Practical Vitamin D Supplementation Using Machine Learning

Tyler Fugere, Zhongning Jim Chen, Issam Makhoul

Division of Hematology and Oncology, Department of Internal Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Corresponding author

Issam Makhoul
Division of Hematology and Oncology,
Department of Internal Medicine, University
of Arkansas for Medical Sciences, 4301 West
Markham Street #890, Little Rock, AR 72205,
USA
Tel: +1-501-686-8511
Fax: +1-501-686-6342
E-mail: makhoulissam@uams.edu

KSBMR
Korean Society for Bone and Mineral Research

Background: Patients with breast cancer are at increased risk of developing osteoporosis. Maintaining normal levels of vitamin D may decrease the risk of osteoporosis, and vitamin D levels must be corrected in patients who develop osteoporosis before beginning bone modifying agents. Therefore, it is important to correct insufficient vitamin D levels in a timely manner. In clinical practice, current guidelines for replacement regimens often fail to rapidly correct vitamin D levels. The goal of this study was to review data in order to predict what replacement regimen(s) were most effective at repleting vitamin D levels. **Methods:** For this retrospective cohort study, data was collected from medical records of 2,164 female patients with breast cancer with Institutional Review Board approval. Total level change per week was the primary outcome and was compared for the most commonly used vitamin D replacement regimens adjusted for age, race, body mass index, creatinine clearance, endocrine therapy, and initial level. **Results:** Higher weekly doses of vitamin D supplementation had a more significant impact on the rate of correction compared to lower daily doses. Generalized linear model was used to develop an online calculator that predicts time to vitamin D level correction adjusted for significant patient characteristics for 5 common replacement regimens as well as no intervention. **Conclusions:** When choosing a vitamin D replacement regimen for patients with vitamin D deficiency, we recommend clinicians use the online calculator to ensure that the chosen regimen will enable the patient to reach vitamin D sufficiency in a timely manner.

Key Words: Breast neoplasms · Osteoporosis · Vitamin D

INTRODUCTION

Vitamin D deficiency results in decreased calcium absorption from the intestines, which leads to increased osteoclast activity and enhanced mobilization of calcium from the bone. If vitamin D deficiency is not corrected, calcium continues to be pulled from the bone, and osteoporosis can occur.[1]

Patients treated for breast cancer have additional risk factors for osteoporosis. Many patients are postmenopausal and have been treated with endocrine therapy or chemotherapy. Aromatase inhibitor-associated bone loss occurs at a rate at least 2-fold higher than bone loss seen in healthy, age-matched postmenopausal women, resulting in a significantly higher fracture incidence.[2] Chemotherapies such as taxane, doxorubicin, 5-fluorouracil, cyclophosphamide, methotrexate, and cisplatin cause an increase in bone resorption independent of bone metastasis

and a reduction in bone structure.[3] Due to these risk factors, breast cancer patients have an increased risk of osteoporosis and fracture. Clinicians should ensure that these patients maintain adequate vitamin D levels to mitigate this risk.

Maintaining normal levels of vitamin D may decrease the risk of osteoporosis as well as fractures and falls in older adults. One study gave participants 100,000 IU of vitamin D3 or placebo once every 4 months for 5 years (15 doses total). Participants in the vitamin D treatment group had a 22% lower rate of fracture at any site and a 33% lower rate of fracture at any major osteoporotic site (hip, wrist, forearm, or vertebrae).[4] Despite the large prevalence of vitamin D insufficiency and the importance of maintaining vitamin D sufficiency in breast cancer patients, there is limited data to support a proper dosing regimen to correct and maintain sufficient vitamin D levels. According to the Institute of Medicine guidelines established in 2011, the recommended daily allowance of vitamin D for adults up to age 70 years is 600 IU/d and 800 IU/d for adults aged 71 years and older with an upper limit of 4,000 IU/d for all adults age 18 and older.[5] However, we have observed in clinical practice that patients often require at least 50,000 IU weekly to reach therapeutic vitamin D levels. One study found that the most effective regimen to correct vitamin D deficiency (25-hydroxy-vitamin D [25(OH)D] levels <20 ng/mL) and vitamin D insufficiency (25[OH]D <30 ng/mL) was ergocalciferol 50,000 IU 3 times a week for 6 weeks, and only regimens with a total dose >600,000 IU administered for a mean time of 60 ± 40 days of vitamin D raised 25(OH)D levels in the majority of patients. This study also reported no cases of vitamin D toxicity.[6] Another study gave patients increasing doses of vitamin D_3 from 28,000 IU/wk up to 280,000 IU/wk, ten times the recommended upper limit, with no adverse effects,[7] suggesting that it may be safe for patients to take much more than the recommended upper limit of 4,000 IU/d. The purpose of this study is to review treatment regimens prescribed at our institution to determine if any regimen is superior to other regimens for correcting vitamin D levels in breast cancer patients and if any patient characteristics predictably influence the regimen necessary to reach sufficiency. This information was used to design an online calculator that would guide the choice of the initial dose of vitamin D and the time needed to reach the target level.

