Supplementary Online Content

Burt LA, Billington EO, Rose MS, Raymond DA, Hanley DA, Boyd SK. Effect of high-dose vitamin D supplementation on volumetric bone density and bone strength: a randomized clinical trial. *JAMA*. doi:10.1001/jama.2019.11889

This supplementary material has been provided by the authors to give readers additional information about their work.

Study Investigators

Principal investigators
Dr. David A Hanley
Dr. Steven K Boyd

Additional Methods

Statistical analysis

Visual inspection of individual profile plots was done to assess the potential of a non-linear effect and if evident was included in the model as a quadratic treatment effect.

A constrained linear mixed effects model was used to analyze each outcome (both primary and secondary) variables. Baseline values were constrained to be the same in each of the three treatment groups, since these were measured prior to randomization. The dependent variable included all measurements taken post baseline. This is described in more detail below.

Fixed effects include, treatment group and a treatment by time interaction (which if significant was indicative of efficacy). Potential random effects included Time (both intercept and slope), when appropriate. The correlation within multiple observations on each participant was modelled as an autoregressive error of order 1. Diagnostic residual plots were examined for deviation from the assumptions underlying the model. We started with the full model. Unnecessary quadratic terms were removed as were unnecessary random slopes or intercept terms to arrive at a parsimonious well-fitting model for each outcome variable. Non-significant treatment by time interactions were retained in the model and reported to describe the lack of observed treatment effect. Missing values were accounted for by using the linear mixed effects model which is considered a superior method to that of using Last Observation Carried Forward (LOCF).

The results of the regression modelling are described in table format (eTable 4) where the Likelihood Ratio Statistic (LRS) for the quadratic term on 1 df (before removal in cases where this term was unnecessary) and the LRS on 2 df for the treatment group by time interaction, which yield the p-value for the treatment effect. Since coefficients for both quadratic terms and interaction terms are hard to interpret, results of the regression models were presented as the mean (with 95% CI) fixed effects calculated from the coefficients.

Details of the statistical model

For the analysis of the primary and secondary variables we used a constrained multilevel mixed linear regression model. This is described in more here.

Let Y_{ijt}represent the observation of the i^{th} *(i= I,…,n)* participant in the j^{th} treatment Group *(j = 1,2,3)* at Time *t* months *(t = 0,6,12,24,36).* Then the model can be written as

$$Y_{ijt} = \beta_0 + \beta_1 \times T_{ij} + \beta_2 \times T_{ij}^2 + \beta_{3j} \times G_j \times T_{ij} + \mu_{i0} + \mu_{i1} T_{ij} + \varepsilon_{ijt}$$

Because Group is a 3-category variable (400 IU, 4000 IU and 10,000 IU) it is equivalent to two dummy variables and therefore β_3 will consist of two different coefficients, $\beta_{31}\ and\ \beta_{32}$, where β_{31} is the difference between 4000 IU Group and the baseline 400 IU group and β_{32} is the difference between 10,000 IU Group and the baseline 400 IU group. It is implied that the coefficient for the baseline category (in this case 400 IU) $\beta_{30} = 0$.

The Model is considered a **multilevel model** because there are multiple observations on each participant (i.e. repeated measures). In a model with multiple observations on each individual it is necessary to address the correlation structure between the observations. An **autoregressive error of order one** means that each observation is more highly correlated to the one before it than to any other observation.

The Model is considered **Mixed** because it has both **fixed** and **random** effects. Treatment Group are considered a fixed effect, but the regression on time is a random effect. We considered both random intercepts and random slopes for Time. This is equivalent to fitting an individual line for each participant and therefore allowing the slopes and intercepts to vary. The amount of variation is captured by the error terms $\mu_{i0}\ and\ \mu_{i1}$ which are assumed to be Normally distributed around zero with a variance that captures the amount of variation in the sample. Therefore, the coefficients $\beta_0\ and\ \beta_1$ can be thought of as the average intercept and slope.

The reason that we used a **constrained** linear regression model was to ensure that each treatment group had the same estimated intercept (i.e. Baseline average value). Since this measurement was taken prior to randomization and taking a vitamin D dose, these values are not statistically significantly different (unless due to chance alone) and the constrained model fits them to be exactly the same, so that direct comparisons at follow-up times can be made between the treatment groups [1,2].

To interpret the model and understand the **constrained** part of the model it is easier if we write the regression in a different form which allows us to drop the error terms. In this case, $\hat{Y}_{jt}$ is the **estimated, predicted or fitted** value of the outcome variable for an individual in group *j* at time *t.* Theoretical or population values of the coefficients β_i are differentiated from the estimated counterparts by putting a "hat" on top to indicate the estimated coefficient (from the data) $\hat{\beta}_i$.

Then the regression equation can be written as

$$\hat{Y}_{ijt} = \hat{\beta}_0 + \hat{\beta}_1 \times T_{ij} + \hat{\beta}_2 \times T_{ij}^2 + \hat{\beta}_{3j} \times G_j \times T_{ij}$$

So, for participants in Group 0 (400 IU)

$$\hat{Y}_{1t} = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 T^2$$

Gives the estimated value at month *t* and for participants in Group 1 (4000 IU)

$$\hat{Y}_{2t} = \hat{\beta}_0 + \hat{\beta}_1 T + \hat{\beta}_2 T^2 + \hat{\beta}_{31} T = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_{31}) T + \hat{\beta}_2 T^2$$

And for those in Group 3, the estimated values is

$$Y_{3t} = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_{32}) T + \hat{\beta}_2 T^2$$

So, at a given time *t* months, the difference between the 4000 IU group and the 400 IU group is simply the difference between $\hat{Y}_{1t}$ and $\hat{Y}_{2t}$ which is $\hat{\beta}_{31} t$ and the 95% CI for the difference at time t months is (LL*t, UL*t).

Therefore, if we use Tt.BMD Radius as an example the difference between 400 IU and 4000 IU at 12 months is -1.3 with 95% CI -2.1, -0.4 calculated from 12*-0.108, and 12*-.179 and 12*-0.037. The difference between the 10,000 IU group and the 400 IU group at 12 months is almost twice that at -2.5 with 95% CI -3.3, -1.7.

Statistical Inference

In the analysis of randomized controlled trials, we aimed to keep the hypothesis testing to a minimum for very good reasons. Multiplicity of p-values can arise from many sources. Therefore, we limit the number of primary variables to a minimum and adjust the alpha value at which we are willing to declare statistical significance to avoid the error of declaring a treatment effect when the difference observed has arisen due to chance alone. Another type of multiplicity can occur, such as in this case when we have more than two treatment arms. In the case of a three-arm clinical trial it is appropriate to use a global test, with one resulting p-value, to determine whether there is a difference between the three groups. Deciding where the differences lie is a matter of controversy. Some advocate in favor of more statistical testing to determine where the actual differences lie with a p-value. Many methods have been developed to adjust p-values to try to maintain the experiment wise error rate, but none of these methods are very good, besides the results are simply a set of p-values which tells the reader very little. Others are in favor of reporting the actual differences between the groups with 95% Confidence Intervals which provides more information. The precision of the estimate gives an idea of the statistical significance and allows the reader to determine which groups are different from one another, and whether these differences are clinically relevant. We chose the latter route, that is we examined the treatment effect with a global p-value resulting from the likelihood ratio test (on 2 df) i.e. that there is a difference between the three treatment groups Had we decided to use Wald statistics which involve comparing the estimate with its standard error and thereby producing a p-value for each treatment effect, this would have doubled the number of p-values for our primary analysis, which would have necessitated reducing the alpha level even further and perhaps reducing power in order to collect a large enough sample in a timely manner. We also consider that comparing the 4000 IU group to the 10000 IU group is an unnecessary comparison. In addition, caution is advised in using Wald statistic particularly in mixed models.

