

## Workshop Report

# UK Food Standards Agency Workshop Consensus Report: the choice of method for measuring 25-hydroxyvitamin D to estimate vitamin D status for the UK National Diet and Nutrition Survey

Anne de la Hunty<sup>1\*</sup>, A. Michael Wallace<sup>2</sup>, Sigrid Gibson<sup>3</sup>, Heli Viljakainen<sup>4</sup>, Christel Lamberg-Allardt<sup>4</sup> and Margaret Ashwell<sup>1,5</sup>

<sup>1</sup>Ashwell Associates Limited, Ashwell Street, Ashwell, Hertfordshire SG7 5PZ, UK

<sup>2</sup>Department of Clinical Biochemistry, Glasgow Royal Infirmary, Glasgow G4 0SF, UK

<sup>3</sup>Sig-Nurture Limited, Guildford, Surrey GU1 2TF, UK

<sup>4</sup>Department of Food and Environmental Sciences, University of Helsinki, Helsinki FI-00014, Finland

<sup>5</sup>Oxford Brookes University, Oxford OX3 0BP, UK

(Received 25 March 2010 – Revised 20 April 2010 – Accepted 20 April 2010)

The consensus workshop, organised on behalf of the Food Standards Agency, was convened to recommend the most appropriate and secure method for measuring vitamin D status in the UK. Workshop participants (the Expert Panel) were invited on the basis of expertise in current 25-hydroxyvitamin D (25OHD) assays, or expertise in vitamin D nutrition and metabolism or detailed knowledge and experience in the National Diet and Nutrition Survey (NDNS). A decision support matrix, which set out the particular criteria by which the different options were scored and evaluated, was used to structure the discussion. The Expert Panel agreed that five methods for measuring 25OHD should be evaluated according to eleven criteria, selected on the basis of their relevance to the NDNS. All three of the evaluating subgroups of the Expert Panel produced similar total scores over the eleven criteria for the different methods; they scored LC–MS/MS and HPLC–UV similarly highly, while the scores for the immunoassay methods were lower. The Expert Panel recommended that an LC–MS/MS method should be the preferred method for the NDNS. A detailed specification for the method will be required to ensure comparability between NDNS and the National Health and Nutrition Examination Survey in the US facilitating future comparisons. The Expert Panel also recommended that the method should be carried out in a laboratory with appropriate expertise, competency and history of records of good performance. The method should be standardised against the National Institute of Standards and Technology SRM 972. If the recommended LC–MS/MS is adopted, the Expert Panel indicated that the method should be able to discriminate the C-3 epimer of 25OHD<sub>3</sub>, especially if used to measure 25OHD in young infants in the forthcoming Diet and Nutrition Survey of Infants and Young Children, who are known to have high circulating concentrations of the C-3 epimer.

### 25-Hydroxyvitamin D: Vitamin D: Measurement: Immunological methods: HPLC

The Food Standards Agency (FSA) needs to be able to assess the adequacy of vitamin D status in the population as a whole, and in specific subgroups of the population, and to know the proportion of each group falling above and below cut-offs for adequate status. Status is assessed by measuring plasma 25-hydroxyvitamin D (25OHD) in samples from the National Diet and Nutrition Survey (NDNS). Concern about the wide variability of 25OHD measurements made using different methods and in different laboratories<sup>(1–3)</sup> led the Scientific Advisory Committee on Nutrition to advise that ‘there is an urgent need to standardise laboratory methodologies’<sup>(4)</sup>. FSA required a robust method

for use in future NDNS and other national surveys, which would also allow comparisons of circulating concentrations of 25OHD reported in earlier NDNS and other UK population studies with those reported in other countries.

A workshop was convened in November 2009 to recommend the most appropriate and secure method for measuring vitamin D status in the NDNS, and to identify methodological issues which need serious consideration. Workshop participants (the Expert Panel) were invited on the basis of expertise in current 25OHD assays, or expertise in vitamin D nutrition and metabolism or detailed knowledge

**Abbreviations:** 25OHD, 25-hydroxyvitamin D; ALTM, all laboratory trimmed mean; CDC, Centers for Disease Control and Prevention; DEQAS, vitamin D External Quality Assessment Scheme; FSA, Food Standards Agency; NDNS, National Diet and Nutrition Survey; NHANES, National Health and Nutrition Examination Survey; NIST, National Institute of Standards and Technology; RMP, reference measurement procedure; SRM, standard reference material.