METHODS

1. Data collection

This retrospective cohort study looked at breast cancer patients who are being followed at the outpatient clinic. Eligible patients were female patients with breast cancer who had vitamin D level checked between 2011 and 2018. Variables that were collected included, age, race, status of endocrine therapy, status of osteoporosis, vitamin D regimens, initial and follow up vitamin D levels and dates collected. Patients who were male, under the age of 18, those with missing variables, those who had never been diagnosed with breast cancer, those who did not have vitamin D level checked or had it checked only once, and those with follow up dates more than 6 months from initiation of vitamin D replacement therapy were excluded. All patient data were obtained with Institutional Review Board approval.

2. Model generating

In order to estimate the change in vitamin D levels in response to different regimens adjusted for different variables, various regression models were tested to assess the accuracy of the response. Seven different factors were analyzed in order to predict the rate of vitamin D level changes (LCs) per week. Four continuous variables were evaluated: age, body mass index (BMI), creatinine, and initial vitamin D levels. Three categorical non-ranked variables included status of hormone modulation therapy, race, and vitamin D replacement regimen. Because the categorical variables are not ranked, they are expanded into either one or zero based on the presence or absence of possible outcome of each categorical variable. Total LC per week was the primary outcome to be predicted by various machine learning algorithms from R package Caret: (1) support vector machines using a radial basis function kernel with tune length of 15; (2) elastic net regression with tune grid alpha ranging from 0 to 1 by steps of 0.05 and lambda by powers of tenths from 0 to 0.1; (3) generalized linear model (GLM) with tune length of 15; (4) partial least square with tune length of 15; (5) Earth Model with degree of 1 and prune length between 2 to 50; and (6) cubist regression with committee ranging from 0 to 100, and neighbors range between 0 to 9. All data passed into the models were scaled and centered via cox-box transformation.

