Pitfalls with Vitamin D Research in Musculoskeletal Disorders and Recommendations on How to Avoid Them

Gary M. Kiebzak¹, Kevin M. Neal², Pooya Hosssienzadeh³, Robert C. Olney⁴, Michael A. Levine⁵,
¹Department of Orthopedics and Sports Medicine, 13535 Nemours Parkway, Nemours Children's Hospital Orlando, FL
²Department of Orthopedic Surgery, Nemours Children's Specialty Care 807 Children's Way Jacksonville, FL
³Department of Orthopaedic Surgery, Washington University School of Medicine 4 Children's Place, Suite 4S60 St Louis, MO
⁴Department of Endocrinology Nemours Children's Specialty Care 807 Children's Way Jacksonville, FL
⁵Center for Bone Health Division of Endocrinology and Diabetes Children's Hospital of Philadelphia 34th & Civic Center Blvd., 11 Northwest Tower, Suite 30 Philadelphia, PA

Abstract
Reports suggesting that vitamin D may have extraskeletal roles have renewed interest in vitamin D research and stimulated publication of an increasing number of new studies each year. These studies typically assess vitamin D status by measuring the blood concentration of 25-hydroxyvitamin D (25[OH]D), the principal circulating metabolite of vitamin D. Unfortunately, variations in assay format, inconsistency in interpreting 25(OH)D concentrations, cohort bias (age, BMI, race, season of measurements, etc.) and failure to measure critical variables needed to interpret study results, makes interpreting results and comparing studies difficult. Further, variation in reporting results (reporting mean values vs. percent of the cohort that is deficient, no clear statement as to clinical relevance of effect size, etc.) further limits interstudy analyses. In this paper, we discuss many common pitfalls in vitamin D research. We also provide recommendations on avoiding these pitfalls and suggest guidelines to enhance consistency in reporting results.

Keywords: Vitamin D, 25-hydroxyvitamin D, deficiency

Introduction
Interest in nonconventional actions of vitamin D remains high many years after the first reports linking vitamin D to a variety of extraskeletal actions. According to a Pubmed.gov search on the term "vitamin D" (December 2018) there were 4497 publications in 2018, and that does not include books, symposium proceedings or media articles on vitamin D. This interest naturally
extends to pediatric orthopedic conditions, where studies have attempted to identify associations between vitamin D status and fracture risk, fracture severity, fracture healing, and skeletal disorders such as adolescent idiopathic scoliosis (AIS). Studies that assess vitamin D status do so by measuring the concentration of 25-hydroxyvitamin D (25(OH)D) in blood. This is because 25(OH)D is the principal circulating form of vitamin D, and is the precursor to the biologically active metabolite, 1,25-dihydroxyvitamin D (1,25(OH)2D). The concentration of 25(OH)D in blood reflects both the amount of parent vitamin D3 that is generated in the skin upon exposure to ultraviolet B radiation as well as the amounts of vitamin D2 and D3 that are obtained from the diet and vitamin supplements. Despite enormous interest in vitamin D research, there are many pitfalls that cloud interpretation of study results. This is due in part to the many variables that affect 25(OH)D concentration.

Our objective was to identify and discuss common design and data presentation problems in vitamin D study results. Reviews of the pediatric vitamin D literature, namely studies associating vitamin D status to fracture risk and studies assessing vitamin D status in AIS, are used as examples of the lack of clarity in presentation of study results. Finally, we will make recommendations about how these common pitfalls can be avoided.

Categories of Pitfalls

Common problems that confound vitamin D research studies can be divided into three categories: 1) design issues, 2) inconsistent presentation of results, and 3) failure to account for variables that are known to influence serum 25(OH)D concentrations. In addition, a lack of assay standardization further confuses the picture, particularly when reports do not provide adequate information about how 25(OH)D was assayed. When such problems are present, they can make it difficult to generate direct comparisons with other studies and can confound interpretation of results.

