

Vitamin D Assay Standardisation and the Deficiency Threshold: How Method Choice Moves the Diagnosis

Solveig Braatveit^{1*}, Nadia Bellofiore-Trani² and Jonas Wredenhagen³

¹Department of Clinical Biochemistry, Valdres Centre for Clinical Biochemistry, Fagernes, Norway. ORCID: [0000-0003-5724-6741](https://orcid.org/0000-0003-5724-6741)

²Centre for Evidence Synthesis, Garfagnana Institute of Laboratory Medicine, Lucca, Italy. ORCID: [0000-0003-9129-9564](https://orcid.org/0000-0003-9129-9564)

³Department of Laboratory Medicine, Taunus Centre for Clinical Biochemistry, Bad Homburg, Germany. ORCID: [0000-0001-2188-3920](https://orcid.org/0000-0001-2188-3920)

*Correspondence to: Solveig Braatveit, Department of Clinical Biochemistry, Valdres Centre for Clinical Biochemistry, Fagernes, Norway, E-mail: s.braatveit@vccb.no

Received: 23 May 2025; Accepted: 14 October 2025; Published: 03 February 2026

Copyright: © 2026 Solveig Braatveit, et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Despite two decades of standardisation effort through the Vitamin D Standardization Program (VDSP), clinically meaningful differences between 25-hydroxyvitamin D assay platforms persist, and these differences are large enough, relative to the narrow band separating commonly used deficiency and insufficiency thresholds, to change which diagnostic category an individual patient's result falls into depending on which platform measured it. We systematically reviewed method comparison studies published since VDSP standardisation efforts began, quantifying inter-platform bias at the specific concentrations corresponding to the 20 ng/mL and 30 ng/mL thresholds most widely used to define deficiency and insufficiency. Inter-platform bias at these thresholds ranged from -15% to +12% across the platforms reviewed even among assays certified as VDSP-traceable, sufficient to reclassify a meaningful proportion of patients near either threshold. We discuss what this means for a laboratory reporting a categorical deficiency classification rather than a numeric result alone, and for interpreting a patient's vitamin D trend across a platform change.

Keywords: Vitamin D; 25-hydroxyvitamin D; Assay standardisation; Deficiency threshold; Method comparison

Introduction

The Vitamin D Standardization Program was established specifically to address well-documented inter-platform variability in 25-hydroxyvitamin D measurement, providing reference measurement procedure traceability that manufacturers can certify their assays against.^[1] Two decades on, standardisation has measurably reduced variability compared with the pre-VDSP era, but has not eliminated clinically meaningful inter-platform differences, and the clinical significance of any remaining bias is amplified by how narrow the gap is between the deficiency and insufficiency thresholds most widely used in practice -- commonly 20 ng/mL and 30 ng/mL respectively, a 10 ng/mL band that a modest percentage bias can meaningfully encroach on.^[2]

We systematically reviewed method comparison studies published since VDSP standardisation efforts began, focusing specifically on bias magnitude at the concentrations corresponding to these two thresholds rather than bias averaged across the full measurement range, since average bias across a wide range can understate the clinically most consequential bias occurring specifically near a diagnostic threshold.

The clinical stakes of this residual inter-platform variability are substantial given how widely 25-hydroxyvitamin D is now tested: it is among the most frequently ordered specialised chemistry assays in many health systems, used to guide supplementation decisions, bone health assessment, and, increasingly, population-level deficiency prevalence estimates that inform public health policy, meaning even a modest proportion of threshold-adjacent misclassification translates into

a large absolute number of individual clinical decisions and policy inputs potentially affected.^{[2][8]}

Methods

We identified 28 method comparison studies reporting 25-hydroxyvitamin D results from at least one VDSP-certified platform against a reference measurement procedure or another certified platform, published since 2011. For each study, bias at 20 ng/mL and 30 ng/mL was extracted or, where not directly reported, estimated from the study's own regression equation applied at these two concentrations.^[3]

Studies were included regardless of whether the comparator platform was itself VDSP-certified, since comparisons against a non-certified but widely used platform remain clinically relevant to laboratories currently operating that platform, though bias estimates were analysed separately for certified-versus-certified and certified-versus-non-certified comparisons to assess whether certification status itself predicted bias magnitude.^[4]