Table 1. Demographic data

	0 units (n=100)	50,000 units weekly (n=345)	100,000 units weekly (n=46)	50,000 units monthly (n=27)	1,000 units daily (n=31)	2,000 units daily (n=46)	5,000 units daily (n=21)	Overall (n=616)
Age (yr)								
Mean±SD	58.9±14.2	61.4±12.5	56.3±12.5	63.1±14.8	61.2±13.6	68.7±11.6	63.7±11.9	61.3±13.1
Median (min, max)	59.0 (20.0, 93.0)	62.0 (22.0, 94.0)	59.0 (25.0, 80.0)	65.0 (36.0, 94.0)	63.0 (42.0, 87.0)	69.5 (28.0, 86.0)	63.0 (35.0, 87.0)	62.5 (20.0, 94.0)
Race								
White	56 (56.0%)	224 (64.9%)	33 (71.7%)	12 (44.4%)	25 (80.6%)	31 (67.4%)	15 (71.4%)	396 (64.3%)
Non-white	44 (44.0%)	121 (35.1%)	13 (28.3%)	15 (55.6%)	6 (19.4%)	15 (32.6%)	6 (28.6%)	220 (35.7%)
BMI (kg/m²)								
Mean±SD	31.4±7.38	31.3±7.26	30.4±80.8	29.5±7.90	29.6±6.95	31.0±6.82	30.2±4.14	31.0±7.23
Median (min, max)	29.6 (19.5, 57.0)	29.5 (16.9, 57.0)	27.0 (18.4, 51.8)	26.3 (21.0, 48.4)	27.0 (20.2, 48.2)	30.9 (18.4, 49.0)	29.6 (22.2, 37.4)	29.5 (16.9, 57.0)
Creatinine clearance (mL/min)								
Mean±SD	84.2±37.3	90.4±37.1	106±42.3	96.4±63.5	83.1±35.0	75.8±38.0	102±40.0	89.8±39.5
Median (min, max)	87.7 (13.1, 190)	88.1 (19.6, 285)	98.8 (36.0, 225)	87.5 (13.2, 285)	88.4 (27.4, 140)	72.3 (23.2, 224)	83.8 (49.4, 225)	88.2 (13.1, 285)
Endocrine therapy								
No therapy	32 (32.0%)	59 (17.1%)	8 (17.4%)	6 (22.2%)	10 (32.3%)	8 (17.4%)	7 (33.3%)	130 (21.1%)
Had therapy	68 (68.0%)	286 (82.9%)	38 (82.6%)	21 (77.8%)	21 (67.7%)	38 (82.6%)	14 (66.7%)	486 (78.9%)
Initial level (ng/mL)								
Mean±SD	24.8±14.2	25.8±15.8	20.0±11.7	34.2±17.7	26.5±9.86	29.5±14.4	24.8±9.36	25.8±15.0
Median (min, max)	23.0 (5.00, 89.6)	22.4 (4.20, 94.0)	18.5 (4.20, 59.8)	26.8 (10.8, 94.0)	23.8 (11.4, 52.8)	29.0 (5.50, 103)	25.2 (7.70, 46.1)	23.2 (4.20, 103)
Duration (week)								
Mean±SD	87.0±50.6	101±37.3	100±29.6	90.6±46.0	99.3±47.5	114±42.4	111±41.1	99.7±41.1
Median (min, max)	87.5 (9.00, 179)	98.0 (8.00, 178)	94.5 (56.0, 170)	88.0 (9.00, 175)	103 (9.00, 175)	124 (31.0, 178)	115 (15.0, 176)	98.0 (8.00, 179)

BMI, body mass index; SD, standard deviation.

In order to assess the strength of each model, the data was split (4:1) into a *training set*, Tr and testing set, Ts (or validation set). All models are trained using only the training set and tested in a separate testing set in order to avoid overfitting. To create each model, a method called *20-fold cross validation* is done on the training set. This is achieved by first partitioning the Tr into n randomly chosen subsets labeled {s(1),...,s(20)}. One by one, each subset (s(i) for each i between 1 to 20) , is taken out of the Tr and the remaining Tr without s(i) (denoted Tr\s(i)) is used to compute parameters of a particular model. The LC of the s(i) (Lci) is compared with the calculated LC (Lc*) derived from the model. The root mean square estimate (RMSE) is then computed to assess its predictive strength. After each iteration, the parameters of the model are adjusted to achieve the minimum RMSE. This is done for each partition of Tr and the whole process is repeated 5 times. Afterwards, Ts, which was never used in the training of the model, was used to compute the predicted weekly LC in the testing set. The Pearson's correlation (r) between the observed and computed weekly LCs were computed. The RMSE of each model as well as the r were used to determine the strength of each model. For the sake of reproducibility, all random seeds were set to 42 prior to any random sampling. Computations were carried out in R (version 3.5.3; The R Foundation for Statistical Computing, Vienna, Austria) using Applied-PredictiveModeling version 1.1-7 and caret version 6.0-84.

3. Calculator

Using GLM, we have created an online calculator for clinicians to use to determine the expected LCs per week and the length of therapy required to reach sufficiency if a patient is given a replacement regimen input by the clinician. A website was created dedicated to this task that was written in Python 3.5.3 using the Django version 2.2.6. Because of the versatility of R for data analysis and potential for new tools to be created, the GLM is stored and computed in R. The 2 languages were connected using python's rpy2 version 2.9.4.

