

Product availability may vary from country to country and is subject to varying regulatory requirements. Please contact your local representative for availability.

#### Local Contact Information

**Siemens Healthcare**  
Laboratory Diagnostics  
511 Benedict Avenue  
Tarrytown, NY 10591-5005  
USA  
Phone: +1 914-631-8000  
[siemens.com/healthcare](http://siemens.com/healthcare)

**Siemens Healthcare Headquarters**  
Siemens Healthcare GmbH  
Henkestr. 127  
91052 Erlangen  
Germany  
Phone: +49 9131 84-0  
[siemens.com/healthcare](http://siemens.com/healthcare)

**SIEMENS**  
**Healthineers**



[siemens.com/vitaminD](http://siemens.com/vitaminD)

# Understanding the Challenges of Vitamin D Harmonization

Historically, there has been variation in vitamin D values amongst different manufacturers and methodologies due to the lack of an official vitamin D standard. The following experts from France, Germany, Italy, the United States, and the United Kingdom address questions regarding the impact of liquid chromatography-mass spectrometry (LC-MS/MS) on the present and future of vitamin D harmonization, how laboratories can feel confident in their results, and if assay- or country-specific cutoffs are needed.

**Professor William Fraser, MD**  
Norwich, United Kingdom

**Deborah French, PhD, and Alan Wu, PhD**  
San Francisco, USA

**Markus Herrmann, MD**  
Bolzano, Italy; Sydney, Australia

**Jean-Claude Souberbielle, PhD**  
Paris, France

**Pablo Roth**  
Heidelberg, Germany



## How can we improve the harmonization of LC-MS/MS? What is the role of LC-MS/MS in terms of assay harmonization?

### Fraser:

Harmonization of LC-MS/MS results can be achieved by all laboratories taking part in an approved external quality assurance scheme (e.g., Vitamin D External Quality Assessment Scheme [DEQAS]) and by all laboratories utilizing the same calibrator material assigned with specific concentrations (e.g., National Institute of Standards & Technology [NIST] aligned material). This is not necessarily producing “correct” results, but will decrease the variability in results obtained between laboratories. LC-MS/MS methods aligned using a suitable calibrator material could be utilized to produce secondary calibrator material with 25(OH) vitamin D concentrations across a wider concentration range that would be suitable for different methods.

### French and Wu:

The harmonization of LC-MS/MS methods to measure 25(OH) vitamin D could be improved if all methods were developed to the same standards in terms of precision, linearity, and accuracy, if all laboratories used the same calibration material and used the same mass to charge ratio for monitoring this analyte.

LC-MS/MS could improve 25(OH) vitamin D assay harmonization if it was used as the gold standard, the results of which all other developed assays should be compared to.

### Herrmann:

The main problem with LC-MS/MS methods is their diversity. They can differ in sample preparation, calibration, instrumentation, ionization, and instrument settings. Some laboratories perform a liquid-liquid extraction with subsequent drying and reconstitution of the sample. Others use a simple acetonitrile protein crash to release 25(OH) vitamin D and to eliminate serum proteins. Calibration can be performed with commercial calibrators and self-prepared calibrators. Furthermore, the LC protocol can be designed to resolve the 3-epi-25(OH)D<sub>3</sub> or not. Ionization can also differ. Although electrospray ionization is most commonly used, other techniques—such as atmospheric pressure photoionization—are used as well.

Harmonization of LC-MS/MS methods will require standardization of all these aspects. Extraction, chromatography, ionization, and detection procedures need to be matched to the assay requirements and similar procedures should be used in all laboratories. One of the key factors requiring harmonization is the chromatography procedure. Longer columns and separation times can enhance the detection of small amounts of analyte by removing contaminants and by reducing internal quenching. Furthermore, this allows for the separation of vitamin D isomers. Gradient conditions also need to be harmonized to ensure assay results can be compared effectively.