Results

Inter-platform bias at the 20 ng/mL and 30 ng/mL thresholds ranged from -15% to +12% across the 28 studies reviewed, and this range persisted even when restricting to comparisons between two VDSP-certified platforms, indicating that VDSP certification substantially narrows but does not eliminate clinically relevant inter-platform differences at these specific concentrations.^{[4][5]}

Several studies additionally reported that bias magnitude was not constant across the measurement range for a given platform pair, with some platform combinations showing larger proportional bias near the lower deficiency threshold than at higher concentrations, meaning a single average bias figure for a platform pair can understate the bias specifically relevant to deficiency classification.^[6] Simulation based on the reported bias distributions suggested that between 8% and 19% of patients with a true value within 2 ng/mL of either threshold could be classified into a different diagnostic category depending on which of the reviewed platforms measured the sample.^[7]

Comparison Type	Bias at 20 ng/mL	Bias at 30 ng/mL
Certified vs certified	-9% to +7%	-6% to +5%
Certified vs non-certified	-15% to +12%	-11% to +9%
Immunoassay vs LC-MS/MS	-13% to +10%	-8% to +7%

Table 1: Range of Reported Inter-Platform Bias at the Two Diagnostic Thresholds, by Comparison Type.

Discussion

A bias magnitude sufficient to reclassify a meaningful minority of borderline patients across a platform change has two practical implications for laboratory practice. First, a categorical deficiency or insufficiency classification reported alongside the numeric result should be treated by clinicians as somewhat platform-dependent near the threshold boundaries, particularly for a patient whose result falls within a few ng/mL of a threshold, rather than as an absolute cutoff.^[8] Second, a laboratory changing vitamin D assay platform should anticipate that some patients' longitudinal trend across the platform change will show an apparent shift attributable to the platform bias rather than to a genuine change in vitamin D status, and informing referring clinicians of the platform change and its approximate bias direction at the time of transition would help avoid misinterpreting this as a clinical change.^[9]

This review is limited to studies reporting results at or near the two thresholds specifically, and cannot characterise bias behaviour at other clinically relevant concentrations such as the higher thresholds sometimes used in specific populations; it is also limited by the relatively small number of platform pairs with published head-to-head comparison data, meaning some currently marketed platforms are not represented in the reviewed literature.^[10]

Why Certification Narrows But Does Not Eliminate Bias

The finding that even certified-versus-certified platform comparisons retained bias of up to 9% at the 20 ng/mL threshold warrants explanation, since VDSP certification is specifically intended to establish traceability to a common reference measurement procedure.

VDSP certification establishes traceability at the level of overall method accuracy against the reference measurement procedure across a defined validation panel, but does not guarantee identical bias behaviour at every individual concentration across the full measurement range, particularly for immunoassay platforms whose antibody-based detection can show concentration-dependent or matrix-dependent bias that a single-point or limited-panel certification exercise may not fully characterise.^{[5][6]}

This distinction between overall method traceability and concentration-specific bias behaviour is not unique to vitamin D assays and is recognised more broadly in analytical chemistry as a limitation of single-point or limited-range certification exercises, but it is particularly consequential for vitamin D given how narrow the clinically relevant diagnostic band is relative to the overall measurement range the assay is validated across.^[10]

Comparison Between Immunoassay and Mass Spectrometry Platforms

The comparison between immunoassay and liquid chromatography-tandem mass spectrometry (LC-MS/MS) platforms deserves separate discussion, since these two measurement principles are mechanistically distinct in ways that plausibly explain some of the observed bias pattern.

LC-MS/MS directly measures 25-hydroxyvitamin D₂ and D₃ separately via mass-to-charge ratio detection following chromatographic separation, generally considered to more closely approximate the reference measurement procedure's own methodology, whereas immunoassay platforms detect 25-hydroxyvitamin D via antibody binding, which can show differential cross-reactivity with the D₂ and D₃ forms and with other structurally related vitamin D metabolites, a source of bias not present in the equivalent form for a well-validated LC-MS/MS method.^{[9][10]}

This mechanistic difference is consistent with this review's finding that immunoassay-versus-LC-MS/MS comparisons showed somewhat wider bias ranges than certified-versus-certified immunoassay comparisons, though the overlap between the two comparison categories was substantial enough that measurement principle alone does not fully predict bias magnitude for any individual platform pair.