RESULTS

In total 2,164 breast cancer patients had a vitamin D 25(OH) level checked between 2011 and 2018. Typically, patients had been on multiple regimens. The seven most common regimens used included 50,000 units weekly, 100,000 units weekly, 50,000 units monthly, 1,000 units daily, 2,000 units daily, 5,000 units daily, and no therapy. After the exclusionary criteria were applied, a total of 616 cases among 379 patients were included in the study. Demographics are shown in Table 1.

As demonstrated in Table 2, several of the models had comparable strength. Because of the ease of interpretation of the impact of each variable, GLM was chosen. Table 3 looks at the breakdown of GLM. Variables with significant

Table 2. Root mean square estimate

	RMSE[a]	r cross validation[b]	RMSE test set[c]	r test set[c]
SVM	1.209	0.567	1.284	0.588
GLM	1.189	0.580	1.270	0.636
Elastic net	1.402	0.469	1.273	0.637
Cubist	1.189	0.589	1.209	0.647
PLS	1.187	0.581	1.269	0.634
Earth model	1.186	0.585	1.228	0.654

Each model attempts to predict weekly vitamin D level changes based on treatment doses of vitamin D and various clinical and demographic features. A 20 fold cross validation with 5 repeated cycles was used to create the model from the training set. The hyperparameters of each model are adjusted to find the minimum RMSE. [a]RMSE training set, which is the average of the minimum RMSE for each training model. [b]Pearson's correlation between the observed and the computed weekly level changes. [c]RMSE and R of the test set. Independent variables were age, body mass index, creatinine, and initial vitamin D level. Adjusted variables were race (white or non-white), endocrine therapy, and the different regimens. The training set to test set ratio was 4:1.
RMSE, root mean square estimate; SVM, support vector machines; GLM, generalized linear model; PLS, partial least square.

Table 3. Generalized linear model with respect to variables

	Estimate	Standard error	T value	P-value
Intercept	1.482	0.529	2.801	0.005[b]
Age	0.014	0.006	2.28	0.023[a]
Body mass index	-0.028	0.01	-2.766	0.006[b]
Creatinine	-0.002	0.002	-1.062	0.289
Initial level	-0.062	0.004	-16.913	$<2e-16$[c]
Non-white	0.001	0.138	0.006	0.995
Hormone therapy	0.151	0.162	0.934	0.351
50,000 units weekly	1.436	0.184	7.824	2.16e-14[c]
100,000 units weekly	2.098	0.284	7.388	4.75e-13[c]
50,000 units monthly	0.498	0.347	1.433	0.152
1,000 units daily	0.377	0.336	1.124	0.261
2,000 units daily	0.646	0.293	2.205	0.028[a]
5,000 units daily	0.761	0.395	1.928	0.054

[a]Strong evidence, [b]stronger evidence, and [c]the strongest evidence according to the *P*-value.
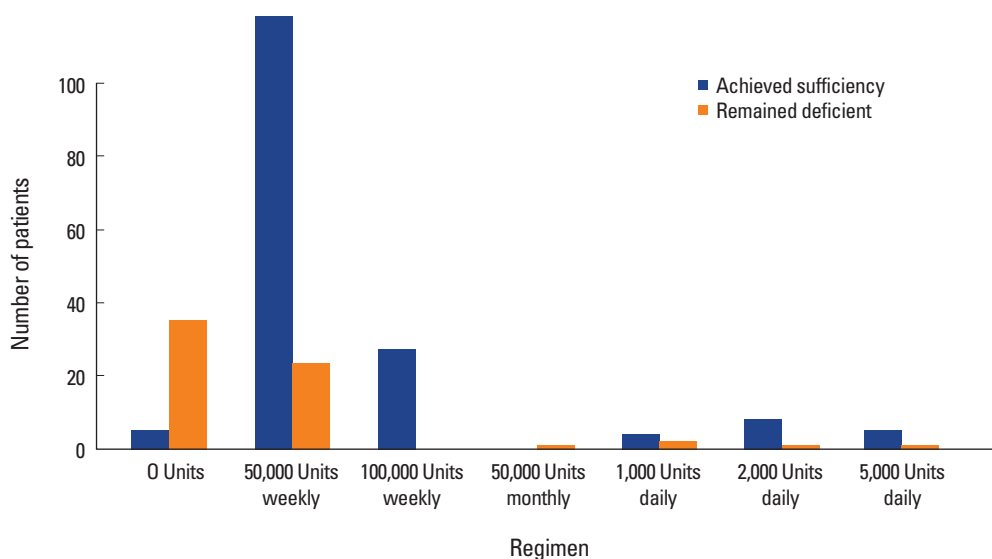
Fig. 1. Response to supplementation by regimen.