\* **Corresponding author:** A. de la Hunty, email [annedelahunty@btinternet.com](mailto:annedelahunty@btinternet.com)

and experience in the NDNS. Background information which had been collated during FSA project N08029 was pre-circulated to participants. The workshop was structured to explore the relevant issues, allow detailed discussion of them and finally come to a consensus recommendation. A decision support matrix, which detailed the criteria by which different options could be scored and evaluated, was used to structure the discussion. The Expert Panel agreed upon the criteria, options and scoring methods.

### The current methodology used in the National Diet and Nutrition Survey

Dr Alison Tedstone, FSA, described the rationale for commissioning the project, and Dr Christine Clewes, MRC-HNR Cambridge, described the current methodology used in NDNS. Assessment of vitamin D status has been included in the NDNS programme since the 1990s, and is planned for inclusion in the Diet and Nutrition Survey of Infants and Young Children. The NDNS currently uses the Diasorin Liaison Total assay for the measurement of 25OHD. There are 800  $\mu$ l of serum available for 25OHD analysis. In 2008–9, 534 samples were analysed for 25OHD out of a study population of 1646 men and women, of whom 575 provided blood samples.

### Current methods available for measuring 25-hydroxyvitamin D

Professor Mike Wallace, Glasgow Royal Infirmary, presented a literature review on the different methods currently available for measuring 25OHD. The literature review identified sixty-eight publications on the measurement of 25OHD since the year 2000, which were categorised as being either quantitative (thirty-five papers), if they included comparisons of the performance of two or more methods, or qualitative (thirty-three papers), if they reviewed or commented on the results of other publications<sup>(5)</sup>.

There are currently two main types of methods used routinely: competitive immunoassays and methods based on chromatographic separation followed by non-immunological direct detection (HPLC-UV and LC-MS/MS). HPLC-UV and LC-MS/MS methods are able to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> independently with good recoveries. Most immunoassays rely on antisera that can detect both 25OHD<sub>2</sub> and 25OHD<sub>3</sub>, but the proportion of 25OHD<sub>2</sub> detected is variable<sup>(6)</sup>. Little detail is available on how commercial immunoassays are standardised. For HPLC-UV and LC-MS/MS, there is a lack of consistency in the way these procedures are standardised. In most HPLC-UV and LC-MS/MS methods, extraction and procedural losses are corrected for by the inclusion of an internal standard. Most procedures have the required sensitivity to identify severe vitamin D deficiency (25OHD < 25 nmol/l). RIA and LC-MS/MS are the more sensitive. In general, the precision of immunoassays, HPLC-UV and LC-MS/MS is comparable, with room for improvement. The main assay weaknesses highlighted in scientific publications were poor specificity, matrix interferences and the lack of a common standard<sup>(3,7,8)</sup>. The main strengths identified were the ability of HPLC-UV and LC-MS/MS to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately, and the convenience and high throughput using immunoassays performed

on automated platforms. A more detailed account of current methodology can be found in a recent review<sup>(5)</sup>.

A. d. I. H., Ashwell Associates, presented the findings of a survey of UK laboratories registered with the vitamin D External Quality Assessment Scheme (DEQAS) to determine users' opinions on the main strengths and weaknesses associated with the different 25OHD methods (A de la Hunty, AM Wallace, S Gibson, *et al.*, unpublished results). This had a response rate of 52%. Respondents scored seventeen different criteria on a scale of 1–4. Strengths were defined as criteria with an average score of 3 or more, while weaknesses were defined as criteria scoring <2. The overall assessment for immunoassays using non-isotopic labels, HPLC-UV and LC-MS/MS, was similar, but LC-MS/MS had more strengths and fewer weaknesses identified by users compared with other methods. The inability to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately was the main criterion distinguishing opinions on immunoassays from HPLC-UV and LC-MS/MS.

### Current variability in 25-hydroxyvitamin D measurements

Julia Jones, DEQAS, reminded workshop participants that DEQAS has been assessing the performance of 25OHD assays internationally for 20 years. Five samples of human serum are distributed quarterly, and individual laboratory performance is determined by comparison of the results with an all laboratory trimmed mean (ALTM), the current target value. The performance of different methods is assessed by comparing the method mean with the ALTM.

The majority of participants in the DEQAS (85%) use commercial non-chromatographic immunoassay kits, but the use of chromatographic methods, HPLC-UV and, in particular, LC-MS/MS, is slowly increasing (14%). Since DEQAS was launched, there has been a gradual improvement in overall inter-laboratory agreement, from a mean annual inter-laboratory CV of over 25% in the 1990s to 16.3% in 2009.

