Determinants of plasma 25-hydroxyvitamin D and development of prediction models in three U.S. cohorts

Kimberly A. Bertrand[1,2], **Edward Giovannucci**[1,2,3], **Yan Liu**[3], **Susan Malspeis**[1], **A. Heather Eliassen**[2], **Kana Wu**[3], **Michelle D. Holmes**[1,2], **Francine Laden**[1,2,4], and **Diane Feskanich**[2]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA

[2]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA

[3]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA

[4]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA

Abstract

Epidemiologic and other evidence suggests that vitamin D may be protective against several chronic diseases. Assessing vitamin D status in epidemiologic studies, however, is challenging given finite resources and limitations of commonly used approaches. Using multivariable linear regression, we derived predicted 25-hydroxyvitamin D [25(OH)D] scores based on known determinants of circulating 25(OH)D, including age, race, ultraviolet-B radiation flux at residence, dietary and supplementary vitamin D intakes, body mass index, physical activity, alcohol intake, postmenopausal hormone use (women only), and season of blood draw, in three nationwide cohorts: the Nurses' Health Study (NHS), NHSII, and the Health Professionals Follow-up Study. The model r^2 for each cohort ranged from 0.25 to 0.33. We validated the prediction models in independent samples of participants from these studies. Mean measured 25(OH)D levels rose with increasing decile of predicted 25(OH)D score, such that differences in mean measured 25(OH)D between extreme deciles of predicted 25(OH)D were 8.7–12.3 ng/mL. Substituting predicted 25(OH)D scores for measured 25(OH)D in a previously published case-control analysis of colorectal cancer yielded similar effect estimates with odds ratios of approximately 0.8 for a 10-ng/mL difference in either plasma or predicted 25(OH)D. We conclude that these data provide reasonable evidence that a predicted 25(OH)D score is an acceptable marker for ranking individuals by long-term vitamin D status and may be particularly useful in research settings where biomarkers are not available for a majority of a study population.

Keywords

Epidemiologic evidence suggests that vitamin D may protect against colorectal, prostate, breast, and oropharyngeal cancers[1–6] and other chronic diseases such as cardiovascular disease[6, 9–9], diabetes[6, 9–11], and hip fractures[12–14]. Plasma 25-hydroxyvitamin D [25(OH)D], the primary circulating form of vitamin D[6, 15], is an accepted biomarker for measuring vitamin D status in clinical settings[16]; however, it is strongly dependent on season of blood draw. Although 25(OH)D is fairly reproducible over 2–3 years[17, 18], one

Corresponding author: Kimberly A. Bertrand Department of Epidemiology Harvard School of Public Health 677 Huntington Avenue Boston, MA 02115 USA telephone: (617) 525-2140 fax: (617) 525-2008 kbertran@hsph.harvard.edu.

measurement is weaker in characterizing longer term exposure[19, 20]. Furthermore, measuring 25(OH)D requires the availability of blood samples and monetary resources for laboratory assays, limiting the feasibility of this approach for many large-scale epidemiologic studies.

Individual determinants of vitamin D status, such as latitude and regional estimates of solar radiation[21–23] or dietary assessment[24–26], have been used as surrogates of vitamin D exposure, but each alone contributes a small proportion of 25(OH)D. An alternative approach to assess vitamin D status is to combine known determinants of circulating 25(OH)D to derive a predicted score from questionnaire data using measurements of plasma 25(OH)D available for a subset of the study population; reported r^2 from such predictive models have ranged from 0.21 to 0.42[18, 27–29]. Although the r^2 between predicted and measured 25(OH)D has been used to assess the "validity" of the predicted 25(OH)D approach, the r^2 in this context has limitations given that a single measure is not a true gold standard of long-term average 25(OH)D concentration. Such an "alloyed" or imperfect gold standard will underestimate true "validity"[30–32]. More importantly, a comparison of high versus low circulating 25(OH)D level in the population, which may be estimated by high and low predicted 25(OH)D, may be the more relevant factor for testing exposure-disease hypotheses in epidemiologic studies.

In this paper, we describe the development and validation of regression models to predict 25(OH)D based on determinants of vitamin D status in three cohorts: the Nurses' Health Study (NHS), the Nurses' Health Study II (NHSII), and the Health Professionals Follow-Up Study (HPFS). Predicted 25(OH)D scores have been used in several analyses within these cohorts[8, 18, 33, 34] but, with the exception of HPFS[18], no formal validation has been conducted previously and the specific prediction models varied with each analysis. We also evaluated reproducibility of plasma 25(OH)D over 10–11 years in NHS participants.

METHODS

Study population

Participants were selected from three U.S. prospective cohort studies. NHS was established in 1976, when 121 700 female nurses aged 30–55 years completed a self-administered questionnaire on risk factors for cancer and other diseases. From 1989–1990, 32 826 participants provided blood samples for analysis. NHSII began in 1989 when 116 671 female nurses aged 25–42 years completed and returned a baseline questionnaire. Between 1996 and 1999, 29 611 participants (aged 32–54 years) provided blood samples. HPFS comprises 51 529 male dentists, optometrists, osteopaths, podiatrists, pharmacists, and veterinarians aged 40–75 years at baseline in 1986. Blood samples were provided by 18 225 of these men between 1993 and 1994. Blood samples have been stored in liquid nitrogen freezers (−130°C) since collection. For all three cohorts, biennial questionnaires are sent to participants to update information on risk factors and to identify newly diagnosed diseases. Diet is assessed by a validated semiquantitative food frequency questionnaire approximately every 4 years[35–38].