Table 4. Odds ratio of vitamin D level increased based on regimen and median weekly level change without adjusting for demographic variables

	Odds ratio	2.50%	97.50%	P-value	Median weekly change
50,000 units weekly	8.0278	5.000	13.108	2.09E-17[b]	0.954
100,000 units weekly	23.500	8.705	82.584	1.77E-08[b]	1.676
50,000 units monthly	1.222	0.509	2.841	0.644659	-0.259
1,000 units daily	3.167	1.398	7.422	0.006414[a]	0.260
2,000 units daily	4.133	2.006	8.835	0.000164[b]	0.298
5,000 units daily	6.400	2.299	20.886	0.000782[b]	0.513

[a]Strong evidence, [b]the strongest evidence according to the P-value.

effect on vitamin D supplementation include age (0.18, $P<0.01$), initial level (-0.64, $P<0.001$), a regimen of 50,000 U weekly (0.36, $P<0.01$) and regimen of 100,000 U weekly (0.31, $P<0.001$) with an intercept of 0.7 ($P<0.001$). Notable other coefficients that are not statistically significant include race of non-white (-0.1, $P<0.07$) and prior or current use of hormone therapy (0.09, $P<0.1$). Although not statistically significant, hormone therapy was a positive predictor for the rate of correction in response to vitamin D supplementation.

In order to assess the response to supplementation adjusted for regimen, Figure 1 shows the breakdown of patients whose levels corrected and patients who remained insufficient (vitamin D level less than or equal to 20 ng/mL) in response to different regimens. Compared with no treatment, the odds of achieving vitamin D sufficiency from deficiency for 50,000 units weekly has odds ratio (OR) of 8.0278 (95% confidence interval [CI], 5.000-13.108; $P<2.09e-17$) and 100,000 units weekly has OR of 23.500 (95% CI, 8.705-82.584; $P<1.7eE-08$), as demonstrated in Table 4.

DISCUSSION

This study was performed to evaluate the efficacy of various regimens in correcting vitamin D levels in patients with vitamin D deficiency/insufficiency and a history of breast cancer. Because the RMSE of the linear model is very similar to other models, it is reasonable to use it as a basis to assess the impact of each coefficient. As we discussed earlier, higher weekly doses of vitamin D supplementation had a more significant impact on the rate of correction compared to lower daily doses. We hypothesize that this is likely due to total higher dose overall, as patients getting 1,000 IU daily, 2,000 IU daily, or 5,000 IU daily get a total of 7,000 IU, 14,000 IU, or 35,000 IU per week respectively compared to 50,000 IU or 100,000 IU.

Because vitamin D is a fat soluble vitamin, which can be stored in the body, overcorrection has the potential to be harmful as well as undercorrection. Therefore a calculator, Figure 2, was made using the GLM in order to predict the expected weekly response rate to help clinicians determine the duration of therapy and frequency of lab checks based on the patient's clinical features. Until a stronger

Fig. 2. Vitamin D replacement calculator.

model is established from collecting more data, the linear model will be used for prediction because of the ease of interpretation of its coefficients. We recommend clinicians use the online calculator by entering the patient's age, BMI, creatinine clearance, initial vitamin D level, race, and hormone therapy status and selecting the regimen the clinician is considering prescribing in order to determine the amount of time it would take for the patient to reach sufficiency on that regimen. Considering this information for each treatment regimen enables the clinician to choose an appropriate dose that will correct the patient's vitamin D deficiency in a timely manner.