As an example, we reviewed all papers known to us that reported 25(OH)D in patients with AIS (n = 6). The race/ethnicity distributions were different between the six papers. One paper did not report 25(OH)D as the mean ± standard deviation for the study groups. Only three papers presented data as the percent who were deficient, insufficient, and sufficient; two of these defined deficiency <20ng/ml 25(OH)D and one paper as 25(OH)D < 10ng.ml. None of the papers accounted for use of multivitamins nor the use of sunscreen in the study or control groups. Only two studies attempted to correlate vitamin D status with disease severity (based on Cobb angle). One paper did not report the type of assay used. Of the other five papers, four different assay systems were used. Individually, all the studies are interesting and make legitimate contributions, however, they are not directly comparable nor could they likely be used in a meta-analysis, in part because of the variability in study design and data presentation.

Design Problems

Research studies should be hypothesis driven and not simply reporting measurement results. Stating a priori whether vitamin D status is hypothesized as a causative factor in the condition being studied, is secondary to the condition itself (i.e., the condition impacts vitamin D status), or has no relation to the condition, would help frame interpretation of results. Also, it should be stated a priori what difference in the magnitude of percent deficiency or 25(OH)D concentration between two comparative groups is considered to be clinically meaningful. For example, "We hypothesize that low 25(OH)D is a causative factor in spinal curve progression in AIS manifest by at least a 30% difference in 25(OH)D concentration compared to controls." Such a statement would temper overinterpretation of clinically insignificant differences (despite statistically significant differences).
The blood level of 25(OH)D can change substantially in a very short period of time (literally within minutes) after exposure to sunlight, or over a longer time period with the use of oral supplements. Therefore, studies must be designed so that measurement of 25(OH)D occurs at time points that are relevant to the development of the study endpoints. For example, if attempting to correlate fracture risk with vitamin D status, the 25(OH)D should be measured within a very short time period after the patient presents with the fracture. For studies in which an effect might take place over a protracted period of time such as bone mineral density or muscle mass changes, or the effect of a treatment or procedure that may last many months, it would be appropriate to verify chronic vitamin D status. This might require repetitive determination of 25(OH)D levels over an interval of time that corresponds to the duration of time necessary for the outcome to be realized. This is because 25(OH)D levels over time are susceptible to the phenomenon of regression to the mean, whereby a patient may be deficient on one measurement, but insufficient or normal at a subsequent measurement (due perhaps as a consequence of season, lifestyle factors, assay variability, etc.). For example, if assessing the effect of vitamin D status on height over three years, it would be inappropriate to measure a single 25(OH)D concentration at the beginning of the study and then assume that this is a valid reflection of vitamin D status over the entire three year period of study.

Failure to measure all critical variables may confound interpretation of results (Table 1; variables affecting 25(OH)D concentration will be discussed further below.)

For example, if a study hypothesis states that vitamin D status impacts spinal curve progression in scoliosis or is a factor in fracture risk, then it is important to not only measure known variables that affect 25(OH)D concentration (such as BMI, use of vitamin supplements, etc., Table 1) but also known variables that affect curve progression (such as bone mineral density, growth stage, and menarchal status).

