



**Universiteit
Leiden**
The Netherlands

We need to talk about the analytical performance of our laboratory developed clinical LC-MS/MS tests, and start separating the wheat from the chaff

Dirks, N.F.; Ackermans, M.T.; Martens, F.; Cobbaert, C.M.; Jonge, R. de; Heijboer, A.C.

Citation

Dirks, N. F., Ackermans, M. T., Martens, F., Cobbaert, C. M., Jonge, R. de, & Heijboer, A. C. (2021). We need to talk about the analytical performance of our laboratory developed clinical LC-MS/MS tests, and start separating the wheat from the chaff. *Clinica Chimica Acta*, 514, 80-83. doi:10.1016/j.cca.2020.12.020

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3212819>

Note: To cite this publication please use the final published version (if applicable).



We need to talk about the analytical performance of our laboratory developed clinical LC-MS/MS tests, and start separating the wheat from the chaff

Niek F. Dirks^a, Mariëtte T Ackermans^b, Frans Martens^a, Christa M. Cobbaert^c, Robert de Jonge^{a,b}, Annemieke C. Heijboer^{a,b,*}

^a Amsterdam UMC, Vrije Universiteit Amsterdam, Endocrine Laboratory, Department of Clinical Chemistry, Amsterdam Gastroenterology & Metabolism, Amsterdam, Netherlands

^b Amsterdam UMC, University of Amsterdam, Endocrine Laboratory, Department of Clinical Chemistry, Amsterdam, Netherlands

^c Leiden University Medical Center, Department of Clinical Chemistry and Laboratory Medicine, Leiden, Netherlands

ARTICLE INFO

Keywords:

Laboratory Developed Test (LDT)
IVDR
Quality
Method performance
Method validation

ABSTRACT

With the upcoming EU regulation on the use of in-vitro diagnostic devices, a critical evaluation of the current status of our in-house developed LC-MS/MS methods is timely and of great relevance. Recently, much attention has been devoted to the need for better specification of analytical and clinical performance. Appropriate reporting of the actual achieved analytical performance is an important determinant of the clinical performance and subsequent clinical effectiveness of a test. We advocate for the application of CLSI C62-A guidelines for method validation and suggest some adaptations for analytical validation of in-house developed LC-MS/MS methods for endogenous substances. Additionally, we underline the importance of well-equipped reviewers and standardized method description, including the presentation of figural evidence of obtained method performance. Achieving this ensures future quality of our in-house developed LC-MS/MS methods.

More and more, mainly academic, clinical laboratories nowadays develop and validate their own LC-MS/MS assays, replacing the traditionally used immunoassays. Those in-house developed LC-MS/MS methods are so-called Laboratory Developed Tests (LDT). According to the Food and Drug Administration (FDA), “an LDT is a type of in vitro diagnostic test that is designed, manufactured and used within a single laboratory” [1]. The new EU regulation (2017/746) on in-vitro diagnostic devices (IVDR) states that LDTs may only be used if there is no equivalent (commercial) device available on the market with an appropriate level of analytical and clinical performance [2]. And so, in anticipation of the new regulation, the question we have to ask ourselves is: Do our current LDTs perform well enough to make the cut? The answer depends on the LDT’s clinical effectiveness, its ability to improve health outcomes over existing tests. Recent publications have focused on the need for better analytical and clinical performance specifications, both important attributes of a method’s clinical effectiveness [3–5]. Here, we would like to stress that presentation of the acquired analytical performance also needs further attention in order to increase the quality

of our LDTs.

Determining the actual analytical performance of a published LC-MS/MS-based LDT is difficult. When searching for a new method for simultaneous quantification of vitamin D metabolites, a number of scientific publications did not describe the analytical conditions in sufficient detail to reproduce the same results, at least in our hands. The publications not only lacked certain details, but they also presented them inconsistently, and did not offer figural evidence of their findings. Although reproducing a published LC-MS/MS-based LDT may seem like an easy task compared to starting from scratch, more often than not, the simple task of reproduction turns into a painful and protracted quest to duplicate someone else’s research. Of course, circumstances may differ to a large extent between locations, may change in time, or with somewhat different equipment or materials, making some adjustments inevitable. For example, when translating an application from a Waters platform to a Sciex platform the source parameters change. Waters’ cone voltage is similar to Sciex’ declustering potential but it requires new empirical optimization to get it just right. Nevertheless, with a

* Corresponding author at: Amsterdam UMC, Department of Clinical Chemistry, Endocrine Laboratory K2-280, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.