Implications for Population-Level Deficiency Prevalence Estimates

Beyond the individual clinical decision-making implications already discussed, the residual inter-platform bias characterised in this review carries implications for population-level vitamin D deficiency prevalence estimates used in public health surveillance and policy.

A national or regional health survey reporting vitamin D deficiency prevalence against a 20 ng/mL threshold will produce a different prevalence estimate depending on which assay platform the survey's laboratory used, independent of any genuine difference in the underlying population's vitamin D status, and comparing prevalence estimates across surveys or time periods that used different assay platforms should account for this source of variation before attributing an observed difference to a genuine change in population vitamin D status.^[8]

Cashman and colleagues' work applying VDSP standardisation protocols retrospectively to a seasonal cohort illustrates this concern concretely, demonstrating a meaningful shift in estimated deficiency prevalence after applying standardisation-derived correction equations compared with the originally reported, uncorrected prevalence figures, underscoring that population-level surveillance data predating or using non-standardised assays should be interpreted cautiously when compared against more recent standardised data.^[4]

Practical Recommendations for Laboratories Reporting Categorical Classification

Drawing together this review's findings, we propose several practical recommendations for laboratories that report a categorical deficiency or insufficiency classification alongside the numeric 25-hydroxyvitamin D result.

Report the numeric result alongside, rather than instead of, the categorical classification, so that a clinician reviewing a borderline result can apply their own clinical judgement about proximity to the threshold rather than relying solely on a binary categorical label that obscures how close the result sits to the boundary.^[8]

Consider a narrow intermediate reporting band -- for example, flagging results within 2 ng/mL of a threshold with a specific comment noting that assay-related measurement variability could place the true result on either side of the threshold -- rather than reporting every result as unambiguously deficient, insufficient, or sufficient regardless of proximity to a boundary.

When changing assay platform, communicate the change and its approximate bias direction, where known from method comparison data, to referring clinicians managing patients with longitudinal vitamin D monitoring, so that an apparent trend shift coinciding with the platform change is not misinterpreted as a genuine change in the patient's vitamin D status.^[9]

Economic Considerations

The findings of this review carry modest but real economic implications for laboratories and health systems managing large-volume vitamin D testing.

A laboratory choosing between candidate assay platforms during procurement could reasonably weight demonstrated threshold-specific bias performance, rather than overall method accuracy alone, as a procurement criterion, potentially avoiding the downstream cost of threshold-adjacent patient reclassification and the associated clinical confusion and repeat testing that can follow an unexplained platform-related trend shift.^[7]

At a population level, the cost of recalibrating public health surveillance trend data to account for known inter-platform bias, as Cashman and colleagues demonstrated is both feasible and informative, is modest relative to the cost of drawing an incorrect policy conclusion from an uncorrected apparent change in population deficiency prevalence.

Study Limitations

This review's reliance on published method comparison studies carries the limitations inherent to systematic reviews of heterogeneous primary studies.

The 28 included studies varied in sample size, population characteristics, and statistical methodology for characterising bias, and the bias estimates extracted at the two specific

diagnostic thresholds in some cases required extrapolation from a regression equation derived from a wider measurement range rather than being directly observed at the threshold concentration itself, introducing some additional uncertainty into the specific bias figures reported in this review.^[6]

Not every currently marketed 25-hydroxyvitamin D assay platform is represented in the published method comparison literature reviewed here, meaning this review's findings characterise the subset of platforms with available comparison data rather than the full range of platforms in current clinical use.

Future Directions

A prospective, coordinated method comparison study specifically designed to characterise bias at the 20 ng/mL and 30 ng/mL thresholds directly, across the full range of currently marketed VDSP-certified platforms using a common, well-characterised sample set, would provide more precise and current threshold-specific bias estimates than this review's reliance on heterogeneous published studies with varying methodology can offer.

Further work developing and validating laboratory reporting formats that explicitly communicate threshold proximity and associated measurement uncertainty, rather than a simple binary categorical classification, and assessing how clinicians interpret and act on such enhanced reporting formats in practice, would help translate this review's findings into a concrete improvement in clinical reporting practice.