Plasma 25(OH)D measurements were available from men and women who served as controls in previous nested case-control studies of chronic diseases. None of the participants had a history of cancer at the time of blood draw. For each cohort, we selected two independent samples: a "training" sample was used to develop the 25(OH)D prediction model and a "test" sample served as a validation dataset. Training samples comprised controls from all completed and on-going nested case-control studies with 25(OH)D assay results when analyses began. Test samples were drawn from more recently established nested case-control studies as this project unfolded and additional plasma 25(OH)D assay

results became available. Prior to exclusions for missing data, the training sets consisted of 2246 women in NHS, 1646 women in NHSII, and 1255 men in HPFS. An additional 818 women in NHS, 479 women in NHSII, and 841 men in HPFS were available for the test sets.

In 2000 and 2001, all women in NHS who gave blood in 1989–1990 and were alive were invited to provide a second blood sample. Of the 18 473 women who participated in the second blood collection, 443 women with no history of cancer had measured 25(OH)D available at both time points. These samples were used to assess within-person variability of plasma 25(OH)D concentrations over 10–11 years.

This study was approved by the institutional review boards of the Harvard School of Public Health and Brigham and Women's Hospital. All participants gave written informed consent at enrollment.

Laboratory analyses

Plasma 25(OH)D levels were determined by radioimmunoassay or chemiluminescence immunoassay, as previously described[39–41], between 1993 and 2010. The time between blood collection and 25(OH)D assay ranged from 3 to 20 years, with the majority of samples assayed within 14 years of blood collection. The stability of 25(OH)D in frozen plasma has been previously demonstrated, even for samples stored >10 years[42]. Intra-assay coefficients of variation (CV) from blinded, replicate, quality control samples were <15% for 23 of 26 laboratory batches; the highest CV was 17.6%. Mean (standard deviation) 25(OH)D concentrations (ng/mL) in training samples were: 28.5 (10.9) (NHS, $n = 2079$); 26.3 (9.8) (NHSII, $n = 1497$); and 25.9 (10.0) (HPFS, $n = 911$).

Statistical analyses

Using the training sample for each cohort, we fit a linear regression model to predict measured plasma 25(OH)D (continuous ng/mL) based on known or suspected determinants[18]. Age (years), season of blood draw, and laboratory batch were included as independent variables in all models to account for known extraneous variation. Other candidate predictor variables were energy-adjusted[43] vitamin D intake from food, vitamin D intake from supplements, average annual UV-B flux – a composite measure of mean UV-B radiation level reaching the earth's surface that takes into account factors such as latitude, altitude, and cloud cover – based on state of residence[44], race/ethnicity, body mass index (BMI), leisure-time physical activity level, alcohol intake, geographic region of residence (North, South, Midwest, West), smoking history, hair color, susceptibility to burn, ability to tan, and number of lifetime sunburns. Menopausal status, postmenopausal hormone (PMH) use, and age at first birth were also considered for women in NHS and NHSII. Data were obtained from questionnaires completed closest to blood draw date. Questionnaires were completed within ±2 years of blood draw for 97% of each sample; the median time period was 5 months prior to blood draw for NHS, 3 months after blood draw for NHS2, and 2 months prior to blood draw for HPFS.

For each cohort, we first fit a multivariable linear regression model with all candidate predictors with $P < 0.05$ in univariate analyses adjusted for laboratory batch and age. Then, we eliminated non-significant (P 0.05) variables from the model one at a time, based on largest P-value. The final multivariable prediction model includes all statistically significant predictors, plus age, season of blood draw, and laboratory batch. The HPFS model is a refinement of one previously published[18]. The general form of the prediction model is: $25(OH)D = \beta_0 + \beta'X'$ where β_0 represents the intercept and β' represents the vector of coefficients associated with the vector of predictors, X' (see Table 1).

Regarding final sets of predictors, we aimed for consistency between cohorts, while allowing for flexibility in cohort-specific models. Factors statistically significant for one cohort were considered for inclusion in models for the other cohorts regardless of statistical significance, given sufficient biological plausibility (e.g., UV-B flux). We excluded individuals with missing values for predictors except PMH use in NHSII for which a missing category was created. The final prediction models were fit to 2079 women ages 42–69 years in NHS, 1497 women ages 32–52 years in NHSII, and 911 men ages 46–81 years in HPFS.

Based on the regression coefficients for each variable in the prediction model, we calculated a predicted 25(OH)D score for each individual in the test samples using personal data for covariates. Age, season of blood draw, and laboratory batch were not used in derivation of predicted 25(OH)D scores. Age is not used in the derivation of predicted 25(OH)D because it is a strong risk factor for many chronic diseases. By excluding age from the derived score, the ability to control finely for potential confounding by age in epidemiologic investigations is retained. Predicted 25(OH)D scores were not calculated if predictor data were missing on the questionnaire closest to blood draw or the previous questionnaire (NHS, $n = 39$; NHSII, $n = 34$; HPFS, $n = 5$). For the test samples, there were 779 women in NHS, 445 women in NHSII, and 836 men in HPFS with available 25(OH)D measurements and predicted 25(OH)D scores.