A major problem with the concept that poor vitamin D status (i.e., low concentration of 25(OH)D) is clinically meaningful is that it is unusual to observe signs or symptoms that can be directly attributed to the "low" 25(OH)D concentrations. In fact, patients frequently have 25(OH)D levels in the deficient range without any obvious symptoms or abnormalities in standard serum chemistries. In addition to the well known detrimental effect of vitamin D deficiency on bone, vitamin D deficiency can also have a detrimental effect on skeletal muscle; these include type II muscle fiber atrophy and metabolic changes manifest as muscle weakness which has been associated with increased risk of falling. Thus, we suggest all studies should attempt to collect bone mineral density, and weight- and age-adjusted measures of muscle strength, which could include grip strength or proximal muscle strength (ability to rise from a sitting position, stair climbing, or speed walking). If such measures are normal in the presence of vitamin D insufficiency then that could argue against the clinical relevance of an observation of suboptimal vitamin D status. A measurement of parathyroid hormone (PTH) may be useful in interpreting the 25(OH)D values. A feedback relationship between serum ionized calcium, PTH, and vitamin D metabolism is well established (as blood calcium levels drop, PTH levels rise and among other effects, stimulate synthesis of 1,25(OH)D). Hence, PTH will increase if 25(OH)D is low enough to impact calcium metabolism, although the set point for this may vary from patient to patient. Theoretically, if a compensatory increase in PTH is not observed in conjunction with a "low" 25(OH)D, then the clinical significance of the "low" 25(OH)D value may be questionable. While these relationships are well established in adults, a compensatory increase in PTH as 25(OH)D falls below a critical threshold may occur at different thresholds in children or elderly adults.
Investigators must carefully select the control group used to compare 25(OH)D concentrations or prevalence of deficiency. This is not easily accomplished but is often directly related to the study hypothesis. For example, if the study hypothesis is that low 25(OH)D is a risk factor for severe pediatric forearm fractures requiring surgical reduction, then a logical control group would be patients with less severe fractures that can be treated conservatively, who are matched for age, sex, BMI, activity level, sun exposure, multivitamin use, etc. It would not be appropriate, for example, to use hospitalized children who may have illnesses that could impact vitamin D status.

**Inconsistent Presentation of Results**

Most papers present serum concentrations of 25(OH)D as the group mean ± standard deviation (SD). However, perhaps more important is the distribution of values within the cohort (the percent that are deficient, insufficient, and sufficient). Some papers do not report this distribution, so it is not possible to fully interpret the mean 25(OH)D value (see discussion of subgroup analyses below). Box and whisker plots would be an effective way to present these data.

One issue related to the presentation of a distribution of values is the definition of cutoffs defining deficiency. Unfortunately, these vary according to which guidelines are followed, but most are trending toward defining vitamin D deficiency as 25(OH)D < 20 ng/ml (Table 2).16-21

**Failure to Account for Variables Known to Affect the Serum Concentration of 25(OH)D**

Interpretation of study results can be influenced by subgroup analysis, and failure to report subgroup results can distort the true findings in a study. Because many factors can affect the 25(OH)D level in blood (either increasing or decreasing the concentration), it is critically important to carefully characterize the study cohort (Table 1). Lifestyle factors known to affect 25(OH)D include sun exposure practices such as exuberant use of sunscreen, diets low in dairy products, and various supplements and medications that influence vitamin D metabolism (multivitamins, calcium supplements with vitamin D, anticonvulsants, etc.). Perhaps the most important patient characteristic that affects 25(OH)D is race/ethnicity. Studies consistently report that nonwhite cohorts (African Americans, Hispanics, Asians) have significantly lower 25(OH)D concentrations than Caucasians. However, the relative contributions of genetics (including skin pigmentation and body fat profiles), socioeconomic status, and culture (including diet low in vitamin D, avoidance of sun exposure, etc.) on these findings remains unclear. Another important demographic variable is BMI, as 25(OH)D is lower in obese subjects compared to normal weight subjects. Failure to carefully match controls, or failure to present subgroup analyses can lead to bias in results and misinterpretation. The implications are obvious. For example, if one cohort has a high percentage of patients who take multivitamins that contain a form of vitamin D and the comparative group does not, then the difference in 25(OH)D between them could be affected by the use multivitamins.