Comparison with Previously Published Reviews of Vitamin D Standardisation

This review's findings build on, and are broadly consistent with, previously published assessments of VDSP standardisation progress, while adding a specific threshold-focused analysis that most prior reviews, concerned primarily with overall method accuracy, have not provided in the same detail.

Binkley and Sempos' foundational description of the standardisation programme's goals and the broader literature characterising baseline assay performance before standardisation both establish that substantial inter-platform variability existed prior to VDSP efforts, and this review's contribution is to show that, even after two decades of standardisation, the residual bias remains large enough, specifically at the clinically relevant diagnostic thresholds, to carry continuing practical significance for individual patient classification rather than being merely a historical pre-standardisation concern.^{[7][3]}

Wise and colleagues' baseline assessment work for the VDSP programme itself anticipated that bias reduction, while substantial, would not reach zero for every platform pair, consistent with this review's finding that even certified-versus-certified comparisons retained meaningful residual bias at the two thresholds examined.

Role of Pre-Analytical and Matrix Effects in Residual Bias

Beyond the antibody cross-reactivity mechanism discussed earlier for immunoassay platforms, pre-analytical and sample matrix effects contribute further to the residual inter-platform bias characterised in this review and warrant separate mention.

Vitamin D binding protein concentration and genotype vary between individuals and can affect how readily 25-hydroxyvitamin D is released from its binding protein during certain assay extraction or pretreatment steps, a matrix effect that may interact differently with different assay chemistries and could plausibly contribute to some of the patient-to-patient variability in bias magnitude observed within, as well as between, the platform pairs characterised in the reviewed studies.^{[9][10]}

Haemolysis, lipaemia, and icterus, recognised general interferents for many immunoassay platforms, have also been reported to affect 25-hydroxyvitamin D immunoassay results to varying degrees depending on the specific platform's antibody and detection chemistry, adding a further source of platform-specific variability beyond the core measurement principle differences already discussed.

Implications for Clinical Guideline Thresholds Themselves

The residual inter-platform bias characterised in this review also raises a more fundamental question about the clinical guideline thresholds themselves: whether a single numeric threshold, uniformly applied regardless of the measuring platform's known bias characteristics, is the most appropriate way to define deficiency and insufficiency in the first place.

The Endocrine Society and other guideline bodies that established the 20 ng/mL and 30 ng/mL thresholds did so based on evidence linking vitamin D status to specific clinical outcomes such as bone health, generally without explicit reference to inter-platform measurement variability of the magnitude this review documents, raising the question of whether future guideline revisions should incorporate platform-aware threshold guidance or measurement uncertainty ranges rather than a single numeric cutoff presented as platform-independent.^[2]

Some guideline bodies and professional laboratory organisations have begun to address this by recommending that a result within a defined margin of a threshold be interpreted with caution rather than as a definitive categorical classification, an approach consistent with this review's own practical recommendation for an intermediate reporting band discussed earlier.

Applicability to Other Threshold-Dependent Analytes

Although this review focuses specifically on 25-hydroxyvitamin D, the underlying problem it characterises -- a diagnostic threshold narrow relative to residual inter-platform measurement bias -- is not unique to vitamin D and has broader relevance to other analytes with similarly narrow clinically relevant diagnostic bands.

Analytes such as parathyroid hormone, where regional reference range and assay-specific calibration differences are well documented, and certain tumour markers with narrow decision thresholds for treatment response assessment, share this general structural vulnerability to platform-related misclassification, suggesting that the threshold-specific bias analysis methodology used in this review could usefully be applied to these other analytes rather than remaining specific to vitamin D alone.^[6]

Governance Recommendations for Laboratories and Manufacturers

Drawing together this review's findings, we propose governance recommendations addressed separately to laboratories and to assay manufacturers.

Laboratories should document, within their local quality management system, which assay platform is in current use, its VDSP certification status, and any known threshold-specific bias characteristics from published method comparison data, updating this documentation whenever the platform changes so that this information remains available to support the clinical communication recommendations made earlier in this review.^[5]

Manufacturers seeking VDSP certification should be encouraged, whether through programme requirements or market expectation, to publish threshold-specific bias data from their certification validation studies rather than only an overall accuracy summary across the full validation range, giving laboratories more direct access to the specific bias information most relevant to their diagnostic reporting practice without requiring reliance on independently published method comparison studies of the kind this review needed to systematically collect.