Difference in bias between methods is one cause of poor agreement between methods. For samples distributed in 2008, bias from the ALTM ranged from –9.0% for the Diasorin Liaison method to +9.5% for LC-MS/MS. Approximately one-third of the participants now use the Diasorin Liaison Total 25OHD assay, which with a negative bias could significantly affect the ALTM. Using the mean of the method means as a target value instead, to give equal weighting to each method, raised the target value and resulted in a shift to a more negative bias for all methods. In 2008, the bias had a range of –9.0 to 9.5% (ALTM) and a range of –11.5 to 6.5% (mean of the method means). No overall change in the spread of bias between methods was observed.

DEQAS has investigated whether the use of a common standard would reduce inter-laboratory imprecision of HPLC-UV and LC-MS/MS assays. There was little effect on the former, but agreement among LC-MS/MS users was significantly better when all participants used the same standard, with the mean inter-laboratory imprecision falling from 16.4 to 10.4% for total 25OHD measurements.

### The relative merits of different methods

The relative merits of four of the main methods were outlined by experts in their use. Dr Tom Hill, University of

Cork, reported that the Immunodiagnostic Systems (IDS) enzyme-linked immunoassay (EIA) has shown a mixed performance within DEQAS. The highest mean percentage bias from the ALTM was reported in 2006, with 28.7 and 23% bias being described for the automated and manual IDS EIA, respectively. This high overestimation of 25OHD subsequently resulted in a re-calibration of the standard curve to improve the comparability of the assay. Since 2006, the performance of the manual IDS EIA, in particular, within DEQAS, has been improving, with a mean annual percentage bias from the ALTM in 2009 (three distributions only) being  $-2\%$ . The validity of re-calibrating an assay to agree with the ALTM or another assay is questionable.

The major strengths of non-isotopic immunoassays include the convenience and high throughput capabilities of the assay. Whether the IDS EIA is significantly susceptible to matrix effects is open to debate, but recent evidence points to a particular effect with the automated platform. Problems associated with operator bias and intermittent batch analysis have been identified with the EIA and warrant attention. A major weakness with all immunoassays is the inability to quantify 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately. This may or may not be important depending on what the analyst requires from the assay. In addition, immunoassays vary in the proportion of 25OHD<sub>2</sub> detected. The two metabolites, 25OHD<sub>2</sub> and 25OHD<sub>3</sub>, assess the intake of vitamin D<sub>2</sub> and intake and production, in skin, of vitamin D<sub>3</sub>, respectively. The circulating 25OHD<sub>2</sub> concentration is normally low or undetectable unless the subject is receiving high amounts in the diet or from supplements. The IDS EIA does not appear to underestimate the presence of 25OHD<sub>2</sub> in DEQAS samples, although the IDS RIA does despite the same antibodies being used in both IDS kits.

The Diasorin Liaison Total method, described by Professor Bruce Hollis, Medical University of Southern Carolina, provides readily accessible, uniform and standardised technology for high volume reference laboratories. The automated Diasorin Liaison Total assay is predicated on the Diasorin RIA, which was used to establish the reference ranges and cut-offs which have directed clinical practice for over 20 years. Under-recovery of 25OHD has, however, been observed in the National Institute of Standards and Technology (NIST) controls (SRM 972) to which 25OHD<sub>2</sub> or 25OHD<sub>3</sub> has been added. This problem was not observed in pools containing only endogenous 25OHD (discussed later).

The greatest advantage (and simultaneously one of the major disadvantages) of HPLC-UV, described by Dr Jacqueline Berry, University of Manchester, is that samples have to be extracted before being put onto HPLC-UV. This means that any potentially interfering substances such as proteins and lipids are removed, eliminating any related interference. The second advantage is the ability to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> metabolites separately and accurately. Further advantages include its reliability, low batch-to-batch variation, its consistency between different operators and low limit of detection (5 nmol/l). These advantages are shared with LC-MS/MS, which has similar work-up procedures. The major disadvantage of HPLC-UV is the time-consuming and labour-intensive extraction procedure that limits the number of samples that can be processed per day. HPLC-UV methods can handle about fifty samples per day; however, several

laboratories have now developed methods to automate the extraction process, so the low throughput may not be a problem in the future.