In developing the best model for each outcome variable, we used some ancillary hypothesis testing to arrive at a parsimonious model that provided the best fit to the data by examining the need for a quadratic term, random intercepts, random slopes and the autoregressive error. We started with a full model and tested each of the terms sequentially. If a term was deemed unnecessary it was removed and the modelling process restarted. We have reported the results of certain tests in the results below. A quadratic term was included in the model to allow for non-linearity, and the Likelihood Ratio Statistic (LRS) on 1 degree of freedom (df) and accompanying p-value is the test for non-linearity. When the (non-significant) p-value is identified (*), this means that the non-linear was tested and then not included in the model.

The Likelihood Ratio Statistic (LRS) on 2 degree of freedom (df) and accompanying p-value is the test for a significant interaction between time and treatment group. A significant p-value (< 0.05) indicates that the trajectory of the mean values of the treatment groups differ over time, resulting in a significant treatment effect. Interpretation of this treatment effect is shown using the manuscript figures.

The Intercept value is the mean value in all the three treatment groups at baseline. This is constrained to be the same under this model, since the groups have been formed by randomization.

Additional Results

A total of 542 participants were screened, and 373 met inclusion criteria and were randomized (pilot cohort: N = 62, main cohort: N = 311). Our pre-defined target was 300 participants, and the reason we included additional participants are two-fold. First, we began recruiting in anticipation of the availability of the second-generation HR-pQCT instrument, but unfortunately, due to the manufacturer's delay in applying for Health Canada approval of the new version of the scanner, baseline measurement of the primary outcome variables was impossible for the first 62 participants, who therefore had baseline measurements on our first-generation HR-pQCT. Because of differences between the first- and second-generation scanner with respect to resolution, we decided against mixing results from two generations of HR-pQCT scanners, and to exclude these participants from our primary analysis. At that time, we decided against mixing results from two generations of HR-pQCT scanners, and to exclude these participants from our primary analysis. We termed this group our 'pilot' cohort, and they completed the three-year trial, albeit without the complete primary outcome measures. Second, during recruitment, a total of 311 subjects were randomized to account for known attrition during the 12-month recruitment phase. We report our findings based on the 311 randomized participants who completed all aspects of the trial.

Due to motion artifact, 12 participants had at least one radius scan removed. Due to region of interest overlap <75%, 13 participants had radius scans at one time-point removed and one participant had all radius scans removed.

The biochemistry raw data are presented in eTable 1. Raw data for the primary (eTable 2) and secondary (eTable 3) variables are presented in the following tables. eTable 4 gives a summary of the statistical models for both the primary and secondary variables.

eFigure 1 shows the change in serum 25-hydroxyvitamin D, parathyroid hormone and C-telopeptide of type 1 collagen over the trial duration. eFigure 2 and eFigure 3 show the change in cortical and trabecular bone mineral density as well as the data distribution. eFigure 4 and eFigure 5 show the change in trabecular number and cortical porosity as well as the data distribution, over the duration of the trial.

eTable 1. Raw data for biochemistry

		N	0 months	3 months	6 months	12 months	18 months	24 months	30 months	36 months
25(OH)D (nmol/L)	400	106	76.30 (21.36), (0.00%)	76.71 (17.69), (0.00%)	73.55 (17.82), (0.94%)	71.31 (14.93), (1.89%)	73.40 (17.54), (3.77%)	76.64 (17.16), (5.66%)	73.27 (15.34), (5.66%)	77.35 (17.91), (5.66%)
	4000	98	81.33 (20.07), (1.02%)	115.25 (22.97), (0.00%)	117.07 (24.21), (1.02%)	116.24 (24.09), (2.04%)	124.30 (28.11), (3.06%)	125.02 (25.65), (4.08%)	128.07 (27.43), (4.08%)	132.16 (27.96), (4.08%)
	10000	101	78.41 (18.40), (0.00%)	188.02 (38.90), (0.00%)	194.27 (41.93), (0.99%)	193.91 (38.02), (0.99%)	200.35 (42.43), (1.98%)	160.41 (46.92), (1.98%)	131.74 (32.06), (4.95%)	144.38 (40.35), (4.95%)
PTH (ng/L)	400	106	22.23 (7.67), (0.00%)	22.33 (7.83), (0.00%)	22.14 (7.93), (1.89%)	21.23 (7.26), (2.83%)	22.48 (7.58), (4.72%)	21.95 (7.78), (5.66%)	22.64 (8.07), (5.66%)	23.39 (8.61), (5.66%)
	4000	98	21.45 (6.58), (1.02%)	20.56 (7.41), (0.00%)	20.31 (6.62), (2.04%)	19.20 (6.35), (2.04%)	19.83 (6.39), (4.08%)	19.08 (6.80), (5.10%)	21.00 (7.41), (4.08%)	21.24 (6.59), (5.10%)
	10000	101	22.11 (7.41), (0.00%)	19.22 (6.42), (1.98%)	17.72 (6.37), (0.00%)	16.64 (5.90), (0.99%)	17.60 (5.99), (1.98%)	18.11 (5.18), (3.96%)	19.58 (6.98), (3.96%)	19.05 (5.93), (4.95%)
CTx (ng/L)	400	106	338.63 (122.64), (0.00%)			332.83 (132.10), (0.94%)		333.85 (131.63), (1.89%)		364.60 (150.63), (5.66%)
	4000	97	339.35 (130.66), (0.00%)			332.55 (130.57), (0.00%)		344.10 (142.83), (1.03%)		351.32 (142.59), (5.15%)
	10000	101	344.85 (126.59), (0.00%)			340.68 (135.08), (0.00%)		360.78 (152.99), (0.99%)		393.80 (153.98), (3.96%)