E-mail address: a.heijboer@amsterdamumc.nl (A.C. Heijboer).

<https://doi.org/10.1016/j.cca.2020.12.020>

Received 25 September 2020; Received in revised form 27 November 2020; Accepted 17 December 2020

Available online 22 December 2020

0009-8981/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

comprehensive description of method details it should not be a bridge too far. Vogeser et al. have proposed a minimal set of fundamental characteristics that should be described and an additional set of variable characteristics that could be described [6]. Lamentably, current method descriptions are often too brief, inconsistent or even technically impossible and turn out to be incompatible with the presented results. Presenting chromatograms of compounds obtained after injecting highly concentrated neat standard solutions and the subsequent assertion of baseline resolution is deceitful representation of results. Obtaining baseline resolution in neat, concentrated solutions, free of possible interfering substances is obviously easier than to accomplish this in human blood samples. Similarly, extra peaks, matrix effects or ion suppression will not be exposed, but rather disguised. By the same token, establishing a limit of detection (LOD) and a lower limit of quantitation (LLOQ) based upon extrapolation of calibrators far above these calculated limits does not provide information on the method's abilities in patient samples. This sloppiness, the absence of information or presence of little inconsistencies—which are difficult to recognize while reading the article and only materialize when the actual effort of reproduction has begun—result in LC-MS/MS-based LDTs with inferior actual performances than those described.

The presence of bad apples among the many LC-MS/MS-based LDTs for 25(OH)D in operation today also becomes apparent when looking at the average quality. External quality assessment (EQA) schemes monitor the performance of all participating methods measuring a specific analyte by comparing their results to an all method mean or target value. This allows for comparison of method variabilities (CVs) and, when a target value is used, method trueness. The Vitamin D EQA scheme (DEQAS) for 25-hydroxyvitamin D assays distributes samples from healthy donors with target values provided by the National Institute of Standards and Technology (NIST) reference method procedure. The scheme shows that while the obtained values in serum of healthy subjects with LC-MS/MS methods are, on average, very close to the NIST target values, interlaboratory CVs vary as much as, or more than, the interlaboratory CVs for the four most used (automated) immunoassays (Fig. 1). This means LC-MS/MS, as a technique, is capable of measuring 25-hydroxyvitamin D more accurately than most immunoassays, and, on average, users indeed measure very close to the target values. Of course, we do have to bear in mind that the DEQAS samples derive from healthy subjects and multiple studies have shown immunoassays underperforming in specific patient populations [7–10]. Even so, while the

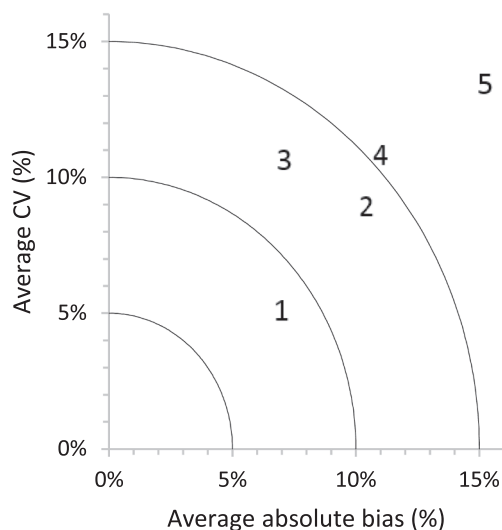


Fig. 1. Average absolute bias and average CV of the five most used assays for 25-hydroxyvitamin D in DEQAS samples from 2018. 1 = Abbot Architect – New (5P02); 2 = Diasorin Liaison Total; 3 = LC-MS/MS; 4 = Roche Total 25(OH)D; 5 = Siemens ADVIA Centaur.

immunoassays had to meet strict requirements before CE marking was obtained, the in-house developed LC-MS/MS-based LDTs are not regulated, and quality largely depends on the varying expertise of the operating laboratory. The DEQAS results clearly show opportunities for improvement: further standardization of 25(OH)D LC-MS/MS method. While the first steps for improvement have been taken with the introduction of certified reference material and reference method procedures, we should now focus on more strict adherence to the practices for assay development, validation and post-implementation monitoring, especially in light of the imminent more stringent regulation for LDTs [11–13]. This calls for leadership of the laboratory specialists to provide guidance in this more demanding landscape [14].