With the UV detection method, there are occasionally substances unrelated to vitamin D that will produce an interfering peak close to that of interest. For example, the C-3 epimer is known to be present at high concentrations in the circulation of some neonates (and also possibly in some adult samples). The C-3 epimer of 25OHD<sub>3</sub> can be resolved using HPLC-UV by adaptation of a reverse-phase column and extending the run time. Different views were expressed by the workshop participants on the utility of quantifying the C-3 epimer, which, to date, has no known function. To gain more information on the C-3 epimer of 25OHD<sub>3</sub>, it will be quantified in National Health and Nutrition Examination Survey (NHANES).

Dr Sandra Rainbow, Northwick Park Hospital, described how LC-MS/MS has, over the past 5 years, become the method of choice for the quantitation of specific steroid-like molecules, including 25OHD, in biological fluids. The technique allows equivalent, or better, detection levels than an immunoassay and greater specificity, and is applicable through semi-automated processing to relatively large sample numbers analysed in batch mode. Several LC-MS/MS methods for the analysis of 25OHD<sub>2</sub> and 25OHD<sub>3</sub> have been published in the last 5 years, which differ in either the sample preparation technique or selection of fragmentation pattern.

The strengths of the LC-MS/MS process include the ability to independently quantify 25OHD<sub>2</sub> and 25OHD<sub>3</sub> and, with suitable adaptation, other related metabolites of interest. Furthermore, sensitivity exceeds that of most immunoassays, and good inter-laboratory precision can be achieved if a common standard is used. Although the equipment is expensive, reagent costs are low and the use of radioisotopes is eliminated. Weaknesses of LC-MS/MS include the lack of a standardised method, the high cost of equipment, the possibility of interference with the C-3 epimer of 25OHD<sub>3</sub>, and the high level of expertise necessary to develop the method and run the equipment.

### The US experience

Dr Mary-Frances Picciano, National Institutes of Health, and Dr Rosemary L. Schleicher, Centers for Disease Control and Prevention (CDC), explained why the US national survey, NHANES, starting with the NHANES 2007–2008, will use an LC-MS/MS method instead of using the Diasorin Liaison immunoassay to measure 25OHD. Dr Karen Phinney, NIST, described the development of the standard reference material (SRM 972) for 25OHD.

CDC discontinued using the Diasorin RIA for NHANES at the end of the NHANES 2005–6, and dedicated several years to the development of an LC-MS/MS method. CDC cited internal and published data<sup>(9)</sup> demonstrating poor specificity of immunoassays compared with chromatographic methods. While HPLC-UV and LC-MS/MS showed outstanding comparability for 25OHD when directly compared<sup>(9)</sup>, isotope dilution tandem MS offers superior sensitivity, accuracy and precision. CDC had concluded that the Diasorin immunoassay lacked the precision necessary for NHANES, and that the changes made to the assay in the late 1990s had affected its performance. The reformulated Diasorin assay used during

NHANES 2000–6 produced 12% lower 25OHD results than the original assay used for NHANES 1988–94. Shifts in quality-control values, as a result of reagent lot variation, also complicated the interpretation of time trend data.

The CDC nutrition laboratory has now developed and validated an automated LC–MS/MS method for monitoring 25OHD<sub>2</sub> and 25OHD<sub>3</sub> in NHANES. A refinement that will allow the separation and measurement of the C-3 epimer of 25OHD<sub>3</sub> is under development. Reconciliation of the values obtained in NHANES in previous years using the Diasorin RIA method with those obtained using the new LC–MS/MS method will involve re-measuring approximately 100 samples from each of seven previous years with the new LC–MS/MS method and calculating regression equations for each year.

NIST has recently developed a SRM for vitamin D metabolites (SRM 972), which consists of four pools of fresh frozen serum<sup>(10)</sup> and these will be used to standardise the calibration of the LC–MS/MS method for NHANES. Each pool has a different concentration of 25OHD<sub>2</sub>, 25OHD<sub>3</sub> or both. One pool also contains a C-3 epimer (3-epi-25OHD<sub>3</sub>), a metabolite that may be present at significant concentrations in infant serum. As part of the value assignment of SRM 972, NIST developed an isotope-dilution LC–MS/MS methodology capable of resolving C-3 epimers. This method is being submitted for approval by the Joint Committee for Traceability in Laboratory Medicine as a reference measurement procedure (RMP).