Data are presented as mean (SD), (% missing). 25(OH)D = 25-hydroxyvitamin D, PTH = Parathyroid hormone, CTx = C-telopeptide of type 1 collagen. Laboratory normal range is: 25(OH)D: 80 – 200 nmol/L; PTH: 7 – 37 ng/L; CTX: 0 – 400 ng/L

eTable 2. Raw data for the primary outcome variables

		N	Baseline	6 Months	12 Months	24 Months	36 Months
TtBMD Radius (mg HA/cm ^{3})	400 IU	104	324.9 (61.5), (0.0%)	325.7 (61.8), (1.9%)	323.2 (61.1), (1.0%)	323.0 (61.8), (5.8%)	320.1 (61.1), (4.8%)
	4000 IU	96	335.9 (65.3), (0.0%)	334.6 (65.1), (2.1%)	333.1 (65.4), (2.1%)	330.4 (66.2), (4.2%)	328.6 (66.1), (6.2%)
	10000 IU	99	329.7 (60.0), (0.0%)	327.4 (60.3), (2.0%)	324.7 (59.7), (2.0%)	320.9 (60.1), (2.0%)	317.3 (61.7), (6.1%)
TtBMD Tibia (mg HA/cm ^{3})	400 IU	105	301.2 (58.3), (0.0%)	302.4 (58.4), (0.0%)	301.6 (58.9), (1.0%)	300.1 (58.4), (4.8%)	299.1 (58.7), (4.8%)
	4000 IU	97	314.1 (52.9), (0.0%)	314.3 (54.0), (1.0%)	312.2 (53.4), (2.1%)	310.3 (53.6), (4.1%)	309.1 (54.8), (4.1%)
	10000 IU	101	306.5 (52.6), (0.0%)	306.4 (52.3), (0.0%)	305.2 (52.7), (1.0%)	303.1 (53.7), (2.0%)	301.5 (54.6), (5.0%)
Failure Load Radius (N)	400 IU	104	2700.7 (1020.7), (0.0%)	2669.5 (1025.0), (1.9%)	2688.1 (989.9), (1.0%)	2672.4 (1015.6), (5.8%)	2694.8 (1022.7), (4.8%)
	4000 IU	96	2580.1 (990.4), (0.0%)	2555.0 (1005.3), (2.1%)	2599.7 (1012.4), (2.1%)	2577.5 (991.2), (4.2%)	2550.5 (994.3), (6.2%)
	10000 IU	99	2556.5 (964.4), (0.0%)	2550.7 (1001.9), (2.0%)	2517.6 (989.0), (2.0%)	2512.4 (1013.3), (2.0%)	2470.6 (1001.5), (6.1%)
Failure Load Tibia (N)	400 IU	105	7831.3 (2420.1), (0.0%)	7811.5 (2380.7), (0.0%)	7794.3 (2437.6), (1.0%)	7740.9 (2482.7), (4.8%)	7785.2 (2443.7), (4.8%)
	4000 IU	97	7660.6 (2001.2), (0.0%)	7595.0 (2023.7), (1.0%)	7621.7 (2029.2), (2.1%)	7536.4 (2057.5), (4.1%)	7576.9 (2125.0), (4.1%)
	10000 IU	101	7533.5 (2209.9), (0.0%)	7512.2 (2185.3), (0.0%)	7452.4 (2206.4), (1.0%)	7418.7 (2261.9), (2.0%)	7413.0 (2265.1), (5.0%)

Data are presented as mean (SD), (% missing). TtBMD = total bone mineral density

eTable 3. Raw data for the secondary outcome variables

		N	Baseline	6 Months	12 Months	24 Months	36 Months
CtBMD Radius (mg HA/ cm^3)	400 IU	104	887.6 (50.0), (0.0%)	884.8 (49.9), (1.9%)	882.2 (51.2), (1.0%)	882.2 (49.4), (5.8%)	879.6 (51.0), (4.8%)
	4000 IU	96	899.2 (51.3), (0.0%)	897.2 (49.4), (2.1%)	890.4 (51.1), (2.1%)	889.2 (52.1), (4.2%)	886.3 (52.4), (6.2%)
	10000 IU	99	904.0 (53.3), (0.0%)	899.3 (52.7), (2.0%)	891.9 (53.5), (2.0%)	889.8 (53.2), (2.0%)	884.9 (54.2), (6.1%)
CtBMD Tibia (mg HA/ cm^3)	400 IU	105	853.9 (61.5), (0.0%)	851.9 (63.2), (0.0%)	848.2 (65.8), (1.0%)	846.6 (67.3), (4.8%)	848.1 (68.8), (4.8%)
	4000 IU	97	868.6 (52.9), (0.0%)	864.9 (55.8), (1.0%)	860.0 (56.8), (2.1%)	856.1 (59.7), (4.1%)	856.6 (62.9), (4.1%)
	10000 IU	101	871.5 (59.0), (0.0%)	868.5 (59.5), (0.0%)	860.8 (64.2), (1.0%)	855.2 (65.4), (2.0%)	853.7 (67.9), (5.0%)
TbBMD Radius (mg HA/ cm^3)	400 IU	104	163.1 (40.3), (0.0%)	164.4 (40.6), (1.9%)	164.4 (41.3), (1.0%)	165.2 (42.1), (5.8%)	164.8 (42.5), (4.8%)
	4000 IU	96	160.4 (39.9), (0.0%)	159.9 (39.8), (2.1%)	160.9 (40.3), (2.1%)	160.9 (41.6), (4.2%)	161.5 (42.0), (6.2%)
	10000 IU	99	155.9 (40.2), (0.0%)	155.7 (40.3), (2.0%)	156.1 (40.3), (2.0%)	155.4 (40.8), (2.0%)	155.1 (40.1), (6.1%)
TbBMD Tibia (mg HA/ cm^3)	400 IU	105	176.4 (37.7), (0.0%)	177.8 (37.7), (0.0%)	177.6 (38.2), (1.0%)	177.4 (39.8), (4.8%)	179.1 (40.3), (4.8%)
	4000 IU	97	174.8 (35.2), (0.0%)	176.0 (36.0), (1.0%)	175.9 (36.5), (2.1%)	175.9 (38.0), (4.1%)	178.3 (39.1), (4.1%)
	10000 IU	101	171.9 (38.7), (0.0%)	172.7 (38.8), (0.0%)	173.0 (38.7), (1.0%)	172.9 (40.3), (2.0%)	175.2 (40.8), (5.0%)
TbN Radius (1/mm)	400 IU	104	1.4 (0.2), (0.0%)	1.4 (0.2), (1.9%)	1.4 (0.2), (1.0%)	1.4 (0.2), (5.8%)	1.4 (0.2), (4.8%)
	4000 IU	96	1.4 (0.2), (0.0%)	1.4 (0.2), (2.1%)	1.4 (0.2), (2.1%)	1.4 (0.2), (4.2%)	1.4 (0.2), (6.2%)
	10000 IU	99	1.4 (0.3), (0.0%)	1.4 (0.3), (2.0%)	1.4 (0.3), (2.0%)	1.4 (0.2), (2.0%)	1.4 (0.2), (6.1%)