For validation, we compared predicted 25(OH)D and actual plasma 25(OH)D measurements in test samples. Laboratory batch-adjusted Spearman correlation coefficients were calculated to assess agreement between predicted score and actual 25(OH)D levels. We examined actual plasma 25(OH)D measurements according to decile of predicted 25(OH)D score[18, 28] and cross-classified individuals by quintile of both predicted and actual 25(OH)D. Using previously published data from a nested case-control study that examined the association between plasma 25 (OH)D and colorectal cancer in NHS and HPFS[4], we calculated odds ratios for colorectal cancer for a 10-ng/mL difference in measured 25(OH)D and then compared results to analyses that used the predicted 25(OH)D score. In these analyses, we derived separate predicted scores at each questionnaire year based on current predictor data and calculated the average predicted 25(OH)D from 1986 - the year predicted scores were first derived - to date of diagnosis (or matched date for controls) as the main exposure variable. For both measured and predicted 25(OH)D, pooled estimates were calculated for NHS and HPFS using a meta-analysis approach described by DerSimonian and Laird[45].

Finally, we evaluated the reproducibility of 25(OH)D measurements over 10–11 years among 443 women in NHS with two blood measures, using a statistical approach previously described[17]. We calculated intraclass correlation coefficients (ICCs) by dividing the between-person variance by the sum of the within- and between-person variances; a 95% confidence interval (CI) also was calculated. Using a mixed model, we adjusted for age (continuous) by including it as a fixed effect. ICC measures the fraction of total variation that is due to between-person variability. A high value for the ICC reflects a low within-person variation.

Among NHS participants with two 25(OH)D measurements 10–11 years apart, we compared average plasma 25(OH)D concentration to average predicted 25(OH)D score over the same time period. We calculated Spearman correlation coefficients based on the residuals of plasma 25(OH)D measurements in each time period from a linear regression model to factor out effects of age and season of blood draw. Because random within-person error can attenuate correlations, we used data from the reproducibility sample to correct for these effects[46, 47].

All statistical tests were two-sided and analyses were performed using SAS version 9 for UNIX (SAS Institute Inc., Cary, NC).

RESULTS

Using multivariable linear regression in the training set within each cohort, we identified the following independent predictors of age-adjusted plasma 25(OH)D levels: race, UVB flux (NHS and HPFS only), dietary vitamin D intake, supplementary vitamin D intake, BMI, physical activity, alcohol intake (NHS and NHSII only), PMH use (NHS only), and season of blood draw (Table 1). Overall, the predictive models explained 25% (NHSII), 28% (HPFS), and 33% (NHS) of the total variability in plasma 25(OH)D concentration. The strongest predictors of circulating 25(OH)D generally were race (a proxy for skin pigmentation) and BMI, followed by physical activity, dietary and supplementary vitamin D intake, and UV-B flux (NHS and HPFS only). Season also was an important predictor of 25(OH)D but is not used in the calculation of predicted 25(OH)D score because it reflects time of blood draw and is not a factor in determining long-term average between-person variation in 25(OH)D. Age was not a significant independent predictor of 25(OH)D in NHS or HPFS, but a modest inverse association was observed in NHSII.

Using the regression coefficients estimated in each training set, we calculated predicted 25(OH)D scores for participants in the corresponding test samples. The batch-adjusted Spearman correlation coefficients between predicted score and actual 25(OH)D level were 0.23 (95% CI: 0.16, 0.29) for NHS, 0.40 (95% CI: 0.32, 0.47) for NHSII, and 0.24 (95% CI: 0.18, 0.30) for HPFS (all *P*-values < 0.0001). After further adjusting for age and season of blood draw, correlations were 0.23 (95% CI: 0.16, 0.29) (NHS), 0.42 (95% CI: 0.34, 0.49) (NHSII), and 0.30 (95% CI: 0.21, 0.37) (HPFS). In all cohorts, actual plasma 25(OH)D levels generally rose with increasing decile of predicted 25(OH)D score (Figure 1). The differences in mean actual 25(OH)D level between extreme deciles of predicted 25(OH)D score were 8.7 ng/mL (95% CI: 5.4, 11.9) for NHS, 12.3 ng/mL (95% CI: 8.7, 16.0) for NHSII, and 8.7 ng/mL (95% CI: 5.5, 11.8) for HPFS.

Because epidemiologic studies often categorize exposures into quantiles for analysis, we cross-classified individuals in the validation samples by quintile of predicted 25(OH)D and measured plasma 25(OH)D levels to determine how well the predicted score performed in ranking individuals with respect to plasma levels. Between 24.8% (NHS) and 29.9% (NHSII) of individuals fell into identical quintiles of predicted and measured 25(OH)D. Using the predicted scores, the majority of individuals were classified in either the same quintile or the adjacent quintile of actual plasma 25(OH)D concentration (NHS: 59.8%, NHSII: 66.5%, HPFS: 61.4%) (Figure 2). Only 5% or less of participants in each cohort were in extreme opposite quintiles according to predicted and actual 25(OH)D. Among women in the lowest quintile (Q1) of actual plasma 25(OH)D in NHS, 33% were categorized in Q1 of the predicted score, 57% were categorized in either Q1 or Q2, and 13% were categorized in Q5. Among women in Q1 of actual plasma 25(OH)D in NHS2, 44% were categorized in Q1 of the predicted score, 66% were categorized in either Q1 or Q2, and 8% were categorized in Q5. Among men in Q1 of actual plasma 25(OH)D in HPFS, 37% were categorized in Q1 of the predicted score, 57% were categorized in either Q1 or Q2, and 10% were categorized in Q5.

Based on data from a previously published case-control study of colorectal cancer in the NHS and HPFS[4], the pooled multivariable odds ratio for a 10-ng/mL difference in measured 25(OH)D was 0.82 (95% CI: 0.66, 1.03). Using the average predicted 25(OH)D score in these analyses yielded an odds ratio of 0.78 (95% CI: 0.41, 1.48).