As indicated earlier, measurement of the recovery of exogenously spiked samples appears not to be suitable for testing the ability of an immunoassay to measure endogenous 25OHD. The use of spiked SRM 972 pools should probably be restricted to assess the accuracy of HPLC–UV and LC–MS/MS methods only. Furthermore, the SRM 972 pool diluted with horse serum may not be suitable for the assessment of immunoassays due to matrix-related interferences in this type of assay. The endogenous human serum, however, does appear to provide comparable results in all assay procedures. A formal commutability study is needed to assess which levels of SRM 972 are appropriate to use with individual assays.

### Lessons from elsewhere

Professor Linda Thienpont, University of Ghent, described the determination of the ideal specifications for a reference measurement system for 25OHD, which includes both routine and reference measurements<sup>(11)</sup>. The specification goals for a RMP should be related to those for routine measurements; the imprecision (CV) of an RMP should be half of that of the goal for routine measurements, while the bias should be one-third of the bias goal for routine measurements. Achievable goals for routine measurements could be based on those determined by biological variation and reference intervals. Thus ideally, routine testing of 25OHD, such as that carried out in the NDNS, should have a precision (CV) ≤ 10% and bias ≤ 5%. These values should be incorporated into the final specification of the method chosen for the NDNS.

Professor Mike Wallace, Glasgow Royal Infirmary, discussed lessons which could be learnt from the measurement of other hard-to-measure hormones. Recent standardisation issues for the measurement of growth hormone have been resolved by the use of a common international standard,

achieved by detailed guidelines and recommendations formulated by the Society for Endocrinology<sup>(12)</sup>. Testosterone, like 25OHD, is a steroid that circulates in nmol/l concentrations bound to a high-affinity specific binding protein. In response to concerns about the validity of some testosterone assays, the Endocrine Society in the United States recently published a position statement, which concluded that direct immunoassays perform poorly at low testosterone concentrations and should be avoided<sup>(13)</sup>. They recommended that extraction and chromatography, followed by MS or immunoassay, were likely to furnish more reliable testosterone results. Similar recommendations for harmonising 25OHD measurements would be helpful.

The requirement for a RMP for 25OHD was highlighted by Graham Carter from DEQAS. There is an urgent need for a definitive method for 25OHD which is accepted to give the 'true' 25OHD concentration in serum samples<sup>(8)</sup>. Historically, GC–MS has been used as the RMP for steroid assays, but several LC–MS/MS methods have been recently promoted as candidate RMP for 25OHD. If an LC–MS/MS method is chosen as an RMP, its status would be greatly enhanced if results were confirmed by GC–MS. DEQAS has commissioned the development of a GC–MS method, which will be used to check the validity of the ALTM as a surrogate for the true value and/or for assigning target values to DEQAS samples. Early results are promising.

### The need for the separate measurement of 25-hydroxyvitamin D<sub>2</sub> and 25-hydroxyvitamin D<sub>3</sub>

Professor Christel Lamberg-Allardt, University of Helsinki, discussed whether the separate measurement of 25OHD<sub>2</sub> and 25OHD<sub>3</sub> was necessary for population monitoring of 25OHD status (C Lamberg-Allardt, unpublished results). There are differences in the biological activity and the toxicity of vitamins D<sub>2</sub> and D<sub>3</sub>. The intestinal absorption of vitamin D<sub>2</sub> is lower than that of vitamin D<sub>3</sub>. The clearance of 25OHD<sub>2</sub> from circulation is faster than that of 25OHD<sub>3</sub>, which could be due to the lower affinity of 25OHD<sub>2</sub> to vitamin D-binding protein. Furthermore, there are some differences in the metabolism between these two forms of vitamin D. Thus, the serum concentration of 25OHD<sub>2</sub> may not be equipotent to that of 25OHD<sub>3</sub>.

There are few sources of vitamin D<sub>2</sub> in the diet: it is mainly found in wild mushrooms. Some studies have shown that irradiation of cultivated mushrooms increases their vitamin D<sub>2</sub> content. Vitamin D<sub>3</sub> is the form mostly used in supplements and fortified foods. Hence, vitamin D<sub>3</sub> is the predominant form, arising from cutaneous synthesis, from food, and from fortified foods and supplements.