		N	Baseline	6 Months	12 Months	24 Months	36 Months
TbN Tibia (1/mm)	400 IU	105	1.3 (0.2), (0.0%)	1.4 (0.2), (0.0%)	1.4 (0.2), (1.0%)	1.4 (0.2), (4.8%)	1.4 (0.2), (4.8%)
	4000 IU	97	1.3 (0.2), (0.0%)	1.4 (0.2), (1.0%)	1.4 (0.2), (2.1%)	1.4 (0.2), (4.1%)	1.4 (0.2), (4.1%)
	10000 IU	101	1.3 (0.2), (0.0%)	1.3 (0.2), (0.0%)	1.3 (0.3), (1.0%)	1.3 (0.2), (2.0%)	1.3 (0.3), (5.0%)
CtPo Radius (%)	400 IU	104	1.0 (0.6), (0.0%)	1.0 (0.6), (1.9%)	1.0 (0.6), (1.0%)	1.0 (0.6), (5.8%)	1.0 (0.6), (4.8%)
	4000 IU	96	0.9 (0.5), (0.0%)	0.9 (0.5), (2.1%)	1.0 (0.6), (2.1%)	1.0 (0.5), (4.2%)	1.0 (0.5), (6.2%)
	10000 IU	99	0.9 (0.6), (0.0%)	0.9 (0.6), (2.0%)	0.9 (0.6), (2.0%)	1.0 (0.6), (2.0%)	1.0 (0.6), (6.1%)
CtPo Tibia (%)	400 IU	105	2.9 (1.3), (0.0%)	2.9 (1.2), (0.0%)	3.0 (1.3), (1.0%)	2.9 (1.2), (4.8%)	2.9 (1.2), (4.8%)
	4000 IU	97	2.9 (1.2), (0.0%)	2.9 (1.2), (1.0%)	3.1 (1.3), (2.1%)	3.1 (1.4), (4.1%)	3.0 (1.4), (4.1%)
	10000 IU	101	2.8 (1.3), (0.0%)	2.8 (1.3), (0.0%)	3.0 (1.4), (1.0%)	3.0 (1.4), (2.0%)	3.0 (1.4), (5.0%)
TH aBMD (g/cm^2)	400 IU	105	1.02 (0.14), (0.95%)		1.02 (0.14), (1.90%)	1.02 (0.14), (4.76%)	1.02 (0.14), (5.71%)
	4000 IU	97	1.04 (0.14), (0.00%)		1.04 (0.15), (1.04%)	1.04 (0.15), (3.12%)	1.03 (0.15), (3.12%)
	10000 IU	101	1.01 (0.14), (0.00%)		1.01 (0.14), (0.99%)	1.01 (0.14), (1.98%)	1.00 (0.14), (5.94%)
TUG (sec)	400 IU	105	7.9 (1.6), (0.0%)		7.6 (1.4), (1.9%)	7.8 (1.4), (7.6%)	7.7 (1.5), (4.8%)
	4000 IU	97	7.6 (1.3), (0.0%)		7.7 (1.4), (1.0%)	7.9 (1.5), (3.1%)	8.0 (1.4), (3.1%)
	10000 IU	101	7.6 (1.3), (0.0%)		7.5 (1.4), (1.0%)	8.3 (5.1), (4.0%)	7.7 (1.4), (5.0%)

		N	Baseline	6 Months	12 Months	24 Months	36 Months
	400 IU	105	34.9 (9.8), (0.0%)		35.1 (10.7), (1.0%)	35.0 (11.0), (4.8%)	34.6 (11.1), (5.7%)
Grip Strength (kg)	4000 IU	97	34.8 (9.9), (0.0%)		34.4 (11.1), (1.0%)	35.1 (11.3), (3.1%)	34.1 (10.9), (3.1%)
	10000 IU	101	34.8 (10.8), (0.0%)		34.5 (12.3), (1.0%)	34.8 (12.4), (2.0%)	34.1 (11.7), (5.0%)

Data are presented as mean (SD), (% missing). CtBMD = cortical bone mineral density, TbBMD = trabecular bone mineral density, TbN = trabecular number, CtPo = cortical porosity, TH aBMD = total hip areal bone mineral density, TUG = timed-up-and-go

eTable 4. Summary of the statistical models

Variable	Coefficient	Estimate	95% CI LL	95% CI UL	LRS	df	P-value
Primary Outcome Variables							
Tt.BMD Radius (mg HA/ cm^3)	$\hat{\beta}_0$	330	323	337			
	$\hat{\beta}_1$	-0.037	-0.098	0.024			
	$\hat{\beta}_2$	-0.002	-0.003	-0.001	18.0	1	<0.001
	$\hat{\beta}_{31}$	-0.108	-0.179	-0.037			
	$\hat{\beta}_{32}$	-0.209	-0.279	-0.139	32.1	2	<0.001
Tt.BMD Tibia (mg HA/ cm^3)	$\hat{\beta}_0$	307	301	313			
	$\hat{\beta}_1$	0.034	-0.021	0.089			
	$\hat{\beta}_2$	-0.002	-0.003	-0.001	9.55	1	0.002
	$\hat{\beta}_{31}$	-0.049	-0.102	0.004			
	$\hat{\beta}_{32}$	-0.114	-0.166	-0.062	15.29	2	< 0.001
Failure Load Radius (N)	$\hat{\beta}_0$	2610	2497	2723			
	$\hat{\beta}_1$	-0.141	-2.322	2.040			
	$\hat{\beta}_2$	-0.006	-0.059	0.048	0.046	1	0.830
	$\hat{\beta}_{31}$	-1.349	-2.741	0.044			
	$\hat{\beta}_{32}$	-1.544	-2.923	-0.165	5.670	2	0.058
Failure Load Tibia (N)	$\hat{\beta}_0$	7679	7431	7927			
	$\hat{\beta}_1$	-6.799	-10.337	-3.260			
	$\hat{\beta}_2$	0.147	0.062	0.232	10.400	1	0.001
	$\hat{\beta}_{31}$	-0.574	-3.064	1.916			
	$\hat{\beta}_{32}$	-2.022	-4.484	0.439	4.200	2	0.123
Secondary Outcome Variables							
CtBMD Radius (mg HA/ cm^3)	$\hat{\beta}_0$	896.48	890.64	902.32			
	$\hat{\beta}_1$	-0.4593	-0.6000	-0.3186			
	$\hat{\beta}_2$	0.0067	0.0037	0.0096	18.96	1	<0.001
	$\hat{\beta}_{31}$	-0.1756	-0.3083	-0.0429			
	$\hat{\beta}_{32}$	-0.2382	-0.3695	-0.1068	13.13	2	0.001
CtBMD Tibia (mg HA/ cm^3)	$\hat{\beta}_0$	864.87	858.23	871.50			
	$\hat{\beta}_1$	-0.7073	-0.8639	-0.5506			
	$\hat{\beta}_2$	0.0138	0.0107	0.0169	70.41	1	<0.001
	$\hat{\beta}_{31}$	-0.1203	-0.2728	0.0323			
	$\hat{\beta}_{32}$	-0.2616	-0.4122	-0.1110	11.26	2	0.004
TbBMD Radius (mg HA/ cm^3)	$\hat{\beta}_0$	159.85	155.32	164.38			
	$\hat{\beta}_1$	0.1030	0.0551	0.1509			
	$\hat{\beta}_2$	-0.0014	-0.0024	-0.0004	8.13	1	0.017
	$\hat{\beta}_{31}$	-0.0264	-0.0724	0.0195			
	$\hat{\beta}_{32}$	-0.0753	-0.1208	-0.0298	10.93	2	0.012