In our reproducibility substudy in NHS, the ICC for plasma 25(OH)D measured over 10–11 years was 0.50 (95% CI: 0.43, 0.57). Among these 443 women, the age- and season-adjusted Spearman correlation coefficient between average measured 25(OH)D based on two blood samples and long-term average predicted 25(OH)D over the same time period was 0.23. We corrected for within-person variation in plasma 25(OH)D to obtain a deattenuated correlation coefficient of 0.28.

DISCUSSION

Using data from three U.S. cohorts, we derived predicted 25(OH)D scores based on various factors that influence circulating levels. The determinants of circulating 25(OH)D we identified generally were consistent with predictors reported by others[27–29, 48–55]. The set of predictors included in the final models explained only a proportion of the total variability in plasma 25(OH)D levels (i.e., 25–33%). The r^2 for our prediction models were generally consistent with previously published models[27–29]. Millen et al. reported a similar multivariable regression model with a comparable r^2 (0.21) and correlation between predicted and actual 25(OH)D (0.45) for the Women's Health Initiative[28]. In the Framingham Offspring Study, Liu et al. developed a model to predict a 25(OH)D score based on a similar set of predictors ($r^2 = 0.26$) and in their validation study observed a correlation of 0.51 between predicted and actual levels[27]. In the Adventist Health Study-2, Chan et al. reported r^2 of 0.22 and 0.33 for white and black populations, respectively (0.42 combined); however, they did not compare predicted and actual 25(OH)D levels in an independent sample[29]. Because only a small proportion of the total variability in plasma 25(OH)D levels is explained by identified predictors, predicted 25(OH)D scores cannot be interpreted as direct blood measurements of 25(OH)D to determine an individual's vitamin D sufficiency, insufficiency, or deficiency status.

Vitamin D prediction models have potential strengths and limitations as exposure assessment tools. Our models and others' have substantial unexplained variability, which likely can be attributed to error in the measurement of predictor variables and plasma 25(OH)D levels and lack of information about other important determinants of vitamin D status such as genetic factors[49, 56] and actual UV exposure. While sun sensitivity characteristics (e.g., ability to tan, susceptibility to burn, and number of lifetime sunburns) were not predictive of 25(OH)D in NHS, data on personal sun exposure and sun behaviors (such as time spent outdoors and use of sunscreen or protective clothing), important determinants of circulating 25(OH)D, were not regularly collected in these cohorts. We examined leisure-time physical activity as a proxy for time spent outdoors and found this to be a significant predictor. The prediction models also include an estimate of average annual UV-B flux, a composite measure of mean UV-B radiation level based on latitude, altitude, and cloud cover, which also was a significant determinant of circulating 25(OH)D.

Millen et al. concluded that predicted 25(OH)D scores "do not adequately reflect serum 25(OH)D concentrations"[28]. While we agree that predicted scores cannot substitute for blood measures in assessing current 25(OH)D level, we view the results of both studies as providing reasonable evidence that predicted 25(OH)D score is an acceptable marker of vitamin D status for the purposes of distinguishing a substantial range of vitamin D exposure in a given study population. In chronic disease epidemiology, the actual contrast between high and low exposure level over years or decades is particularly relevant. We calculated differences in measured 25(OH)D between extreme deciles of predicted score of 9 – 12 ng/mL, which represents the actual contrast in long-term 25(OH)D that can be studied in these populations. This difference corresponds to differences in vitamin D intakes of approximately 1000–1500 IU/day[57], and is considerably larger than what may be estimated using single surrogates of vitamin D exposure, such as dietary vitamin D intake, which

explains a contrast of approximately 3 ng/mL in 25(OH)D between high and low intake (Table 1).

A single blood measurement of 25(OH)D has the advantage of being a direct measure of circulating 25(OH)D; however, it is substantially influenced by recent and acute exposures (e.g., beach vacation, season), which contributes to measurement error in estimating long-term 25(OH)D. Correlations between two direct 25(OH)D measures taken 2–14 years apart range from 0.42–0.72[17–20], reflecting that a single 25(OH)D measurement is not a true gold standard of long-term 25(OH)D level. In NHS, the ICC for plasma 25(OH)D measured 2–3 years apart was 0.72[17]; over 10–11 years, the ICC was 0.50. While an ICC of 0.50 indicates fair to good reproducibility of a biomarker[58], the difference between the 2–3 year ICC and the 10–11 year ICC reflects lower reproducibility over a longer time period. Therefore, in our analyses and those by Millen et al.[28] and Liu et al.[27], correlation coefficients comparing predicted and actual 25(OH)D are likely underestimated because measured 25(OH)D is not a true gold standard and because random within-person error in the measurement of both variables attenuates correlation coefficients[32]. Because circulating plasma 25(OH)D is an imperfect measure of long-term 25(OH)D status, the comparison of mean actual 25(OH)D level by category of predicted score in validation analyses may better reflect the utility of predicted 25(OH)D scores to assess long-term 25(OH)D status. Although we assumed that the average of two plasma 25(OH)D measurements taken 10–11 years apart would better represent long-term 25(OH)D status, correlation coefficients were similar in the NHS sample with repeated measurements.

Another objection commonly raised about the predicted 25(OH)D score is that it may be confounded by its predictors (e.g., physical activity or BMI), which could be independent risk factors for disease[28]. This criticism would also be true for plasma 25(OH)D levels, which inherently incorporate these factors. Importantly, including predictors of vitamin D status as covariates in multivariable models may represent overadjustment because these variables are important determinants of 25(OH)D. Therefore, adjusting for these factors may be inappropriate. A potential advantage of using predicted 25(OH)D scores over measured 25(OH)D in analytic epidemiology is that a sensitivity analysis could be performed in which physical activity (or other predictor) is excluded from the score, thereby removing potential confounding by this factor. In practice, however, we did not observe evidence of confounding of predicted 25(OH)D by BMI or physical activity in previous analyses of colorectal cancer risk in HPFS[18].

