DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, *et al*. Causal relationship between obesity and vitamin D status: Bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. 2013;10(2):e1001383. https://doi.org/10.1371/journal.pmed.1001383

PMid:23393431

13. Larose CY, Camargo JR, Langhammer A, Romundstad P, Mai XM. Factors associated with vitamin D deficiency in a Norwegian population: The HUNT study. J Epidemiol Community Health. 2014;68(2):165-70. https://doi.org/10.1136/jech-2013-202587

PMid:24197920

14. Skaaby T, Husemoen LL, Thuesen BH, Pisinger C, Hannemann A, Jørgensen T, *et al*. Longitudinal associations between lifestyle and vitamin D: A general population study with repeated vitamin D measurements. Endocrine. 2016;51(2):342-50. https://doi.org/10.1007/s12020-015-0641-7

PMid:26024976

15. Tardelli VS, Lago M, Silveira DX, Fidalgo TM. Vitamin D and alcohol: A review of the current literature. Psychiatry Res. 2017;248:83-6. https://doi.org/10.1016/j.psychres.2016.10.051

PMid:28033511

16. Ogunsakin O, Hottor T, Mehta A, Lichtveld M, McCaskill M. Chronic ethanol exposure effects on vitamin D levels among subjects with alcohol use disorder. Environ Health Insights. 2016;10:191-9. https://doi.org/10.4137/EHI.S40335

PMid:27795667

17. Rapuri PB, Gallagher JC, Nawaz Z. Caffeine decreases vitamin D receptor protein expression and 1,25(OH)2D3 stimulated alkaline phosphatase activity in human osteoblast cells. J Steroid Biochem Mol Biol. 2007;103(3-5):368-71. https://doi.org/10.1016/j.jsbmb.2006.12.037

PMid:17223552

18. Koundourakis NE, Margioris AN. Vitamin D and physical activity. In: A Critical Evaluation of Vitamin D-Basic Overview. London: IntechOpen; 2017. https://doi.org/10.5772/65103

19. Nascimento NAP, Moreira PFP, Carvalho VA, Aragão L, Marin-Mio RV, Lazaretti-Castro M, *et al*. Effect of vitamin D level and physical exercise on the physical performance and functional test results in elderly women. J Geriatr Med Gerontol. 2019;5(1):061.

20. Lee CJ, Kim SS, Suh WY, Kim JS, Jung JG, Yoon SJ, *et al*. The effect of education and vitamin D supplementation on the achievement of optimal vitamin D level in Korean postmenopausal women. J Bone Metab. 2019;26(3):193-9. https://doi.org/10.11005/jbm.2019.26.3.193

PMid:31555616

21. Leger-Guist'hau J, Domingues-Faria C, Miolanne M, Peyrol F, Gerbaud L, Perreira B, *et al*. Low socio-economic status is a newly identified independent risk factor for poor vitamin D status in severely obese adults. J Hum Nutr Diet. 2017;30(2):203-15. https://doi.org/10.1111/jhn.12405

PMid:27524803

22. 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. 2020;40(5):3436-47. https://doi.org/10.1016/j.clnu.2020.11.019

PMid:33309415

23. Purdon G, Comrie F, Rutherford L, Marcinkiewicz A."Vitamin D Status of Scottish Adults: Results from the 2010 & 2011," Scottish Health Surveys; 2013.

24. Davis SV. The Relationship Between Socioeconomic Status and Body Mass Index on Vitamin D Levels in African American Women with and without Diabetes Living in Areas with Abundant Sunshine, Graduate Theses and Dissertations, University of South Florida; 2013.

25. Wyskida M, Owczarek A, Szybalska A, Brzozowska A, Szczerbowska I, Wieczorowska-Tobis K, *et al*. Socio-economic determinants of vitamin D deficiency in the older Polish population: Results from the PolSenior study. Public Health Nutr. 2018;21(11):1995-2003. https://doi.org/10.1017/S1368980017003901

PMid:29352837