Variable	Coefficient	Estimate	95% CI LL	95% CI UL	LRS	df	P-value
TbBMD Tibia (mg HA/ cm^3)	$\hat{\beta}_0$	174.64	170.47	178.80			
	$\hat{\beta}_1$	0.03302	-0.0093	0.0753			
	$\hat{\beta}_2$	0.00184	0.0010	0.0027	16.92	1	< 0.001
	$\hat{\beta}_{31}$	0.00562	-0.0314	0.0426			
	$\hat{\beta}_{32}$	-0.03016	-0.0667	0.0064	4.14	2	0.127
TbN Radius (1/mm)	$\hat{\beta}_0$	1.403	1.3776	1.428			
	$\hat{\beta}_1$	0.000	0.0000	0.001			0.905*
	$\hat{\beta}_{31}$	0.000	-0.0004	0.0005			
	$\hat{\beta}_{32}$	-0.001	-0.0010	-0.0001	9.22	2	0.010
TbN Tibia (1/mm)	$\hat{\beta}_0$	1.3237	1.3001	1.3473			
	$\hat{\beta}_1$	0.0025	0.0017	0.0034			
	$\hat{\beta}_2$	0.0000	-0.0001	0.0000	13.37	1	<0.001
	$\hat{\beta}_{31}$	-0.0002	-0.0007	0.0004			
	$\hat{\beta}_{32}$	-0.0005	-0.0011	0.0001	3.18	2	0.204
CtPo Radius (%)	$\hat{\beta}_0$	0.9122	0.8483	0.9762			
	$\hat{\beta}_1$	0.0041	0.0011	0.0071			
	$\hat{\beta}_2$	-0.0001	-0.0001	0.0000	4.11	1	0.043
	$\hat{\beta}_{31}$	0.0017	-0.0007	0.0041			
	$\hat{\beta}_{32}$	0.0012	-0.0012	0.0035	2.00	2	0.367
CtPo Tibia (%)	$\hat{\beta}_0$	2.8559	2.7156	2.9962			
	$\hat{\beta}_1$	0.0137	0.0075	0.2000			
	$\hat{\beta}_2$	-0.0003	-0.0004	-0.0001	15.19	1	<0.001
	$\hat{\beta}_{31}$	0.0003	-0.0046	0.0052			
	$\hat{\beta}_{32}$	0.0012	-0.0036	0.0061	0.26	2	0.877
TH aBMD (g/cm^2)	$\hat{\beta}_0$	1.02E+00	1.01E+00	1.04E+00			
	$\hat{\beta}_1$	1.46E-04	-9.76E-05	3.90E-04			
	$\hat{\beta}_2$	-1.04E-05	-1.61E-05	-4.82E-06	13.00	1	<0.001
	$\hat{\beta}_{31}$	1.97E-05	-1.78E-04	2.18E-04			
	$\hat{\beta}_{32}$	5.27E-05	-1.43E-04	2.49E-04	0.28	2	0.868
Balance (Sway index)	$\hat{\beta}_0$	2.30E+00	2.24E+00	2.36E+00			
	$\hat{\beta}_1$	-1.03E-02	-1.39E-02	-6.64E-03			
	$\hat{\beta}_2$	1.59E-04	7.04E-05	2.47E-04	12.47	1	<0.001
	$\hat{\beta}_{31}$	-1.25E-04	-2.57E-03	2.31E-03			
	$\hat{\beta}_{32}$	-2.31E-03	-4.73E-03	1.12E-04	4.30	2	0.116
Grip Strength (kg)	$\hat{\beta}_0$	34.66	33.50	35.82			
	$\hat{\beta}_1$	-0.013	-0.037	0.011			0.377*
	$\hat{\beta}_{31}$	-0.003	-0.038	0.031			
	$\hat{\beta}_{32}$	-0.007	-0.042	0.027	0.177	2.00	0.915

Variable	Coefficient	Estimate	95% CI LL	95% CI UL	LRS	df	P-value
SF-36 MCS	$\hat{\beta}_0$	55.510	54.846	56.175			0.701*
	$\hat{\beta}_1$	-0.007	-0.038	0.024			
	$\hat{\beta}_{31}$	0.019	-0.022	0.061			
	$\hat{\beta}_{32}$	0.020	-0.022	0.061	1.140	2.00	0.566
TUG (sec)	$\hat{\beta}_0$	7.717	7.517	7.917			
	$\hat{\beta}_1$	-0.002	-0.012	0.008			0.972*
	$\hat{\beta}_{31}$	0.010	-0.003	0.024			
	$\hat{\beta}_{32}$	0.008	-0.005	0.022	2.62	2	0.270

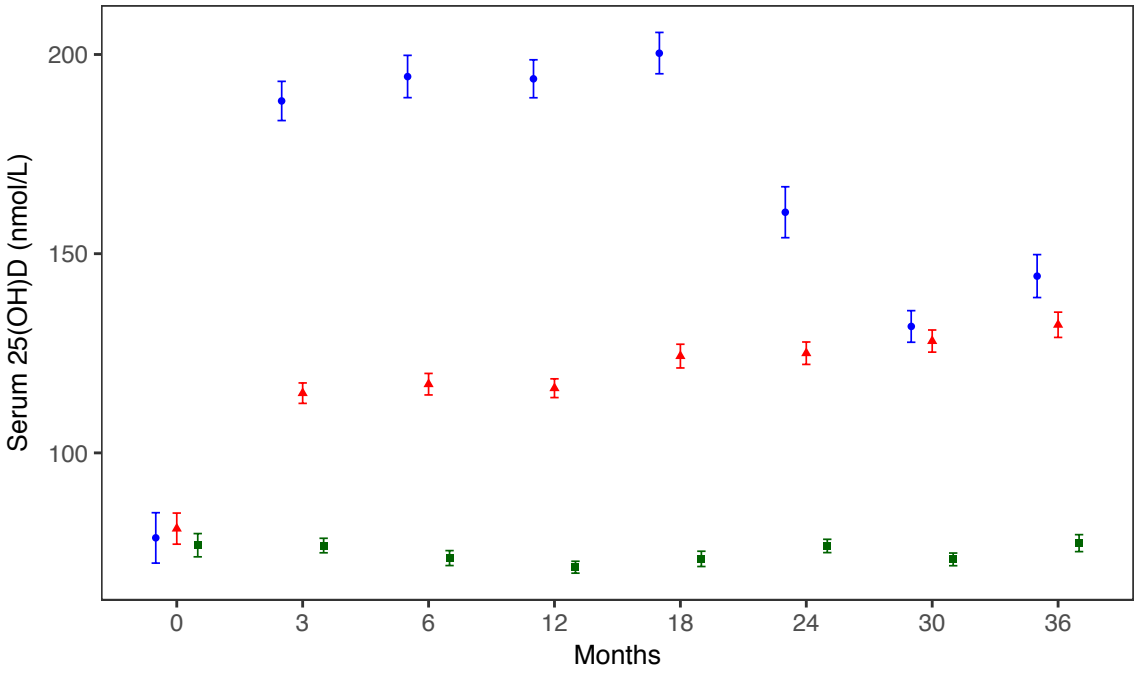
CI = confidence interval, LL = lower limits, UL = upper limits, df = degrees of freedom, LRS = likelihood ratio statistic, TtBMD = total bone mineral density, CtBMD = cortical bone mineral density, TbBMD = trabecular bone mineral density, TbN = trabecular number, CtPo = cortical porosity, TH aBMD = total hip areal bone mineral density, SF-36 = Short Form Health Survey questionnaire, MCS = mental component summary, TUG = timed up-and-go.