Predicted scores were derived based on data not collected for assessing vitamin D status; the predictive ability of derived scores would likely improve if additional determinants of circulating 25(OH)D, such as personal sun exposure behaviors, were incorporated. Random measurement error in predicted 25(OH)D is expected to attenuate measures of association with disease[32, 59]; however, predicted scores should allow investigators to test a sizeable contrast in 25(OH)D between "low" and "high" exposure categories and will still be useful to detect moderate to strong vitamin D-disease associations. In our cohorts, we observe similar associations for various disease endpoints using plasma 25(OH)D and predicted 25(OH)D as the exposure variable, including hypertension[8], colorectal cancer incidence[18] and survival[34, 60], pancreatic cancer incidence[33] (unpublished data), and prostate cancer incidence[18] (unpublished data). For example, although statistical power was reduced when we used average predicted 25(OH)D, we observed similar odds ratios of approximately 0.8 for a 10-ng/mL difference in either plasma or predicted 25(OH)D for colorectal cancer based on data from a previous case-control study in the NHS and HPFS[4]. In a much larger HPFS cohort analysis with 691 colorectal cancer cases, the relative risk for the same increment of predicted 25(OH)D was 0.63 (95% CI: 0.48, 0.83)[18], demonstrating that the loss in precision may be recovered by increasing sample size in analyses using predicted scores.

For analyses of vitamin D and chronic diseases in these specific cohorts, predicted 25(OH)D scores can be derived for each participant at each questionnaire cycle. An advantage of longitudinal data is the availability of updated predictor information, allowing the predicted 25(OH)D score to change over time and potentially better capturing long-term average vitamin D status. Such studies would use data available from the full cohorts and complement biomarker analyses with smaller sample sizes. As noted by others[29], prediction models developed in NHS, NHSII, and HPFS may not apply to other study populations because of underlying population differences and/or availability of data; however, similar models may be developed using the general approach described here and could be useful for investigating hypothesized associations between vitamin D status and disease. It is also possible, however, that the prediction models developed in these cohorts could perform well in populations with similar demographics (e.g., male and female populations with similar age, race, and residential latitude distributions as the NHS, NHSII and HPFS); such applications would benefit from additional validation. In conclusion, predicted 25(OH)D scores may be a practical alternative for studying such associations in international and other settings where large-scale biomarker studies are not feasible.

Acknowledgments

REFERENCES

1. Garland CF, Gorham ED, Mohr SB, et al. Vitamin D and prevention of breast cancer: pooled analysis. J Steroid Biochem Mol Biol. 2007; 103:708–711. [PubMed: 17368188]

2. Giovannucci E. The epidemiology of vitamin D and cancer incidence and mortality: a review (United States). Cancer Causes Control. 2005; 16:83–95. [PubMed: 15868450]

3. Li H, Stampfer MJ, Hollis JB, et al. A prospective study of plasma vitamin D metabolites, vitamin D receptor polymorphisms, and prostate cancer. PLoS Med. 2007; 4:e103. [PubMed: 17388667]

4. Wu K, Feskanich D, Fuchs CS, et al. A nested case control study of plasma 25-hydroxyvitamin D concentrations and risk of colorectal cancer. J Natl Cancer Inst. 2007; 99:1120–1129. [PubMed: 17623801]

5. IARC. IARC Working Group Reports Vol. 5. Vol. Vol. 5. International Agency for Research on Cancer; Lyon: 2008. Vitamin D and Cancer. IARC Working Group Reports.

6. Holick MF. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. Am J Clin Nutr. 2004; 80:1678S–1688S. [PubMed: 15585788]

7. Giovannucci E, Liu Y, Hollis BW, et al. 25-hydroxyvitamin D and risk of myocardial infarction in men: a prospective study. Arch Intern Med. 2008; 168:1174–1180. [PubMed: 18541825]

8. Forman JP, Giovannucci E, Holmes MD, et al. Plasma 25-hydroxyvitamin D levels and risk of incident hypertension. Hypertension. 2007; 49:1063–1069. [PubMed: 17372031]

9. Michos ED, Melamed ML. Vitamin D and cardiovascular disease risk. Curr Opin Clin Nutr Metab Care. 2008; 11:7–12. [PubMed: 18090651]

10. Pittas AG, Lau J, Hu FB, et al. The role of vitamin D and calcium in type 2 diabetes. A systematic review and meta-analysis. J Clin Endocrinol Metab. 2007; 92:2017–2029. [PubMed: 17389701]

11. Pittas AG, Chung M, Trikalinos T, et al. Systematic review: Vitamin D and cardiometabolic outcomes. Ann Intern Med. 2010; 152:307–314. [PubMed: 20194237]

12. Feskanich D, Willett WC, Colditz GA. Calcium, vitamin D, milk consumption, and hip fractures: a prospective study among postmenopausal women. Am J Clin Nutr. 2003; 77:504–511. [PubMed: 12540414]

13. Bischoff-Ferrari HA, Willett WC, Wong JB, et al. Fracture prevention with vitamin D supplementation: a meta-analysis of randomized controlled trials. JAMA. 2005; 293:2257–2264. [PubMed: 15886381]

14. Cauley JA, Lacroix AZ, Wu L, et al. Serum 25-hydroxyvitamin D concentrations and risk for hip fractures. Ann Intern Med. 2008; 149:242–250. [PubMed: 18711154]

15. Holick MF. Vitamin D deficiency. N Engl J Med. 2007; 357:266–281. [PubMed: 17634462]

16. Horst, R.; Reinhardt, T.; Reddy, G. Vitamin D metabolism. In: Feldman, D.; Pike, J.; Glorieux, F., editors. Vitamin D. Elsevier Academic Press; Burlington, MA: 2005. p. 15-36.