Patient Communication Considerations

Beyond the clinician-facing communication recommendations made earlier in this review, patients themselves may encounter and be confused by an apparent change in their own vitamin D classification across a laboratory platform change, and this warrants brief separate consideration.

A patient reviewing their own longitudinal laboratory results, increasingly common given widespread patient portal access to laboratory data, may notice an apparent shift in their vitamin D result or categorical classification coinciding with a change in testing laboratory or assay platform, and without context may reasonably interpret this as a genuine change in their vitamin D status rather than a measurement artefact, potentially prompting

unnecessary concern or an unwarranted self-directed change in supplementation.^[8]

A brief, plain-language explanatory note accompanying a result generated shortly after a known platform change, or available on request, explaining that assay changes can cause an apparent shift in reported values without reflecting a genuine change in vitamin D status, would extend the clinician-facing communication recommendation made earlier in this review directly to patients accessing their own results.

Age and Population-Specific Considerations

The clinical significance of threshold-adjacent misclassification discussed throughout this review may vary across different patient populations, and several age- and population-specific considerations are worth noting explicitly.

Older adults, pregnant women, and patients with conditions affecting bone metabolism are often subject to more frequent vitamin D monitoring and more consequential clinical decisions based on the categorical classification, such as supplementation dose adjustment, meaning the practical impact of threshold-adjacent misclassification is plausibly greater in these populations than in a general population screened less frequently and with less clinical action contingent on the specific threshold classification.^[2]

Populations with naturally lower circulating vitamin D binding protein concentrations, including some pregnant women and patients with certain chronic conditions, may also be disproportionately affected by the matrix-effect-related bias mechanisms discussed earlier, an interaction between population characteristics and assay-specific bias that the aggregate bias figures reported in most method comparison studies, including those reviewed here, are not typically positioned to characterise.

Interaction with Point-of-Care and Direct-to-Consumer Vitamin D Testing

The growth of point-of-care and direct-to-consumer 25-hydroxyvitamin D testing, operating outside traditional central laboratory oversight, introduces a further dimension to the inter-platform bias concern characterised throughout this review.

Point-of-care and direct-to-consumer platforms are not uniformly subject to the same VDSP certification requirements or quality oversight as central laboratory platforms, and published method comparison data for these newer testing modalities against VDSP-certified reference platforms remains considerably more limited than for established central laboratory platforms, meaning the bias ranges characterised in this review's predominantly central-laboratory-focused literature may not fully capture the bias magnitude relevant to these increasingly used alternative testing modalities.^[7]

A consumer receiving a direct-to-consumer vitamin D result classified as deficient or sufficient has essentially no visible means of knowing which assay platform produced their result or

how that platform's bias characteristics compare with the VDSP-certified platforms characterised in published method comparison literature, an information gap considerably larger than even the platform-transparency gap this review identifies for conventional central laboratory testing.

Why This Review Focused Specifically on Threshold-Adjacent Bias

This review's specific focus on bias at the 20 ng/mL and 30 ng/mL thresholds, rather than bias averaged across the full measurement range, reflects a deliberate methodological choice worth making explicit.

Most published method comparison studies report bias summary statistics, such as mean bias or Bland-Altman limits of agreement, calculated across the full range of concentrations included in the comparison, a summary that can obscure concentration-dependent bias variation of the kind several included studies specifically reported, where bias magnitude differed meaningfully between lower and higher concentrations within the same platform pair.^[6]

By specifically extracting or estimating bias at the two clinically relevant threshold concentrations rather than relying on a single full-range summary statistic, this review aimed to characterise the bias dimension most directly relevant to diagnostic misclassification risk, which is this review's central practical concern, rather than overall method agreement across the full measurement range, which is a related but distinct analytical performance question.

Interdisciplinary Coordination Between Laboratory and Endocrinology Services

Effective management of the threshold-adjacent classification uncertainty characterised in this review depends on sustained coordination between laboratory medicine, which generates and reports the categorical classification, and endocrinology and primary care services, which act on it clinically.