The recent endorsement of two reference methods for the measurement of 25(OH) vitamin D is a significant step forward. Furthermore, alignment to the NIST 972 material seems to bring LC-MS/MS methods closer together. However, as long as there is no agreement on sample preparation, the most appropriate LC-MS/MS protocol harmonization will not be attained.

### Souberbielle:

Initially, the reference method (which should be an LC-MS/MS method) must be defined urgently, as well as the official reference standard (there are currently several “candidate methods” that produce somewhat different results). Also, the way the samples are “prepared” (purification, extraction, delipidation...) is of paramount importance. Then, a panel of “reference” serums (say 40 to 50 serums) with concentrations ranging from very low to very high natural, unspiked, 25(OH) vitamin D, including some samples containing 25(OH) vitamin D<sub>2</sub> with concentrations measured (possibly in triplicate) with the reference method, should become available to users of 25(OH) vitamin D assays. Once this is done, it should be easy for every user of an LC-MS/MS method to calibrate their own method against the reference method.

Consequently, what is important in terms of harmonization (not only of LC-MS/MS methods) is to urgently define a consensual reference method (and consensual reference laboratories).

### Roth:

Tandem Mass Spectrometry (LC-MS/MS) is the method of choice for vitamin D testing. There are some difficulties in having LC-MS/MS for routine analysis due to a high workload. There already exists a reference LC-MS/MS method by Dr. Linda Thienpont (Belgium) that is known as isotope dilution (ID) LC-MS/MS for the quantification of different vitamin D metabolites in their native condition. The NIST Standard Reference Material (SRM) 972 is very helpful to harmonize LC-MS/MS methods in the determination of 25(OH) vitamin D. There are matrix effects of three of the four NIST levels in immunoassays not affecting LC-MS/MS but possibly leading to variable results in immunoassays. According to this, a reference panel of human samples defined by means of LC-MS/MS would be more suitable in harmonizing different immunoassay methods.

Furthermore, there is variability of LC-MS/MS results between different laboratories, and only experienced LC-MS/MS labs should run LC-MS/MS for vitamin D with validated reagents and well-defined procedures. The end user’s knowledge and experience in technology will influence the quality of the LC-MS/MS results significantly, and therefore utilization of a reference preparation or reference panel is essential. It will not be easy to harmonize between LC-MS/MS methods because many factors are involved, such as the type of column and instrument, as well as type of solvent. If these are standardized, then it will be possible to bring methods together. All of these individual impacting factors will cause variability of test results, and only a well-defined reference method will give opportunities for harmonizing immunoassays.

# How should labs interpret their own DEQAS results? Should they run both CAP and DEQAS? What is the role of the NIST SRM 972?

## Fraser:

Currently DEQAS results are best interpreted by each laboratory by comparing within their specific method groups. The inter-group comparisons are always interesting to observe as they often highlight problems that can exist for a specific method, and these may need to be investigated further (e.g., 3-epimer interference in LC-MS/MS and lack of 25(OH) vitamin D<sub>2</sub> recognition by immunoassay). Running multiple external quality assurance samples defeats the object of such an exercise and one scheme should suffice for each laboratory. Suitable NIST material or NIST aligned material can be used to improve harmonization of particular methods, but there is a problem with the concentration range and the matrix of the NIST material that makes it unacceptable for universal use in all assays.

## French and Wu:

Labs should use their DEQAS results as part of monitoring the stability of their assay over time once it is live for patient testing, since it monitors the accuracy of your results compared to an all method mean and precision of your results compared to your method peer group.

If you are a College of American Pathologists (CAP)-inspected lab, you are, of course, required to run CAP proficiency. The benefit of also running DEQAS is that the samples provided are human serum and not manufactured like the CAP samples, and are therefore less prone to matrix effects. However, both of these programs only monitor precision and how close you are to your method peer group result. The CAP accuracy-based program that is now available for some analytes, including 25(OH) vitamin D, consists of minimally processed human serum and gives an indication of how close your result is to the result of a reference method. At this time, this is not equivalent to proficiency testing, and so participation in CAP or DEQAS would still be necessary for compliance.