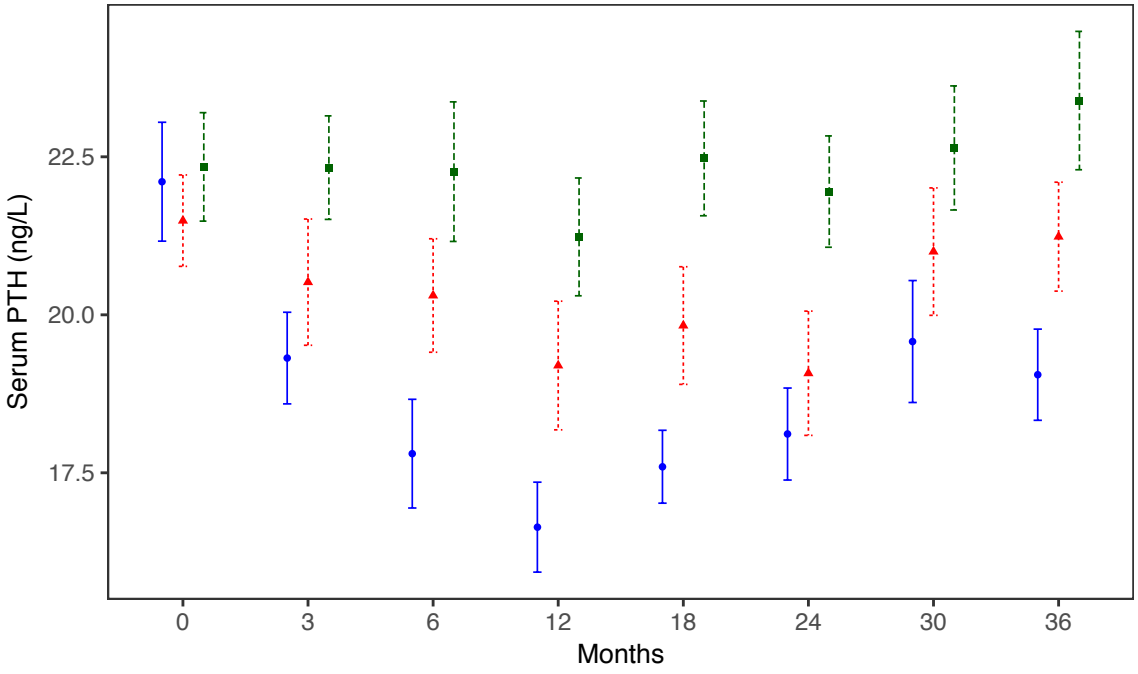
The Likelihood Ratio Statistic (LRS) on 1 degree of freedom (df) and accompanying p-value is the test for non-linearity. When the (non-significant) p-value is identified (*), this means that the non-linear was tested and then not included in the model.

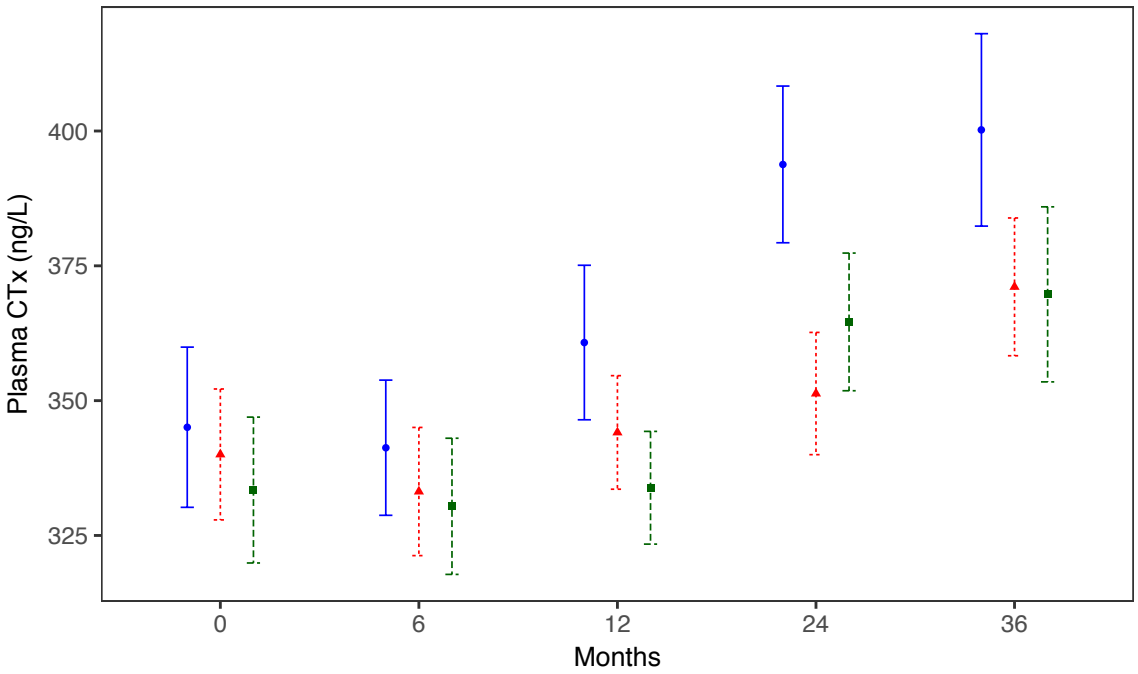
The Likelihood Ratio Statistic (LRS) on 2 degree of freedom (df) and accompanying p-value is the test for a significant interaction between time and treatment group. A significant p-value (< 0.05) indicates that the trajectory of the mean values of the treatment groups differ over time, resulting in a significant treatment effect. Interpretation of this treatment effect is shown using the manuscript figures.

The Intercept value is the mean value in all the three treatment groups at baseline. This is constrained to be the same under this model, since the groups have been formed by randomization.

eFigure 1. Change in serum 25-hydroxyvitamin D, parathyroid hormone and C-telopeptide of type 1 collagen during three years of vitamin D supplementation.