There are few population studies showing detectable circulating concentrations of 25OHD<sub>2</sub>. Unpublished results from the OPTIFORD study show that the proportion of the population with detectable levels of 25OHD<sub>2</sub> is low (about 5%). Nevertheless, at least in Finland, the vitamin supplements given as drops to infants contain vitamin D<sub>2</sub>, and their serum will mainly contain 25OHD<sub>2</sub>. In a recent report from West London in the UK, in which a high proportion of subjects were Asian, about 16% of the patients had detectable serum 25OHD<sub>2</sub> concentrations<sup>(14)</sup>. In conclusion, the need for measuring 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately depends on the future use of vitamin D<sub>2</sub> in food

fortification and as supplements, as the natural food sources are of marginal importance. The measurement of 25OHD<sub>2</sub> could be important at the individual level.

### Using the decision support matrix to reach consensus on methodology

#### Criteria

The purpose of the decision support matrix is to structure discussion and aid the evaluation and comparison of the

**Table 1.** The eleven criteria used to evaluate suitability for measuring 25-hydroxyvitamin D (25OHD) in the National Diet and Nutrition Survey (NDNS)

<i>Ability to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately:</i> the ability of the method to quantify 25OHD <sub>2</sub> and 25OHD <sub>3</sub> separately
<i>Ability to measure 25OHD<sub>2</sub> and 25OHD<sub>3</sub> equally well:</i> the ability of the method to be equally specific for 25OHD <sub>2</sub> and 25OHD <sub>3</sub> in volunteer pools and endogenous samples to give a total 25OHD value
<i>Ability to discriminate the C3-epi-isomer of 25 OHD3</i> within the assay*
<i>Low limit of quantitation:</i> the method can measure very small amounts of 25OHD with a reasonable degree of certainty (say 10 nmol/l ± 5%). This means a low functional sensitivity for immunoassays and a low limit of quantitation for HPLC-UV and LC-MS/MS†
<i>Trueness:</i> the closeness of agreement between values measured in an assay. In practice, this means a small percentage bias from either the ALTM or the MOMM. Close to the truth as we know it now‡§
<i>Ability of the method to be adjusted by the laboratory to a standard or reference method,</i> for example, the NIST standard and reference method
<i>Precision:</i> a small variation of the measurement within laboratories and between laboratories. Within-laboratory variation is measured with the intra-assay CV and the inter-assay (or total) CV. Between-laboratory variation is measured as the Annual CV for each method by DEQAS
<i>Reliability:</i> the technology works as expected, is not prone to breakdown and does not require adaptations to get it running reliably
<i>Good continuity over time:</i> results using this method can reliably be compared over time, and the method is not prone to changes in calibration, precision and sensitivity
<i>Ability to control standardisation:</i> the analyst has control over the assay calibration, but if a commercial supplier is involved, imminent changes in calibration are notified in advance
<i>Feasibility of setting up the method in the UK</i> with minimal adaptations to the standard method required¶

ALTM, all laboratory trimmed mean; MOMM, mean of method means; NIST, National Institute of Standards and Technology; DEQAS, vitamin D External Quality Assessment Scheme; VIM, *Vocabulaire International de Métrologie*.

\* The Expert Panel agreed that the ability to discriminate the C-3 epimer should be included as a criterion since the method would also be used in the forthcoming Diet and Nutrition Survey of Infants and Young Children.

† The Expert Panel agreed that a limit of quantitation in the region of <10 nmol/l (± 5%) was sufficient for the purposes of the NDNS.

‡ As there is no 'gold standard' for 25OHD methods, the Expert Panel agreed that accuracy or 'trueness' could not be assessed against one. It was agreed that assessment of accuracy against the NIST standard reference materials and the NIST candidate reference method procedure was the best currently available approach.

§ Subsequent to the workshop, it was suggested that it would have been better if an internationally accepted definition of trueness had been used. This is trueness: closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value (VIM). In practice, this means a small percentage bias from the 'truth' as we know it now<sup>(15,16)</sup>.

|| It was agreed that the ability of a method to be adjusted to the NIST standard (SRM 972 Level 1) should be a criterion.

¶ It was agreed that operator skills and laboratory facilities should be considered as part of the feasibility of setting up the method in the UK.

different 25OHD methods for the NDNS. The Expert Panel was asked to consider which criteria should be used and which methods should be evaluated. After a plenary discussion, the Expert Panel voted by a substantial majority to use eleven criteria for the evaluation procedure (see Table 1). They did not think it was necessary to weight the criteria because this would over-complicate the matrix. It was, however, made clear that this decision did not mean that all criteria had equal importance.