Joint development of local reporting conventions -- such as the intermediate reporting band proposed earlier in this review -- between laboratory and referring clinical services, rather than the laboratory unilaterally determining reporting format without clinical input, would help ensure the resulting reporting convention is both analytically sound and practically useful at the point of clinical decision-making.^[8]

Laboratories introducing a platform change affecting known threshold-specific bias characteristics should proactively notify major referring endocrinology and primary care services in advance of the transition, rather than relying on individual clinicians to notice and correctly interpret an apparent longitudinal trend shift after the fact, extending the communication recommendation made earlier in this review into a concrete, proactive institutional practice.

Practical Checklist for Clinicians Interpreting Threshold-Adjacent Results

Drawing together this review's findings into a form directly usable by clinicians, we propose the following practical checklist for interpreting a 25-hydroxyvitamin D result close to the deficiency or insufficiency threshold.

Note whether the result falls within approximately 2-3 ng/mL of a threshold, the range within which this review's findings suggest platform-related bias could plausibly change the categorical classification; where available, check whether the reporting laboratory has recently changed assay platform, which may explain an otherwise unexpected shift in a patient's longitudinal trend; and treat a borderline categorical classification as informative but not absolute, particularly when clinical management hinges significantly on which side of the threshold the result falls.^{[9][10]}

For patients where the categorical classification materially changes management, such as initiating or escalating supplementation, consider repeat testing or, where available, confirmation via an LC-MS/MS-based platform generally considered to more closely approximate the reference measurement procedure, particularly for a result sitting very close to a threshold on a platform without extensive published threshold-specific bias characterisation.

Regulatory Context for Assay Certification and Labelling

Regulatory frameworks governing in vitro diagnostic assay approval and labelling generally require manufacturers to demonstrate analytical performance against an appropriate comparator, but the specific regulatory expectations around threshold-specific bias disclosure for assays with narrow clinically relevant diagnostic bands, such as vitamin D, vary considerably between jurisdictions.

Some jurisdictions' regulatory frameworks have moved toward requiring more granular, concentration-specific performance data within product labelling and documentation rather than a single overall accuracy summary, a direction consistent with this review's recommendation that manufacturers publish threshold-specific bias data, though the extent to which equivalent requirements apply uniformly across all markets where a given VDSP-certified platform is sold is beyond the scope of this review to characterise comprehensively.^{[1][5]}

Why This Review Reached a Systematic Rather Than Narrative Approach

The decision to conduct this review systematically, with a defined search and inclusion methodology across 28 studies, rather than as a narrative summary of selected illustrative studies, reflects the practical aim of producing threshold-specific bias range estimates that clinicians and laboratories could reasonably rely on as representative of the published literature rather than as a curated selection potentially subject to selection bias toward the most dramatic reported discrepancies.

A systematic approach also allowed the separate analysis of certified-versus-certified and certified-versus-non-certified comparisons presented earlier in this review, a stratification that would have been difficult to perform reliably on a smaller, non-systematically selected set of illustrative studies, and which this review considers one of its more clinically useful specific findings given how directly it speaks to whether VDSP certification alone is sufficient reassurance against threshold-adjacent misclassification risk.^{[3][4]}

Summary of Recommendations for Implementing Laboratories

Implementing laboratories seeking to act on this review's findings may find it useful to sequence the recommendations made throughout this report roughly in order of implementation ease, beginning with the lowest-cost reporting and communication changes before considering platform-level procurement decisions.

Reporting the numeric result alongside the categorical classification, and documenting current platform certification status and known bias characteristics within the local quality management system, can typically be implemented immediately at negligible cost; introducing an intermediate reporting band for threshold-adjacent results and proactively notifying referring services ahead of a planned platform change require somewhat more coordination but remain comparatively low-cost; while weighting threshold-specific bias performance in future platform procurement decisions is necessarily a longer-term consideration tied to the laboratory's existing equipment replacement cycle.^{[7][9]}

Reflections on Standardisation as an Ongoing Rather Than Completed Process

A broader reflection emerging from this review is that assay standardisation is more accurately understood as an ongoing process requiring continued monitoring and refinement than as a one-time achievement completed once a certification programme is established and manufacturers begin certifying platforms against it.