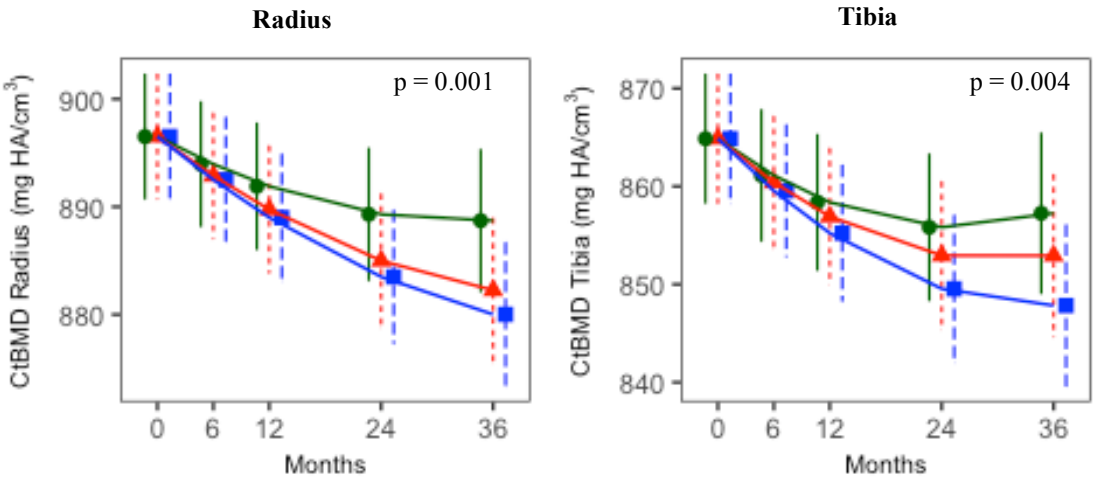




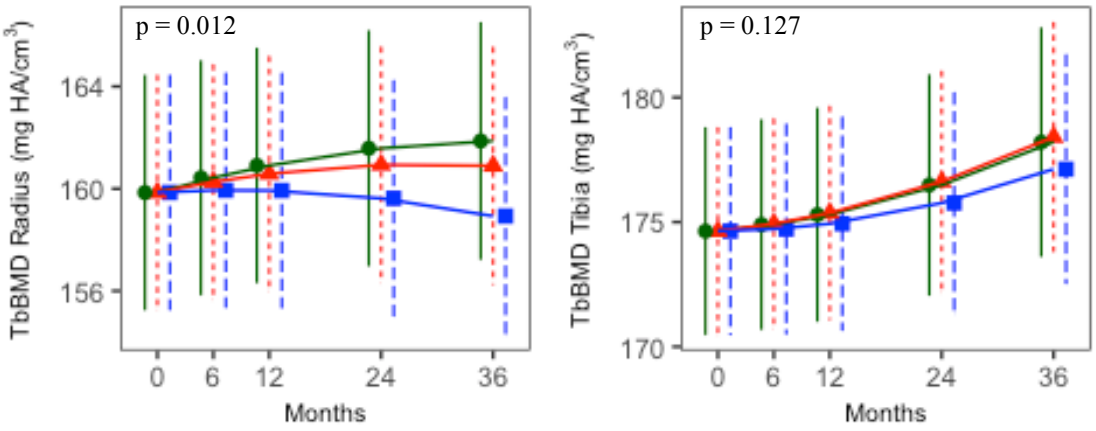
The modelled data show the mean and 95% confidence interval (taking into account the intra-participant correlation due to repeated measures) for the 400 IU group (green square), 4000 IU group (red triangle) and 10 000 IU group (blue circle). 25OHD = 25-hydroxyvitamin D, PTH = parathyroid hormone, CTx = C-telopeptide of type 1 collagen. For 25OHD, from month three onwards, between group differences exist at all months except for the 4000 and 10 000 IU groups at month 30.

.

eFigure 2. Change in cortical and trabecular bone mineral density during three years of vitamin D supplementation.



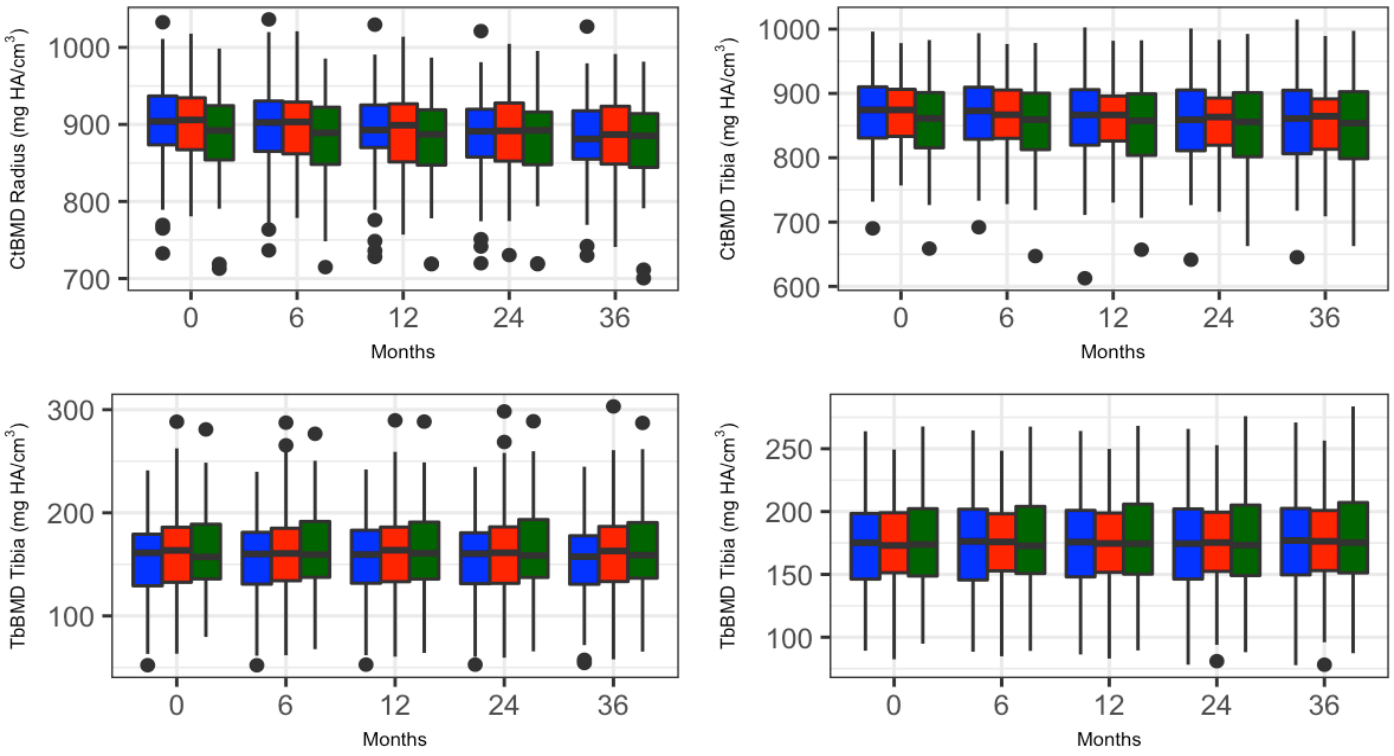
Month	0	6	12	24	36	Month	0	6	12	24	36
10000 IU	99	97	97	97	93	**10000 IU**	101	99	99	99	95
4000 IU	96	94	94	92	90	**4000 IU**	97	96	95	93	93
400 IU	104	102	103	98	99	**400 IU**	105	103	104	99	100



Month	0	6	12	24	36	Month	0	6	12	24	36
10000 IU	99	97	97	97	93	**10000 IU**	101	101	100	99	96
4000 IU	96	94	94	92	90	**4000 IU**	97	96	95	93	93
400 IU	104	102	103	98	99	**400 IU**	105	105	104	100	100

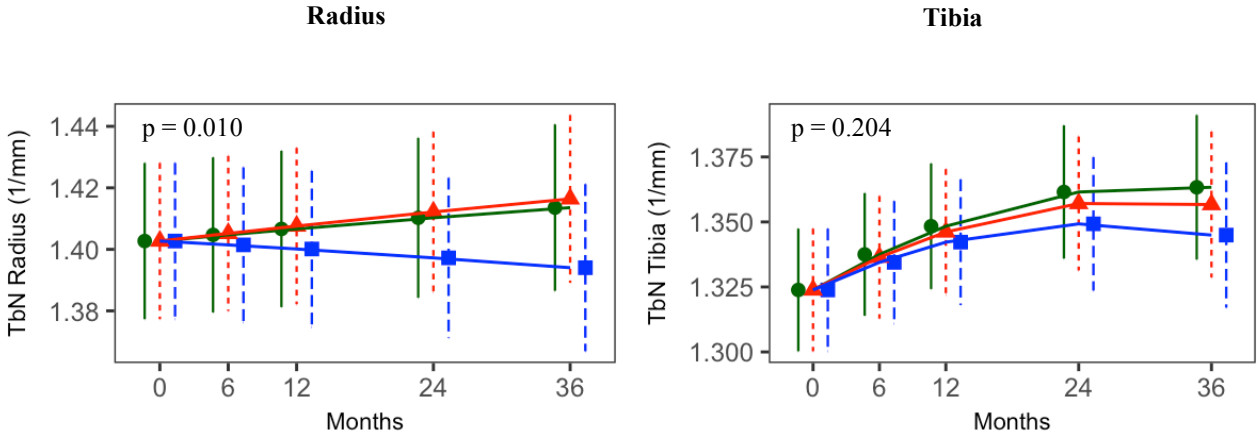
The modelled data show the mean and 95% confidence interval (for the predicted values using the random effects model) for the 400 IU group (green circle), 4000 IU group (red triangle) and 10 000 IU group (blue square). Plots of the radius are on the left and tibia on the right. CtBMD = cortical bone mineral density, TbBMD = trabecular bone mineral density. P-value represents the group by time interaction effect. The number of participants included in each model at the respective month is located under each figure.