17. Kotsopoulos J, Tworoger SS, Campos H, et al. Reproducibility of plasma and urine biomarkers among premenopausal and postmenopausal women from the Nurses' Health Studies. Cancer Epidemiol Biomarkers Prev. 2010; 19:938–946. [PubMed: 20332276]

18. Giovannucci E, Liu Y, Rimm EB, et al. Prospective study of predictors of vitamin D status and cancer incidence and mortality in men. J Natl Cancer Inst. 2006; 98:451–459. [PubMed: 16595781]

19. Hofmann JN, Yu K, Horst RL, et al. Long-term variation in serum 25-hydroxyvitamin D concentration among participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Cancer Epidemiol Biomarkers Prev. 2010; 19:927–931. [PubMed: 20332255]

20. Jorde R, Sneve M, Hutchinson M, et al. Tracking of serum 25-hydroxyvitamin D levels during 14 years in a population-based study and during 12 months in an intervention study. Am J Epidemiol. 2010; 171:903–908. [PubMed: 20219763]

21. Garland CF, Garland FC. Do sunlight and vitamin D reduce the likelihood of colon cancer? Int J Epidemiol. 1980; 9:227–231. [PubMed: 7440046]

22. Garland FC, Garland CF, Gorham ED, et al. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. Prev Med. 1990; 19:614–622. [PubMed: 2263572]

23. Grant WB, Mohr SB. Ecological studies of ultraviolet B, vitamin D and cancer since 2000. Ann Epidemiol. 2009; 19:446–454. [PubMed: 19269856]

24. Anderson LN, Cotterchio M, Vieth R, et al. Vitamin D and calcium intakes and breast cancer risk in pre- and postmenopausal women. Am J Clin Nutr. 2010; 91:1699–1707. [PubMed: 20392891]

25. Martinez ME, Giovannucci EL, Colditz GA, et al. Calcium, vitamin D, and the occurrence of colorectal cancer among women. J Natl Cancer Inst. 1996; 88:1375–1382. [PubMed: 8827015]

26. Oh K, Willett WC, Wu K, et al. Calcium and vitamin D intakes in relation to risk of distal colorectal adenoma in women. Am J Epidemiol. 2007; 165:1178–1186. [PubMed: 17379616]

27. Liu E, Meigs JB, Pittas AG, et al. Predicted 25-hydroxyvitamin D score and incident type 2 diabetes in the Framingham Offspring Study. Am J Clin Nutr. 2010; 91:1627–1633. [PubMed: 20392893]

28. Millen AE, Wactawski-Wende J, Pettinger M, et al. Predictors of serum 25-hydroxyvitamin D concentrations among postmenopausal women: the Women's Health Initiative Calcium plus Vitamin D clinical trial. Am J Clin Nutr. 2010; 91:1324–1335. [PubMed: 20219959]

29. Chan J, Jaceldo-Siegl K, Fraser GE. Determinants of serum 25 hydroxyvitamin D levels in a nationwide cohort of blacks and non-Hispanic whites. Cancer Causes Control. 2010; 21:501–511. [PubMed: 20012182]

30. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. Am J Epidemiol. 1993; 137:1251–1258. [PubMed: 8322765]

31. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an "alloyed gold standard". Am J Epidemiol. 1997; 145:184–196. [PubMed: 9006315]

32. Willett, W. Nutritional Epidemiology. 2nd Edition. Oxford University Press; New York: 1998.

33. Bao Y, Ng K, Wolpin BM, et al. Predicted vitamin D status and pancreatic cancer risk in two prospective cohort studies. Br J Cancer. 2010; 102:1422–1427. [PubMed: 20389298]

34. Ng K, Wolpin BM, Meyerhardt JA, et al. Prospective study of predictors of vitamin D status and survival in patients with colorectal cancer. Br J Cancer. 2009; 101:916–923. [PubMed: 19690551]

35. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol. 1985; 122:51–65. [PubMed: 4014201]

36. Willett WC, Sampson L, Browne ML, et al. The use of a self-administered questionnaire to assess diet four years in the past. Am J Epidemiol. 1988; 127:188–199. [PubMed: 3337073]

37. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol. 1992; 135:1114–1126. discussion 1127–1136. [PubMed: 1632423]

38. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. J Am Diet Assoc. 1993; 93:790–796. [PubMed: 8320406]

39. Hollis BW. Quantitation of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D by radioimmunoassay using radioiodinated tracers. Methods Enzymol. 1997; 282:174–186. [PubMed: 9330287]

40. Gallicchio L, Helzlsouer KJ, Chow WH, et al. Circulating 25-Hydroxyvitamin D and the Risk of Rarer Cancers: Design and Methods of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. Am J Epidemiol. 2010; 172:10–20. [PubMed: 20562188]

41. Wagner D, Hanwell HE, Vieth R. An evaluation of automated methods for measurement of serum 25-hydroxyvitamin D. Clin Biochem. 2009; 42:1549–1556. [PubMed: 19631201]

42. Hollis BW. Measuring 25-hydroxyvitamin D in a clinical environment: challenges and needs. Am J Clin Nutr. 2008; 88:507S–510S. [PubMed: 18689391]

43. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr. 1997; 65:1220S–1228S. discussion 1229S–1231S. [PubMed: 9094926]

44. Scotto, J.; Fears, TR.; Fraumeni, JF, Jr.. Solar radiation. In: Schottenfeld, D.; Fraumeni, JF., Jr., editors. Cancer Epidemiology and Prevention. Oxford University Press; New York: 1996. p. 355-372.

45. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; 7:177–188. [PubMed: 3802833]

46. Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr. 1979; 32:2546–2559. [PubMed: 506977]

47. Liu K, Stamler J, Dyer A, et al. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis. 1978; 31:399–418. [PubMed: 711832]

48. McCullough ML, Weinstein SJ, Freedman DM, et al. Correlates of Circulating 25-Hydroxyvitamin D: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. Am J Epidemiol. 2010; 172

49. Shea MK, Benjamin EJ, Dupuis J, et al. Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr. 2009; 63:458–464. [PubMed: 18030310]

50. Kelly JL, Friedberg JW, Calvi LM, et al. A case-control study of ultraviolet radiation exposure, vitamin D, and lymphoma risk in adults. Cancer Causes Control. 2010; 21:1265–1275. [PubMed: 20373010]

51. Benjamin A, Moriakova A, Akhter N, et al. Determinants of 25-hydroxyvitamin D levels in African-American and Caucasian male veterans. Osteoporos Int. 2009; 20:1795–1803. [PubMed: 19280273]

52. Holick MF, Siris ES, Binkley N, et al. Prevalence of Vitamin D inadequacy among postmenopausal North American women receiving osteoporosis therapy. J Clin Endocrinol Metab. 2005; 90:3215–3224. [PubMed: 15797954]

53. Burgaz A, Akesson A, Oster A, et al. Associations of diet, supplement use, and ultraviolet B radiation exposure with vitamin D status in Swedish women during winter. Am J Clin Nutr. 2007; 86:1399–1404. [PubMed: 17991652]

54. Jacques PF, Felson DT, Tucker KL, et al. Plasma 25-hydroxyvitamin D and its determinants in an elderly population sample. Am J Clin Nutr. 1997; 66:929–936. [PubMed: 9322570]

55. Brock K, Huang WY, Fraser DR, et al. Low vitamin D status is associated with physical inactivity, obesity and low vitamin D intake in a large US sample of healthy middle-aged men and women. J Steroid Biochem Mol Biol. 2010; 121:462–466. [PubMed: 20399270]

56. Engelman CD, Fingerlin TE, Langefeld CD, et al. Genetic and environmental determinants of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels in Hispanic and African Americans. J Clin Endocrinol Metab. 2008; 93:3381–3388. [PubMed: 18593774]

57. Heaney RP, Davies KM, Chen TC, et al. Human serum 25-hydroxycholecalciferol response to extended oral dosing with cholecalciferol. Am J Clin Nutr. 2003; 77:204–210. [PubMed: 12499343]

58. Rosner, B. Fundamentals of biostatistics. Duxbury; Belmont, CA: 2006.

59. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease. Relationships and methods of correction. Annu Rev Public Health. 1993; 14:69–93. [PubMed: 8323607]

60. Ng K, Meyerhardt JA, Wu K, et al. Circulating 25-hydroxyvitamin d levels and survival in patients with colorectal cancer. J Clin Oncol. 2008; 26:2984–2991. [PubMed: 18565885]
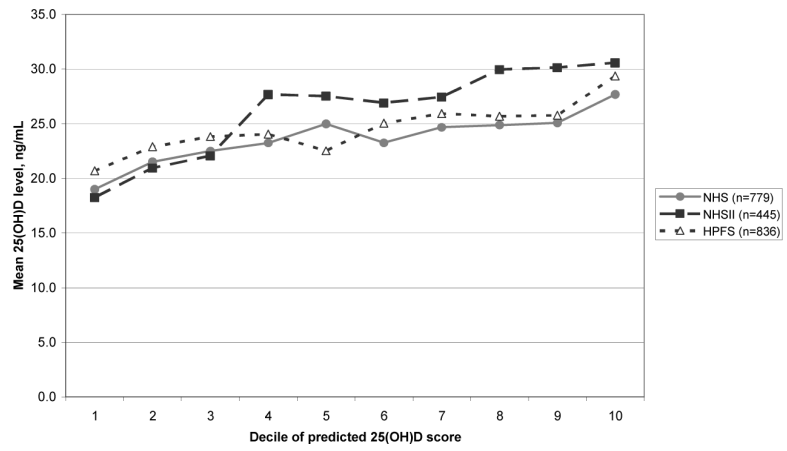
Figure 1.
Mean actual 25(OH)D level by decile of predicted 25(OH)D score in the Nurses' Health Studies (NHS and NHSII) and the Health Professionals Follow-Up Study (HPFS) validation samples.
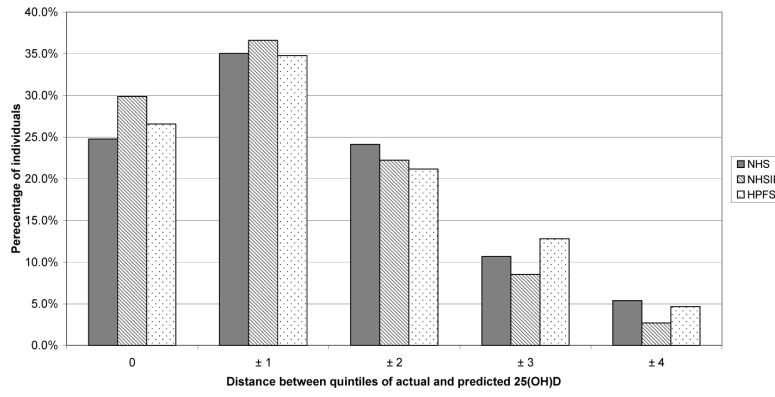
Figure 2.
Percentage of individuals classified by quintiles of actual and predicted 25(OH)D in the Nurses' Health Studies (NHS and NHSII) and Health Professionals Follow-Up Study (HPFS) validation samples.