To aid in validation of new bioanalytical methods, the European Medicines Agency (EMA), the Food and Drug Administration (FDA) and the Clinical and Laboratory Standard Institute (CLSI) have all published validation guidelines that meet the requirements set by the IVDR for LDTs [1,2,15,16]. The EMA and FDA guidelines were drafted for the validation of bioanalytical methods measuring drug concentrations in the drug discovery process or during clinical trials. These are less well-suited for use in clinical method validation of endogenous substances, as these are designed for the validation of methods determining exogenous substances, for which obtaining negative samples of the appropriate biological matrix is fairly simple. The CLSI C62-A document is specifically tailored to be used for clinical LC-MS/MS assays and builds upon the backbone of the FDA and EMA guidelines. It discusses minimal performance specifications, validation practices and post-implementation requirements. It still does not, unfortunately, fully acknowledge the difficulty of obtaining negative samples for the validation of methods measuring endogenous substances like 25-hydroxyvitamin D. This might be a reason for the unwarranted differences in method performance between the 25-hydroxyvitamin D LC-MS/MS-based LDTs revealed by DEQAS, as none of the guidelines are easily interpreted in light of this complication and adherence might therefore be poor. Nonetheless, with some adaptations, the prescribed validation assessments from CLSI C62-A can be followed without a negative matrix. The documents discuss the following performance parameters: Calibration standards, accuracy/trueness, imprecision, sensitivity, matrix effects, specificity/selectivity/interferences, carryover, stability, dilutions, recovery, QC and linearity. The guidelines on assessment of imprecision, sensitivity, stability, recovery, QC and linearity do not require negative samples and are straightforward to follow. The preparation of adequate calibration standards in a surrogate matrix is vital when no negative matrix is available. The traceability of the calibration standards to certified reference material (CRM) is to be mentioned, as well as a description of the characteristics of the surrogate matrix, such as pH, specific gravity and protein concentration. Once a surrogate matrix has been selected, it should be used throughout the process of method performance validation and prove to behave similar. For determination of accuracy or trueness, CLSI C62-A prescribes to perform at least two out of three of the following validation practices: comparison to a reference measurement procedure (RMP), analysis of commutable CRMs or spike and recovery analysis. While spike and recovery analysis is more complex without a negative matrix, as the amount of the analyte endogenously present needs to be subtracted, the other two practices can be performed without problem. Of course, RMPs and CRMs do not exist for every analyte, which means other ways of determining method accuracy should be explored to the best of abilities. Showing method performance in an EQA scheme or comparison with another similar method are well-advised as an assessment of trueness. To assess possible matrix effects, CLSI C62-A suggests comparing native matrix samples spiked with analyte post-extraction versus analyte spiked in neat solution. A slightly more complex procedure is required when negative matrix samples are not available. The procedure as provided by Matuszewski can be slightly modified to assess both recovery and possible matrix effects [17]. The original procedure prescribes measuring three sets of samples with the analyte (also to be done

for the internal standard), one in a neat solution, one in which the analyte was added at the beginning of the extraction process, and one in which the analyte was added after extraction of the samples. This way the matrix effect (ME) for each analyte can be calculated as (analyte spiked in after extraction)/(analyte in neat solution), the recovery efficiency (RE) as (analyte spiked in before extraction)/(analyte spiked in after extraction) and the overall process efficiency (PE) as $ME \times RE$. Likewise, by using the analyte to internal standard ratio as the ‘analyte’ in the calculations, an internal standard normalized ME, RE and PE can be calculated. A sample that contains an unknown amount of endogenous analyte can be fortified with extra analyte and measured both with and without the spiked amount in the three sets. Using this strategy, the CV of matrix effects, recovery efficiency and process efficiency in six samples, covering the quantitation range, could be assessed. To study specificity/selectivity/interferences, the guidelines suggest evaluating a high concentration of potential interfering substances in matrix with and without analyte present. Alternatively, we could evaluate samples with or without a high concentration of potential interfering substances present, which does not require negative samples. To assess the presence of carryover, CLSI C62-A suggests injecting extracted negative samples after samples with increasing concentrations of the analyte. The extracted negative sample may be replaced by a sample with a very low amount of analyte and a negative surrogate matrix. Similarly, dilution integrity can be easily assessed with the surrogate negative matrix.