Table 3 shows results of a study conducted to document vitamin D status in children with radius fractures (mean age, 9.8 ± 3.4 years; 65% were boys). These previously unpublished data show the potential for misleading reporting of study findings. The total cohort mean is less than the cutoff of 30ng/ml that many labs and some guidelines define as vitamin D sufficiency. But, the subgroup of Caucasian subjects had a mean 25(OH)D that was 25% higher (and in the "sufficient" range, >30 ng/ml), compared to all nonwhites (despite no significant difference in BMI, a variable that can affect 25(OH)D concentration). Thus, we suggest it would be misleading to report only the total group mean 25(OH)D value given this significant subgroup difference. Table 4 shows the covariable confounder of BMI on the risk for severe fracture, in
which both a high BMI (classifying patients as obese or overweight) and 25(OH)D deficiency were independent risk factors for having a severe distal radius fractures requiring surgical management. Failure to report and discuss the impact of BMI would be misleading, possibly overweighting the impact of the vitamin D status as risk factor for severe fractures.

25(OH)D Assay Issues
Multiple assay systems are available for the measurement of 25(OH)D in blood. These assays can be grouped into two general categories: 1) immune based and 2) chromatography based (ultraviolet or mass spectrometric detection). The mass spectroscopy systems are currently favored as the standard. In fact, the liquid chromatography tandem mass spectroscopy method is used by the Center for Disease Control and Prevention as the reference measurement system. However, these systems require sophisticated equipment and technical expertise and are difficult to automate. Hence, immunoassays are more commonly available.

Unfortunately, there are well described problems with variation and a lack of congruity between different assays systems, with considerable differences in the 25(OH)D concentration reported when the same sample is assayed by different systems. Further, there are differences between assays with respect to their ability to measure both 25(OH)D2 and 25(OH)D3, and some assays may cross-react and measure other metabolites. There can be a clinically important difference if one assay system measures 25(OH)D 10-15% lower than another, as it could impact the number of patients reported as deficient. These assay differences may be problematic when attempting to compare studies or do meta-analyses. Thus, it is important to know detail about the assay system used when evaluating the results of a study (some papers do not report the type of assays used). Efforts are in progress to standardize 25(OH)D assays, and to have lab certification of assay systems to mitigate the various issues associated with multiple assays types. However, at the time of this writing, such a solution has not been implemented.

Solutions and Recommendations
Professional societies or journals that publish vitamin D-related studies could help standardize reporting of results by providing guidelines and checklists for ensuring some degree of study standardization. This especially applies to standardizing the presentation of cutoff values for defining deficiency.

Design
- Investigators should state a priori what difference between the main study cohort and the comparative group is hypothesized to be clinically relevant to frame later interpretation of obtained results.
- Investigators should measure and report all variables known to affect 25(OH)D and make adjustments in data, or discuss the absence of such information in a paragraph regarding study limitations and the possible impact on interpretation.
- Measurements of 25(OH)D should occur over the time interval during which an effect or outcome is expected, not merely at the beginning or end of a study.
- While it is important for the sake of generalizability to have diversity in study cohorts (and institutional review boards may require it), it may be problematic for vitamin D studies. Study designs should either consider enrolling homogeneous patient groups or anticipate the need for subgroups analyses. The latter would require a randomized block design for prospective studies or matched patient selection in retrospective studies, with large enough sample size to allow for subgroup analyses.

Presentation of Data
The mean 25(OH)D alone does not provide enough information about the true vitamin D status in a cohort. All vitamin D studies should report not only the mean 25(OH)D but also the percent deficiency (the range of values can be eloquently presented using box and whisker plots). However, guidelines defining deficiency are not consistent. We suggest that a solution for minimal standardization is for journals to require that results be presented for a sliding scale of cutoff points, reporting percent deficiency at <10, <20, and <30ng/ml (as simple bar graphs). Not only would that allow readers to judge for themselves what is true deficiency, but would be beneficial for comparing results between papers and documenting important detail in the literature in the likely event of future adjustments of cutoff values.

*25(OH)D Assay*

- Details of the assay must be presented in publications, and if possible, whether the lab and assay have been certified per the National Institute of Standards and Technology ([NIST], an agency of the U.S. Department of Commerce, Gaithersburg, MD), or the Vitamin D Standardization Program of the National Institutes of Health Office of Dietary Supplement.