As the NIST SRM 972 has certified concentrations of 25(OH) vitamin D<sub>2</sub> and D<sub>3</sub>, it can be used as a material to which you can calibrate your LC-MS/MS assay. Immunoassay manufacturers could also standardize their assay calibration to NIST SRM 972. This is one way in which assays could become more standardized.

## Herrmann:

The all method mean of external QAP programs, such as DEQAS, is typically biased by the dominant method. In many instances this is the DiaSorin LIAISON® method. To avoid this problem, laboratories should compare their results to the results obtained with LC-MS/MS and RIA. Both methods are not significantly affected by matrix effects and may thus provide a solid estimate of bias. However, even these higher order methods may differ in their ability to measure various isoforms of vitamin D and thus give different results for individual samples. Furthermore, laboratories should compare their results with peers using the same method to assess if the test performs as good as in other labs. It is also important to consider how the concentrations of QAP samples relate to the cutoffs used. Lastly, knowledge about the sample matrix is essential for the interpretation of QAP results. Samples that are not based on serum may affect the solubility of vitamin D and interfere with extraction procedures and the assay itself.

I do think that participation in one program is sufficient. Laboratories should ensure that the external quality assurance program that is used suits their needs. Important points to consider are sample composition, transport, turnaround time of results, 25(OH) vitamin D levels that are tested, number of participants, and distribution of the various methods in the program.

The NIST standard reference material has helped to improve the concordance of 25(OH) vitamin methods. LC-MS/MS methods have particularly benefitted from the release of this standard. In a recent study, we have compared two different LC-MS/MS methods performed in independent laboratories in Australia and the USA. Both methods showed excellent agreement. SRM 972 appears to be less useful for the harmonization of immunoassays. This is mainly due to the matrix of SRM 972. Only level one is native serum. Level 2 is diluted with horse serum. Levels 3 and 4 are spiked with 25(OH) D<sub>2</sub> and level 4 with 3-epi-25(OH)D<sub>3</sub>, respectively. LC-MS/MS methods remove proteins before measuring 25(OH) vitamin D and are thus not affected from potential matrix effects provided that an adequate carrier is present to ensure availability of analyte prior to extraction into organic solvent. Furthermore, LC-MS/MS methods have the capability of detecting D<sub>2</sub>, D<sub>3</sub>, and 3-epi-25(OH)D<sub>3</sub>.

## Souberbielle:

I am not sure I understand the question. Currently the DEQAS results are compared to the "all methods trimmed mean." If the reference method is defined soon, I think that DEQAS will give results obtained with this reference method in reference labs. Until that happens, we have to continue as before. Regarding the NIST, I don't remember the numbers of the 4 NIST standards, but I have the notion that only one is a "real" serum. Only this one could be used as a standard, I believe.

## Roth:

Labs should use an available, reliable standard. Currently, the NIST standard is well defined, but unfortunately the current lot is depleted. Hopefully the new one will be available soon. This standard should be an integral part of a well-performed LC-MS/MS analysis. There should be no need to run both DEQAS and CAP. DEQAS has a great deal of experience with EQA and they provide human-based samples. It should be borne in mind that DEQAS is a European scheme and CAP is a U.S. scheme. There may well be differences between U.S. LC-MS/MS methods and EU LC-MS/MS methods, as immunoassay companies claim standardization to LC-MS/MS, but significant differences have been observed not only between immunoassays but also between LC-MS/MS in our lab.

## Where do you see the future of the vitamin D standard going? What is likely to be the consensus standard in the future?



### Fraser:

Suitable commutable material needs to be produced covering a wide range of 25(OH) vitamin D<sub>3</sub> and D<sub>2</sub> concentrations that can be utilized by all assay manufacturers and all LC-MS/MS systems. This may have had the initial concentrations assigned by a gas chromatography-mass spectrometry (GC-MS/MS) system or by a consensus mean of LC-MS/MS users that is satisfactory to all laboratories. This should be produced under the auspices of an IFCC standardization or CAP committee.