With the above suggested adaptations, summarized in Table 1, the CLSI C62-A guidelines can be translated to serve for method validation of LC-MS/MS-based LDTs for endogenous substances without the need for negative samples. Hopefully, adhering to these suggestions will enable anyone to perform proper validation of their LC-MS/MS-based LDTs, raising the overall quality of methods and improving the average variability. Notwithstanding, to additionally allow for uncomplicated and successful reproduction of published LC-MS/MS-based LDTs we also need to encourage uniform representation of method performance. To avoid concluding the published LC-MS/MS method you have been trying to reproduce is erroneous at the eleventh hour, journals will have to be more stringent on the minimal requirements for method description and should solicit well-acquainted reviewers in order to uphold a certain standard for method validation description, enabling easy method reproduction and hampering misrepresentation of results. We should therefore not only establish guidelines for validation of analytical performance, but also minimal criteria for description of that validation in literature. To ensure adherence to the guidelines on analytical performance validation, stating the CLSI C62-A guidelines, or any other guideline, has been followed, should not be enough. Neither should just providing a table with figures corresponding to the different validation parameters. What is required to ensure validation of method performance has been properly executed, is figural evidence, particularly raw extracted ion chromatograms of genuine samples and not processed peak chromatograms from secondary software or chromatograms of calibrators. Chromatograms of samples showing the method is actually able to measure at the level of the LLOQ and proving baseline resolution is indeed accomplished. Recently, a format for standardized method description of clinical LC-MS/MS-based LDT has been proposed which should facilitate easy reproducibility [6].

Improving the overall analytical performance and the transparency of reporting of our LC-MS/MS-based LDTs may be the only way of proving that we are able to develop assays ourselves that meet the more stringent requirements of the new EU IVDR, effective May 26, 2022. We need to take action in order to do so and start separating the wheat from the chaff. This way, we can guarantee that our clinical LC-MS/MS-based LDTs are fit-for-purpose, and can be relied on for adequate interpretation and diagnosis in patient care.

In conclusion, we propose that all clinical LC-MS/MS-based LDTs should be validated according to the CLSI C62-A guideline. We have suggested adaptations to the CLSI C62-A guideline for analytical validation of LC-MS/MS-based LDTs for endogenous substances. We believe

Table 1
Suggested adaptations to adopt the CLSI C62-A guidelines for method validation of LC-MS/MS methods for endogenous substances.

LC-MS/MS method validation parameters	CLSI C62-A Guidelines	Our recommendations when negative samples are not available
Calibration standards	Select blank, zero, and 6–8 calibration standards (across the quantitation range), prepared in the same biological matrix as the samples.	Select surrogate negative matrix that resembles the patient matrix and test it throughout validation experiments.
Accuracy/trueness	Perform at least two of the following: Compare to an RMP using ≥ 40 samples, analysis of CRMs and/or spike and recovery analysis	Compare to EQA scheme or another similar method is also a legitimate option.
Imprecision	Imprecision should be assessed using 20 determinations across the quantitation range.	No adaptations necessary.
Sensitivity	LLOQ should be established using ≥ 40 replicates from 3 to 5 different samples over ≥ 5 runs	No adaptations necessary.
Matrix effects	Compare peak areas for 5 native matrix samples spiked with analyte postextraction vs analyte spiked into neat solution across the quantitation range.	Use both fortified and unfortified samples in matrix effect experiments (Matuszewski, see text).
Specificity/selectivity/interferences	Evaluate a high concentration of potential interferent in matrix (preferably a patient sample) with and without analyte present	Evaluate samples with or without a high concentration of potential interfering substances present.
Carryover	Inject extracted negative samples after samples with increasing concentrations of analyte	Inject a sample with a very low amount of analyte and a negative surrogate matrix sample after injecting a very high sample.
Stability	Recommended experiments for freeze and thaw stability, short-term temperature stability, long-term stability, stock solution stability, and postpreparative stability (i. e., time in autosampler); experiments should be done with ≥ 3 aliquots of a low and high concentration	No adaptations necessary.
Dilutions	Verify dilutions of ≥ 5 separate replicates.	Dilute with negative surrogate matrix.
Recovery	Compare the analytical results for extracted samples at 3 concentrations (low, medium, and high) with unextracted standards that represent 100% recovery. For endogenous analytes spiked into matrix that is not blank, the technique of standard additions should be used. Proficiency testing, QC, and external reference samples should be included in recovery experiments.	No adaptations necessary.
QC	Test ≥ 3 QC concentrations (1 around $3 \times$ the LLOQ, 1 in midrange, and 1 near the upper limit of quantification’ in duplicate.	No adaptations necessary.
Linearity	Analyze 9–11 concentrations with 2–4 replicates each. Evaluate linearity with the	No adaptations necessary.