If investigators plan to do long-term data collection or multiple vitamin D studies, they should consider preparing an internal standard that can be remeasured over time or in serial studies to account for assay drift. This can be done by collecting a large sample of blood and then freezing aliquots that can be serially measured. This would be particularly important in multisite studies if a central assay lab could not be used. Reference samples of 25(OH)D can also be purchased commercially. Alternatively, samples can be held until the end of the study and run together as a single batch to avoid issues with assay drift.

**Summary**

We suggest that guidelines should be available to standardize studies of vitamin D and data presentation to allow for direct comparisons and for high-quality meta-analyses. The authors believe that the suggestions made here are relatively easy to implement. Professional societies and especially journal editorial boards should consider checklists to ensure critical data elements are presented in published vitamin D studies.

**References**

doi: 10.1186/1748-7161-9-S1-045


doi: 10.1097/BPO.0000000000000587
Table 1. Variables Known to Affect Serum 25(OH)D Concentration

<table>
<thead>
<tr>
<th>Variable</th>
<th>Impact on Vitamin D Metabolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Limited evidence that 25(OH)D decreases going into teen years. Many reports of decreased levels in institutionalized frail elderly men and women. 22-24</td>
</tr>
<tr>
<td>BMI or age- adjusted classifications as overweight, obese</td>
<td>Well established that 25(OH)D is lower in obese subjects compared to normal weight subjects. 25-28</td>
</tr>
<tr>
<td>Race</td>
<td>Well established that nonwhite subjects (African Americans, Hispanics, Asian) have lower 25(OH)D compared to Caucasian subjects. 8,22,29-31</td>
</tr>
<tr>
<td>Sun exposure</td>
<td>25(OH)D increases with increasing unprotected sun exposure (the response can very rapid). However, because of the difficulty in obtaining reliable estimates of sun exposure, investigators should evaluate data for possible impact of season of 25(OH)D measurement. With respect to study design in studies with multiple comparison groups, ensure that samples are collected equally by season (i.e., if a study has two groups of patients, and one group had most samples collected in summer and the other had most samples collected in winter, then there could be bias in the 25(OH)D concentrations). 2,9,32</td>
</tr>
<tr>
<td>Sunscreen use or skin protected by clothing</td>
<td>Well established that 25(OH)D decreases with excessive use of sunscreen or clothing protection (which may be cultural). 2,9,32</td>
</tr>
<tr>
<td>Dietary intake of vitamin D</td>
<td>25(OH)D decreased if diet is devoid of foods that contain vitamin D. But the effect of diet on 25(OH)D can rarely be effectively isolated, and patient self-reporting of dietary intake is generally unreliable. However, study enrollment criteria can exclude those patients with complete avoidance of dairy products or fatty fish. 1,2</td>
</tr>
<tr>
<td>Use of vitamin supplement</td>
<td>Higher 25(OH)D compared to subjects that do not use a vitamin D supplement. 2,33</td>
</tr>
<tr>
<td>Use of medications known to affect vitamin D metabolism</td>
<td>Various medications affect vitamin D metabolism directly or indirectly (via effects on calcium balance). For example, use of seizure drugs phenobarbital and phenytoin, and anti-tuberculosis drugs (isoniazid) result in decreased 25(OH)D synthesis in the liver. Study exclusion criteria should include current use of such medications. 2,34</td>
</tr>
<tr>
<td>Hospitalization or medical conditions limiting sun exposure, associated with poor diet, leading to frailty, wasting etc.</td>
<td>Patients that are ill or frail with reduced exposure to sunlight and/or have vitamin D deficient diets will have decreased 25(OH)D compared to healthy subjects. This is especially true if there are impairments in liver of kidney function, which should be an exclusion criterion in a study. 2,33,35</td>
</tr>
</tbody>
</table>
Table 2. Variation in Guidelines Defining Vitamin D Status