### French and Wu:

There is a definite need for a vitamin D standard. The NIST SRM 972 was a great start, but since matrix effects affect the recovery of this standard in some methods, particularly using immunoassays, there is a need for a new standard. It seems like the consensus standard would have to be one consisting of as close to a human serum matrix as possible so that all methods can obtain accurate results when measuring it.

### Herrmann:

This is a difficult question to answer. I think that most manufacturers and laboratories hope for a standard that is exclusively based on human serum and has not been spiked. Furthermore, the role of the 3-epi-25(OH)D<sub>3</sub> requires further investigation. The inclusion of 3-epi-25(OH)D<sub>3</sub> in future SRMS will depend on the outcomes of these further investigations.

### Souberbielle:

Clearly, I don't know (but I do not participate in these discussions). What I know, however, is that it is time to decide (arbitrarily?) which preparation will be the definitive international standard against which the reference method will be calibrated!

### Roth:

An International Reference Preparation (IRP) would be highly desirable, which could be established through the IFCC/WHO for which the primary testing standard should be ID LC-MS/MS. A secondary standard should be derived from the IRP for daily use. This secondary standard could be a defined set of patient samples whose vitamin D values had been tested in the primary method. Many factors can influence the results, such as pools from different countries with variable binding proteins, which could have an effect on separating vitamin D from its binding protein. Also, the type of sample derived from healthy or end-stage renal disease patients could influence results. Although both serum and plasma samples would ideally be used, initially serum would be the best choice of sample type. Different cross-reactivity of antibodies in immunoassays gives rise to variability, too, especially against 25(OH) vitamin D<sub>2</sub> and D<sub>3</sub>, 24,25(OH)<sub>2</sub> vitamin D<sub>3</sub> and D<sub>2</sub>, and the 3-epi-25(OH)D<sub>3</sub> has to be considered for IRP preparation.

*“There is a definite need for a vitamin D standard.”*



## How can labs feel confident in their results given the current variation?

### Fraser:

Confidence will come from improving performance within an external quality assurance scheme and observing a reduction of the variability observed in method comparisons. Variability in results is bound to persist between methods until a suitable means of standardizing all methods and harmonizing methods is available.

### French and Wu:

There is still variation in results between methods, but this has been improving over the past few years. In order to remain confident in their results, labs could participate in the CAP accuracy-based survey and, of course, proficiency testing. Labs could also run patient comparison samples with another lab that uses a different method.

### Herrmann:

Because of the inherent difficulties in measuring 25(OH) vitamin D, some uncertainty is inevitable. However, if labs validate their method of choice against an established RIA or an NIST aligned LC-MS/MS method, they can obtain a good understanding of the capabilities of their assay. In settings with a high proportion of pediatric samples, a comparison against an LC-MS/MS method that separates the 25(OH) vitamin D epimers is preferable. Laboratories should validate their assay at low levels and carefully select the lower limit of the reportable range. In my laboratory, we do not report quantitative results below 20 nmol/L. Based on our experience, quantification down to 20 nmol/L is sufficient in most clinical situations, and comparison studies with RIA and LC-MS/MS have shown that the accuracy of our test is acceptable at levels above 20 nmol/L.

In addition, I would recommend confirming questionable immunoassay results by LC-MS/MS or RIA.

### Souberbielle:

I don't think that the 25(OH)D assays are of poorer value than other assays for other steroid compounds. The problem with harmonization of 25(OH)D results is, however, a more important concern than for assays of other compounds. Indeed, 25(OH)D levels are usually interpreted as being above or below clinical cutoffs (that are not consensual, but this is another debate) that are currently independent of the assay used. If we say, for example, that vitamin D insufficiency corresponds to a 25(OH)D level <30 ng/mL, it is obvious that all assays should be, reasonably, standardized at this concentration (produce almost the same value).