#### Methods for measuring 25-hydroxyvitamin D

The Expert Panel considered eight potential methods on which they felt there was adequate published information to evaluate (see Table 2). They decided to exclude three methods for the following reasons:

1. The Roche assay as it only measures 25OHD<sub>3</sub>.
2. The two RIA methods as the NDNS had already moved away from using an RIA method. Current UK Health and Safety policies restrict their use, and it is likely they will be phased out in future.

The Expert Panel agreed to evaluate five methods (Diasorin Liaison, IDS EIA automated, IDS EIA manual, HPLC-UV and LC-MS/MS) against the eleven agreed criteria.

### Evaluating the different methodologies according to the decision support matrix

The participants in the Expert Panel were divided into three breakout groups, chosen to ensure a range of expertise in each group. The breakout groups were asked to score each of the five methods from 1 to 4 against the eleven agreed criteria (1, not good; 2, acceptable; 3, good; 4, very good). Groups were encouraged to give a consensus score and to record their comments or any conditions for each score; the scores from each breakout group were then pooled. Breakout groups were given the option of changing their score if the plenary discussion revealed that they had interpreted the criteria in a slightly different way. The final criteria, and agreed definitions against which the different methods were evaluated, are those listed in Table 1.

The scores for the individual groups were very similar. All three groups scored LC-MS/MS and HPLC-UV similarly highly, while the scores given by all groups for the

**Table 2.** The eight potential methods considered and the five methods (in italics) chosen to evaluate in the decision support matrix

Manual immunoassay
<i>IDS manual EIA</i>
<i>DiaSorin RIA</i>
IDS RIA
Automated immunoassay
<i>DiaSorin Liaison Total</i>
<i>IDS EIA on automated platform</i>
Roche
Direct detection methods
<i>HPLC-UV</i>
<i>LC-MS/MS</i>

IDS, Immunodiagnostic Systems; EIA, enzyme-linked immunoassay.

**Table 3.** Pooled scores\* from all three breakout groups evaluating five methods against eleven criteria

	Immunoassay			HPLC-UV	LC-MS/MS
	Liaison Total	IDS EIA (automated platform)	IDS EIA (manual)		
Ability to measure 25OHD <sub>2</sub> and 25OHD <sub>3</sub> separately	1.0	1.0	1.0	4.0	4.0
Ability to measure 25OHD <sub>2</sub> and 25OHD <sub>3</sub> equally well	4.0	2.0	2.0	4.0	4.0
Ability to discriminate the C-3 epimer within the assay	1.0	1.0	1.0	4.0	4.0
Low limit of quantitation	2.7	3.0	3.3	4.0	4.0
Trueness	2.3	2.0	2.0	3.0	3.7
Adjustable (by laboratory) to reference method (NIST, etc.)	1.0	1.0	1.0	4.0	4.0
Precision	2.7	2.7	2.7	2.7	3.0
Reliability	2.3	2.3	2.7	2.7	2.3
Good continuity over time	2.0	2.0	2.0	3.0	3.0
Ability to control standardisation	1.3	1.3	1.3	3.7	3.7
Feasibility of setting up the method in the UK	4.0	4.0	3.7	3.3	3.3
<b>Total score</b>	<b>24 (24, 23, 26)</b>	<b>22 (23, 23, 21)</b>	<b>23 (22, 24, 22)</b>	<b>38 (37, 39, 39)</b>	<b>39 (38, 40, 39)</b>

IDS, Immunodiagnostic Systems; EIA, enzyme-linked immunoassay; 25OHD, 25-hydroxyvitamin D; NIST, National Institute of Standards and Technology.

\* Scoring: 1, not good; 2, acceptable; 3, good; 4, very good.

immunoassays were lower. The total 'plenary' score for each method (as listed in Table 3) therefore reflected these differences. It was agreed that the scores for each method reflected the discussions about the strengths and weaknesses of the different methods, in relation to the particular requirements of the NDNS. The majority view was that only two methods (HPLC-UV and LC-MS/MS) should go on for further discussion. There was little to discriminate between the scores of the three immunoassays, and of the three methods, the Diasorin Liaison Total method scored the highest. Discussion thereafter focussed on LC-MS/MS and HPLC-UV. Table 4 summarises their strengths and weaknesses.