(continued on next page)

Table 1 (continued)

LC-MS/MS method validation parameters	CLSI C62-A Guidelines	Our recommendations when negative samples are not available
	polynomial regression method. Serial dilutions should be avoided.	

journals can improve LC-MS/MS method publications by soliciting well-equipped reviewers and demanding elaborate method description, such as described by Vogeser et al. [6]. This will improve the overall quality of methods, ease of reproduction, and is paramount in anticipation of the new EU IVDR. Together with a new generation of laboratory specialists, familiar with the latest test evaluation requirements, maintaining access to our LDTs is very much achievable with the proper evidence of their effectiveness.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank dr. Emma L. Walker for allowing us to use DEQAS data in this paper.

References

- [1] Draft Guidance for Industry, Food and Drug Administration Staff, and Clinical Laboratories. FDA Notification and Medical Device Reporting for Laboratory Developed Tests (LDTs), 2014.
- [2] REGULATION (EU) 2017/746 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, Official Journal of the European Union, 2017.
- [3] S. Sandberg, C.G. Fraser, et al., Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine, *Clin. Chem. Lab. Med.* 53 (6) (2015) 833–835.
- [4] P.J. Monaghan, S. Robinson, et al., Practical guide for identifying unmet clinical needs for biomarkers, *EJIFCC* 29 (2) (2018) 129–137.
- [5] S.J. Lord, A. St John, et al., Setting clinical performance specifications to develop and evaluate biomarkers for clinical use, *Ann. Clin. Biochem.* 56 (5) (2019) 527–535.
- [6] M. Vogeser, C. Schuster, A proposal to standardize the description of LC-MS-based measurement methods in laboratory medicine, *Clin. Mass Spectr.* 13 (2019) 36–38.
- [7] E. Cavalier, P. Lukas, et al., Analytical and clinical evaluation of the new Fujirebio Lumipulse(R)G non-competitive assay for 25(OH)-vitamin D and three immunoassays for 25(OH)D in healthy subjects, osteoporotic patients, third trimester pregnant women, healthy African subjects, hemodialyzed and intensive care patients, *Clin. Chem. Lab. Med.* 54 (8) (2016) 1347–1355.
- [8] B. Depreter, A.C. Heijboer, M.R. Langlois, Accuracy of three automated 25-hydroxyvitamin D assays in hemodialysis patients, *Clin. Chim. Acta* 415 (2013) 255–260.
- [9] E. Elsenberg, E. Ten Boekel, et al., Standardization of automated 25-hydroxyvitamin D assays: How successful is it? *Clin. Biochem.* 50 (18) (2017) 1126–1130.
- [10] A.C. Heijboer, M.A. Blankenstein, et al., Accuracy of 6 routine 25-hydroxyvitamin D assays: influence of vitamin D binding protein concentration, *Clin. Chem.* 58 (3) (2012) 543–548.
- [11] K.L. Lynch, Accreditation and quality assurance for clinical liquid chromatography-mass spectrometry laboratories, *Clin. Lab. Med.* 38 (3) (2018) 515–526.
- [12] M. Vogeser, C. Seger, Quality management in clinical application of mass spectrometry measurement systems, *Clin. Biochem.* 49 (13–14) (2016) 947–954.
- [13] R.J.A.C. Roelofs-de Beer, J. Wielders, et al., Validation and verification of examination procedures in medical laboratories: opinion of the EFLM Working Group Accreditation and ISO/CEN standards (WG-A/ISO) on dealing with ISO 15189:2012 demands for method verification and validation, *Clin. Chem. Lab. Med.* 58 (3) (2019) 361–367.
- [14] A.R. Horvath, S.J. Lord, et al., From biomarkers to medical tests: The changing landscape of test evaluation, *Clin. Chim. Acta* 427 (2014) 49–57.
- [15] K.L. Lynch, CLSI C62-A: A new standard for clinical mass spectrometry, *Clin. Chem.* 62 (1) (2016) 24–29.
- [16] EMA, Guideline on Bioanalytical Method Validation, EMA, Committee for Medicinal Products for Human Use (CHMP), London, UK, 2011.
- [17] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, *Anal. Chem.* 75 (13) (2003) 3019–3030.