### Roth:

There is one advantage with vitamin D is that an exact "true" value is not as essential as, for example, measuring therapeutic drug values. Insufficiency, sufficiency, and toxic levels have to be defined correctly. There is a need to have a good "estimate" of vitamin D in a patient, and the clinical situation has to be taken into account, too. Factors to be considered include age, demographic data, environment, medication, and underlying diseases. Most immunoassays provide a reliable vitamin D value. Nonetheless, there is definitely a need for proper standardization, improved accuracy and traceability, and harmonization.

Available immunoassays can have a good role.



## Should labs adopt country-specific cutoffs, assay-specific cutoffs or use what is in the literature?

### Fraser:

With the clear variability in methods observed in DEQAS and reported in several publications, assay specific cutoffs should be utilized until there is better harmonization of all methods.

### French and Wu:

As we have seen from the multitude of reports in the literature regarding 25(OH) vitamin D concentrations in different populations, there is a large variation around the world and vitamin D deficiency has been associated with almost every known disease. It is important to note, however, that the consensus guidelines that have been released are all based upon the optimal 25(OH) vitamin D concentration required to support bone health only. In this regard, my recommendation would be to follow these consensus guidelines in setting the lab cutoffs, as long as the method in use can accurately quantify 25(OH) vitamin D at these concentrations.

### Herrmann:

No. The recommended cutoffs for 25(OH) vitamin D do not represent a traditional reference range. They are based on parathyroid hormone (PTH) measurements and clinical observations, such as bone mineral density and fractures. While the prevalence of 25(OH) vitamin D deficiency may vary between countries, any negative impact on biological functions (bone related or not) can be expected to start at the same 25(OH) vitamin level worldwide. Therefore, vitamin D targets should be the same around the globe.

However, it should be acknowledged that the biological systems affected by vitamin D are also influenced by other factors, including nutrition, exercise, stress, medication, and underlying diseases. Depending on these factors the same vitamin D level may have slightly different biological effects. Although some of these factors vary between different geographic regions, our current knowledge is insufficient to advocate for country-specific cutoffs.

I would not argue for assay-specific cutoffs. While such an approach may be scientifically correct, it can be expected that this would cause substantial confusion amongst clinicians. Furthermore, in most cases such cutoffs cannot directly be linked to clinical effects, as appropriate studies are lacking. Harmonizing the performance of 25(OH) vitamin D assays is probably a more feasible approach to overcome the difficulties in 25(OH) vitamin D measurement.

### Souberbielle:

This is an important question. I don't think that there should be country-specific cutoffs. However, this should be the case as (see above) the definition of vitamin D insufficiency/deficiency is not consensual, and, in a given country, the cutoff value depends primarily on what the opinion leaders think. This may become complicated when the opinion leaders think differently (see the recent debate in the U.S. after the IOM report for example). Having said that, and whatever the cutoff chosen, it is clear that the choice of the cutoff depends on what is found in the literature.

In my opinion (only my opinion!) the strongest way to define an optimal 25(OH) vitamin D level is to consider which 25(OH) vitamin D levels were obtained in intervention studies (RCTs) that show positive effects of vitamin D. A good example is the effect of vitamin D supplementation on peripheral fractures as reviewed by Bischoff-Ferrari. When she interpreted the data (and I agree with her), she realized that in most studies (but not all!) the 25(OH) vitamin D assay used was the DiaSorin RIA. She even tried to express the results of the studies that did not use the DiaSorin RIA in «DiaSorin equivalent». What I want to say is that the different cutoffs that are used for defining vitamin D efficiency/insufficiency are based mostly on the DiaSorin RIA. If this RIA does not appear to be perfectly similar with what is given by the reference method, it should be recalibrated—then the cutoffs will have to be revisited.

### Roth:

No assay-specific or country-specific cutoff values should be used. There have been masses of publications in the literature in recent years and these can be used to interpret results. Cutoff values of 30 or 40 ng/mL have been noted as important in observational studies and 30 ng/mL is very reliable, based on the literature, for distinguishing between sufficiency or insufficiency of vitamin D.