<table>
<thead>
<tr>
<th>Organization/Society Guidelines for Defining Vitamin D Status</th>
<th>Vitamin D Status (Values are ng/ml)</th>
<th>Severe Deficiency</th>
<th>Deficiency</th>
<th>Insufficiency</th>
<th>Sufficiency</th>
<th>No added benefit</th>
<th>Possible harm (toxicity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute of Medicine, 2011</td>
<td>&lt;12</td>
<td>12-20</td>
<td>21-30</td>
<td>31-50</td>
<td>&gt;50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endocrine Society, 2011</td>
<td>&lt;20</td>
<td>20-29</td>
<td>30-100</td>
<td></td>
<td>&gt;100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Academy of Pediatrics (Pediatric Endocrine Society), 2008</td>
<td>&lt;5</td>
<td>5-15</td>
<td>16-20</td>
<td>21-100</td>
<td>101-149 (“excess”)</td>
<td>&gt;150</td>
<td></td>
</tr>
<tr>
<td>Kidney Disease Outcomes Quality Initiative, 2017</td>
<td>&lt;5</td>
<td>5-15</td>
<td>16-30</td>
<td></td>
<td>&gt;30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mayo Medical Laboratories, 2017</td>
<td>&lt;10</td>
<td>10-24</td>
<td>25-80</td>
<td></td>
<td>&gt;80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quest Diagnostics (commercial lab)</td>
<td>&lt;20</td>
<td>20-29</td>
<td>&gt;30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Subgroup Effects on Reported Mean 25(OHD) Concentration in 100 Children With Radius Fractures (Previously Unpublished Data)

<table>
<thead>
<tr>
<th></th>
<th>All Patients</th>
<th>All African Americans and Hispanics (n = 77)</th>
<th>Caucasian (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index</td>
<td>19.5 ± 4.2</td>
<td>19.8 ± 3.85</td>
<td>18.5 ± 4.92</td>
</tr>
<tr>
<td>25OHD, ng/ml</td>
<td>27.5 ± 8.28</td>
<td>26.0 ± 7.17*</td>
<td>32.5 ± 3.89</td>
</tr>
</tbody>
</table>

Data are mean ± SD.

25(OHD) was assayed within 30 days of the forearm fracture using a blood-spot card technique (ZRT Laboratory, Beaverton, OR). Samples were assayed using a liquid chromatography-mass spectrometry method.

*Unpaired t tests compared to Caucasian group, P < 0.0001.
Table 4. BMI as a Significant Confounder in Study Results (Previously Unpublished Data).

<table>
<thead>
<tr>
<th></th>
<th>No reduction or closed reduction (n = 88)</th>
<th>Surgical reduction (n = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>9.52 ± 3.2</td>
<td>12.07 ± 1.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P = 0.008*</td>
</tr>
<tr>
<td>Body mass index</td>
<td>18.9 ± 4.0</td>
<td>23.9 ± 4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P = 0.0001*</td>
</tr>
<tr>
<td>25OHD, ng/ml</td>
<td>28.1 ± 8.08</td>
<td>23.3 ± 8.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS, P = 0.057*</td>
</tr>
<tr>
<td>Percent deficient</td>
<td>17</td>
<td>50</td>
</tr>
<tr>
<td>25(OH)D &lt; 20 ng/ml</td>
<td></td>
<td>P = 0.0172**</td>
</tr>
</tbody>
</table>

Both BMI (and status as overweight and obese) and vitamin D deficiency were significant independent risk factors for surgical reduction.

* Compared to non-operative management group, unpaired t-test.

** Compared to non-operative management group, Fishers Exact test.

Data are mean ± SD.

25(OHD) was assayed within 30 days of the forearm fracture using a blood-spot card technique (ZRT Laboratory, Beaverton, OR). Samples were assayed using a liquid chromatography-mass spectrometry method.)