Table 1

Predictors of Plasma 25(OH)D Level from Multiple Linear Regression Models in the Nurses' Health Studies (NHS & NHSII) and the Health Professionals Follow-Up Study (HPFS).[*]

Predictor	NHS (n = 2079; model r^2 = 0.33)		NHSII (n = 1497; model r^2 = 0.25)		HPFS (n = 911; model r^2 = 0.28)	
	Difference in 25(OH)D, ng/mL (β)	P-value	Difference in 25(OH)D, ng/mL (β)	P-value	Difference in 25(OH)D, ng/mL (β)	P-value
Intercept	22.69		35.78		31.94	
Age (years)[†]	0.07	0.07	−0.23	<0.0001	−0.04	0.24
Race		<0.001		<0.001		0.03
White	0 (referent)		0 (referent)		0 (referent)	
Black	−11.30		−6.42		−4.89	
Asian	---		−5.55		−4.67	
Hispanic	---		−6.83		---	
Other	−1.63		1.98		−1.48	
UV-B flux category[††]		<0.0001		0.67		0.002
1 (high)	0 (referent)		0 (referent)		0 (referent)	
2	−2.69		−0.16		−1.89	
3	−1.29		−0.66		−2.54	
4	---		−0.60		−2.66	
5 (low)	---		---		−3.97	
Dietary vitamin D		<0.0001		0.003		0.001
<100 IU/day	0 (referent)		0 (referent)		0 (referent)	
100–199 IU/day	0.92		1.56		−0.32	
200–299 IU/day	2.19		1.87		2.37	
300–399 IU/day	3.43		3.55		1.93	
400 IU/day	3.33		2.49		3.10	
Supplementary vitamin D		<0.0001		<0.001		<0.001
0 IU/day	0 (referent)		0 (referent)		0 (referent)	
1–199 IU/day	2.85		0.76		2.51	
200–399 IU/day	1.57		2.05		0	
400 IU/day	3.15		2.70		2.54	

Predictor	NHS (n = 2079; model r^2 = 0.33) Difference in 25(OH)D, ng/mL (β)	P-value	NHSII (n = 1497; model r^2 = 0.25) Difference in 25(OH)D, ng/mL (β)	P-value	HPFS (n = 911; model r^2 = 0.28) Difference in 25(OH)D, ng/mL (β)	P-value
Body mass index (kg/m²)		<0.0001		<0.0001		<0.001
<19	---		2.22		---	
<22 (19–21.9 in NHSII)	0 (referent)		0 (referent)		0 (referent)	
22–24.9	−0.57		−0.38		−0.39	
25–29.9	−1.95		−2.35		−2.28	
30–34.9	−3.32		−5.09		−3.44	
35	−8.16		−6.17		−7.30	
Quintile of physical activity		<0.0001		<0.0001		<0.0001
1 (low)	0 (referent)		0 (referent)		0 (referent)	
2	1.77		0.99		1.04	
3	1.15		1.20		0.99	
4	2.13		3.07		3.57	
5 (high)	3.66		3.79		3.75	
Postmenopausal hormone use^		0.001		0.12		
1	0 (referent)		0 (referent)		---	
2	−1.66		0.17		---	
3	−2.11		1.94		---	
4	−1.17		1.53		---	
5	−0.66		0.71		---	
Alcohol intake (g/day)		<0.0001		<0.001		
0	0 (referent)		0 (referent)		---	
>0–<5	0.24		1.34		---	
5–<10	1.33		2.38		---	
10	2.62		2.69		---	
Season of blood draw†		<0.0001		<0.0001		<0.0001
Fall	0 (referent)		0 (referent)		0 (referent)	
Summer	1.18		1.33		0.88	
Spring	−2.68		−5.55		−3.08	
Winter	−3.35		−5.61		−4.45	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

*
Adjusted for laboratory batch.

[†]Age and season not used in predicted score calculation.

[††]UV-B flux category (in RB count $\times 10^{-4}$): NHS: 1 is < 113, 2 is 113, 3 is < 113; NHSII: 1 is 145–196, 2 is 115–144, 3 is 108–114, 4 is < 105; HPFS: 1 is 158–196, 2 is 137–154, 3 is 115–133, 4 is 105–113, 5 is <105.

[#]Mean values in extreme quintiles of physical activity (in MET-hrs/week): NHS: 1 is 1.2, 5 is 41.4; NHSII: 1 is 1.6, 5 is 52.8; HPFS: 1 is 3.6, 5 is 87.4.

^
Postmenopausal hormone use: NHS: 1 is premenopausal, 2 is post-menopausal, never user, 3 is post-menopausal, past user, 4 is postmenopausal, current user, 5 is postmenopausal, unknown use; NHSII: 1 is never user, 2 is past user, 3 is recent past user, 4 is current user, 5 is unknown use.