The VDSP programme has demonstrably reduced inter-platform variability relative to the pre-standardisation era, a genuine and clinically meaningful achievement, but this review's finding that residual bias remains clinically significant at the specific concentrations that matter most for diagnostic

classification suggests that declaring standardisation complete, rather than continuing to refine certification requirements toward concentration-specific rather than only overall accuracy targets, would leave a meaningful and actionable source of diagnostic uncertainty unaddressed.^{[1][7]}

Why This Review Centred Threshold Proximity Rather Than Overall Accuracy

This review's consistent emphasis on bias specifically at the diagnostic thresholds, rather than overall method accuracy across the measurement range, reflects a deliberate editorial choice to centre the analysis on the dimension most directly actionable by clinicians and laboratories managing individual patient classification decisions, rather than the dimension most commonly reported in manufacturer validation summaries.

An assay platform can report excellent overall correlation and low mean bias across its full validated measurement range while still carrying the kind of threshold-localised bias this review documents, and a laboratory or clinician relying solely on a manufacturer's overall accuracy summary, without threshold-specific context, could reasonably but mistakenly conclude that diagnostic misclassification risk at the clinically relevant thresholds is negligible when the published method comparison literature reviewed here suggests otherwise.^{[4][6]}

Conclusion

Inter-platform bias in 25-hydroxyvitamin D measurement at the clinical deficiency and insufficiency thresholds ranged from -15% to +12% across VDSP-era studies reviewed, sufficient to reclassify a meaningful minority of borderline patients depending on assay platform. Clinicians interpreting a threshold-adjacent result, and laboratories transitioning between assay platforms, should account for this residual bias rather than treating a categorical deficiency classification as platform-independent.

References

1. Sempos CT, Vesper HW, Phinney KW, et al. "Vitamin D status as an international issue: national surveys and the problem of standardization". *Scand J Clin Lab Invest Suppl* 243(2012): 32-40.
2. Holick MF, Binkley NC, Bischoff-Ferrari HA, et al. "Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline". *J Clin Endocrinol Metab* 96(2011): 1911-1930.
3. Binkley N, Sempos CT. "Standardizing vitamin D assays: the way forward". *J Bone Miner Res* 29(2014): 1709-1714.
4. Cashman KD, Kiely M, Kinsella M, et al. "Evaluation of Vitamin D Standardization Program protocols for standardizing serum 25-hydroxyvitamin D data: a case study of the program's potential for a seasonal cohort". *Am J Clin Nutr* 97(2013): 1235-1242.
5. Sempos CT, Betz JM, Camara JE, et al. "General steps to standardize the laboratory measurement of serum total 25-hydroxyvitamin D". *J AOAC Int* 100(2017): 1230-1233.
6. Wise SA, Phinney KW, Tai SS, et al. "Baseline assessment of 25-hydroxyvitamin D assay performance: a vitamin D standardization program study". *J AOAC Int* 100(2017): 1244-1252.

7. Binkley N, Sempos CT, Vitamin D Standardization Program (VDSP). "Standardizing vitamin D assays: the way forward". J Bone Miner Res 29(2014): 1709-1714.
8. Cashman KD, Dowling KG, Skrabakova Z, et al. "Vitamin D deficiency in Europe: pandemic?". Am J Clin Nutr 103(2016): 1033-1044.
9. Enko D, Fridrich L, Rezanka E, et al. "25-hydroxy-vitamin D status: limitations in comparison and clinical interpretation of serum-level variations due to the immunoassay used". Int J Lab Hematol 36(2014): e57-e60.
10. Farrell CJ, Martin S, McWhinney B, et al. "State-of-the-art vitamin D assays: a comparison of automated immunoassays with liquid chromatography-tandem mass spectrometry methods". Clin Chem 58(2012): 531-542.

Citation: Solveig Braatveit, Nadia Bellofiore-Trani, Jonas Wredenhagen. "Vitamin D Assay Standardisation and the Deficiency Threshold: How Method Choice Moves the Diagnosis." J Curr Trends Clin Med Lab Biochem (2026);4: 009-016