eFigure 3. Data distribution for cortical and trabecular bone mineral density throughout the three-year study.



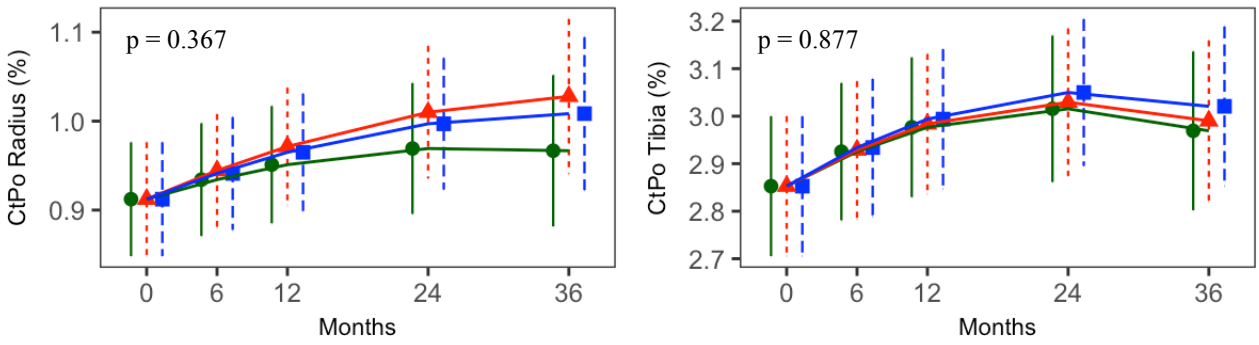
Boxplots showing the median, interquartile range and extreme values for the 10,000 IU group (blue), 4000 IU group (red) and 400 IU group (green). CtBMD = cortical bone mineral density; TbBMD = trabecular bone mineral density.

eFigure 4. Change in trabecular number and cortical porosity during three years of vitamin D supplementation



Month	0	6	12	24	36
10000 IU	99	97	97	97	93
4000 IU	96	94	94	92	90
400 IU	104	102	103	98	99

Month	0	6	12	24	36
10000 IU	101	101	100	99	96
4000 IU	97	96	95	93	93
400 IU	105	105	104	100	100

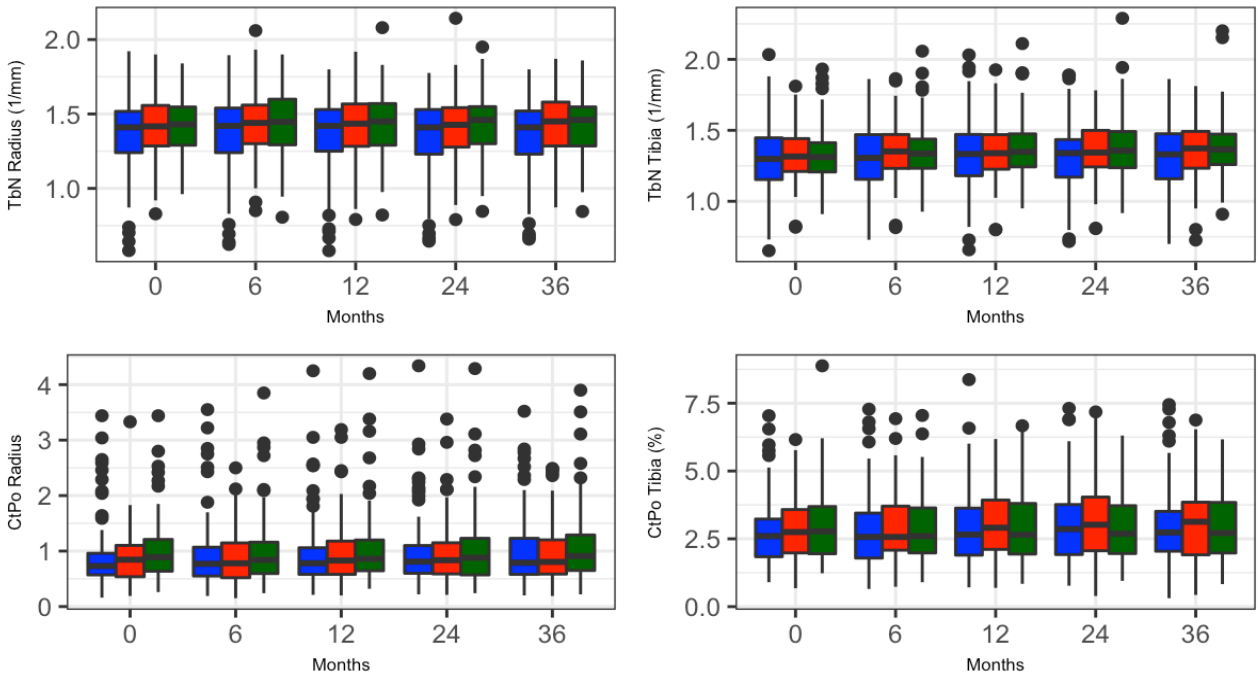


Month	0	6	12	24	36
10000 IU	99	97	97	97	93
4000 IU	96	94	94	92	90
400 IU	104	102	103	98	99

Month	0	6	12	24	36
10000 IU	101	101	100	99	96
4000 IU	97	96	95	93	93
400 IU	105	105	104	100	100

The modelled data show the mean and 95% confidence interval (for the predicted values using the random effects model) for the 400 IU group (green circle), 4000 IU group (red triangle) and 10 000 IU group (blue square). Plots of the radius are on the left and tibia on the right. TbN = trabecular number, CtPo = cortical porosity. P-value represents the group by time interaction effect. The number of participants included in each model at the respective month is located under each figure.

eFigure 5. Data distribution for trabecular number and cortical porosity throughout the three-year study.



Boxplots showing the median, interquartile range and extreme values for the 10,000 IU group (blue), 4000 IU group (red) and 400 IU group (green). TbN = trabecular number, CtPo = cortical porosity.

References

1. Coffman CJ, Edelman D, Woolson RF. To condition or not condition? Analysing "change" in longitudinal randomised controlled trials. *BMJ Open*. 2016;6(12):e013096. doi:10.1136/bmjopen-2016-013096.

2. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M. Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*. 2018;10:80-85. doi:10.1016/j.conctc.2018.03.008.