Time to vitamin D level normalization is an important factor in clinical practice in patients with osteoporosis. Those patients need a bone modifying agent to increase their bone mass. However, it has been shown that initiating this treatment before the correction of vitamin D deficiency

can be detrimental. Hence the need for rapid correction of these levels.[8]

Weaknesses of this retrospective study include inability to ensure vitamin D levels were checked at regular intervals and before regimen change. Some providers prefer regimens that include a dose change, such as 50,000 IU weekly for 6 weeks then 800 IU daily. If the level was not checked after the initial six weeks of 50,000 IU weekly, it is impossible to tell retrospectively if the level peaked after the initial 6 weeks and then declined on the 800 IU daily and will approach deficiency again if 800 IU is continued daily, or if the level has remained steady on 800 IU daily. Many patients were given multiple regimens and changed to new regimens often without level checks, likely due to provider preference in light of the lack of established guidelines for replacement. Patients with multiple regimens were not able to be included in the study.

Due to limitations of retrospective review and limited sample size, we were unable to randomize patients to various treatment regimens. Initial vitamin D levels had a statistically significant impact on the rate of vitamin D correction. It is possible that patients with lower initial levels were reactionarily given higher dosing regimens compared to those who were only mildly deficient. In the future, a study with patients randomized to various treatment regimens controlled for initial vitamin D levels would be warranted. It would be worthwhile to compare 100,000 IU weekly, 50,000 IU weekly, and 10,000 IU daily to evaluate the hypothesis that total dose is the most important factor. In that case, it would be expected that 10,000 IU would correct levels faster than 50,000 IU weekly, but not as rapidly as 100,000 IU weekly.

This study population is limited to breast cancer patients and may not be generalizable to the other patient populations. Additional studies with larger sample sizes that are not limited to breast cancer patients would be beneficial. Future studies to evaluate the accuracy of the online calculator are also warranted.

DECLARATIONS

Acknowledgments

Ethics approval and consent to participate

This study was approved by the University of Arkansas for Medical Sciences Institutional Review Board.

Conflict of interest

No potential conflict of interest relevant to this article was reported.

ORCID

Tyler Fugere   *https://orcid.org/0000-0003-4548-9105*
Issam Makhoul   *https://orcid.org/0000-0001-8646-5658*

REFERENCES

1. Sunyecz JA. The use of calcium and vitamin D in the management of osteoporosis. Ther Clin Risk Manag 2008;4: 827-36.

2. Hadji P, Aapro MS, Body JJ, et al. Management of aromatase inhibitor-associated bone loss (AIBL) in postmenopausal women with hormone sensitive breast cancer: Joint position statement of the IOF, CABS, ECTS, IEG, ESCEO IMS, and SIOG. J Bone Oncol 2017;7:1-12.

3. Makhoul I, Montgomery CO, Gaddy D, et al. The best of both worlds - managing the cancer, saving the bone. Nat Rev Endocrinol 2016;12:29-42.

4. Trivedi DP, Doll R, Khaw KT. Effect of four monthly oral vitamin D3 (cholecalciferol) supplementation on fractures and mortality in men and women living in the community: randomised double blind controlled trial. BMJ 2003; 326:469.

5. Ross AC, Manson JE, Abrams SA, et al. The 2011 report on dietary reference intakes for calcium and vitamin D from the Institute of Medicine: what clinicians need to know. J Clin Endocrinol Metab 2011;96:53-8.

6. Pepper KJ, Judd SE, Nanes MS, et al. Evaluation of vitamin D repletion regimens to correct vitamin D status in adults. Endocr Pract 2009;15:95-103.

7. Kimball SM, Ursell MR, O'Connor P, et al. Safety of vitamin D3 in adults with multiple sclerosis. Am J Clin Nutr 2007; 86:645-51.

8. Carmel AS, Shieh A, Bang H, et al. The 25(OH)D level needed to maintain a favorable bisphosphonate response is >/=33 ng/ml. Osteoporos Int 2012;23:2479-87.