### Recommendations for 25-hydroxyvitamin D methodology for use in National Diet and Nutrition Survey

Both HPLC-UV and LC-MS/MS scored similarly highly, and there was little to discriminate between them according to the criteria against which they had been evaluated. It was agreed, however, that LC-MS/MS should be recommended as the preferred method for the NDNS as this would facilitate future comparisons on 25OHD status with NHANES. The fact that NHANES has recently made the change to LC-MS/MS from immunoassay and has already put reconciliation measures in place strongly influenced the Expert Panel to recommend this methodology to FSA for use in NDNS. If FSA accepts this recommendation, the Expert Panel suggested the following considerations:

1. The particular characteristics the specific method needs to achieve in terms of sensitivity, precision and accuracy should be specified at the outset.
2. The specification should be drawn up in close liaison with US CDC to ensure that the method used for NDNS is similar to that used for NHANES. A visit to the CDC Micronutrient laboratory in Atlanta by the successful contractor is recommended.

3. There is a potential problem with the blood collection tubes used in the NDNS (Sarstedt Monovette™ tubes) as they could introduce ion suppression effects in LC-MS/MS. The method can be adapted to prevent the problem from occurring.
4. The use of the NIST standards (SRM 972) or secondary standards calibrated against NIST standards is recommended.
5. A reconciliation procedure is needed to be able to compare the results measured previously using Diasorin Liaison Total method with those obtained using the

**Table 4.** Strengths and weaknesses of HPLC-UV and LC-MS/MS

Both methods scored highly for their ability to measure 25OHD <sub>2</sub> and 25OHD <sub>3</sub> separately and equally well
Both methods would be able, with suitable adjustment, to discriminate the C-3 epimer. However, there is currently no equivalent internal standard, and CDC and NIST use deuterated 3-β-OH-25OHD <sub>3</sub> as the internal standard for the epimer because the deuterated epimer is unavailable presently
Both methods also scored highly for a low limit of quantitation and ability to adjust to the NIST standard and reference method
Both methods scored relatively poorly for reliability, reflecting the complexity of the instrumentation needed to run the assays and the skill needed for their maintenance. All groups, however, agreed it must be assumed that the measurement would be carried out in an established laboratory, using skilled personnel. It was also suggested that it would be best to batch samples and measure on an annual basis, although the need to report values back to participants in a reasonable time might mean that this is not feasible
HPLC-UV scored less well on precision, based on the between-laboratory precision data provided by DEQAS, and on trueness compared with LC-MS/MS. The group noted that individual laboratories, however, could have good intra- and inter-assay precision data

25OHD, 25-hydroxyvitamin D; CDC, Centers for Disease Control and Prevention; NIST, National Institute of Standards and Technology; DEQAS, vitamin D External Quality Assessment Scheme.

suggested LC–MS/MS method. It will be necessary to run the two methods together for a while to perform a cross-over study. This will enable a correction factor or derived equation (or factors for low/medium/high concentrations) to be determined for values previously measured with the Diasorin Liaison method.

6. Ideally, there should be a limited range of methods being used by FSA contractors measuring 25OHD. The results obtained by other FSA-funded projects should be comparable with the values obtained from the NDNS. As each research contract is put out to tender, the FSA should reserve the option to decide to specify a single laboratory for the measurement of 25OHD for all its research projects.
7. While LC–MS/MS is capable of quantitating 25OHD<sub>2</sub> and 25OHD<sub>3</sub> separately, the FSA should seek advice on whether it needs to report both separately, as will be done in NHANES. The Expert Panel advised that, when reporting values back to participants, only total 25OHD (25OHD<sub>3</sub> + 25OHD<sub>2</sub>) should be reported, in keeping with usual UK clinical practice.
8. The LC–MS/MS method chosen will need to be adjusted to ensure that it is able to discriminate the C-3 epimer of 25OHD<sub>3</sub>, especially in the Diet and Nutrition Survey of Infants and Young Children. The Expert Panel thought that until it was known whether the C-3 epimer was functional, there was probably no need to quantify it.

### Conclusions and recommendations

The Expert Panel recommended that an LC–MS/MS method should be the preferred method for the NDNS. A detailed specification for the method will be required to ensure comparability between NDNS and NHANES in the US facilitating future comparisons. The Expert Panel also recommended that the method should be carried out in a laboratory with appropriate expertise, competency and history of records of good performance. The method should be standardised against NIST SRM 972. If the recommended LC–MS/MS is adopted, the Expert Panel indicated that the method should be able to discriminate the C-3 epimer of 25OHD<sub>3</sub>, especially if used to measure 25OHD in young infants in the forthcoming Diet and Nutrition Survey of Infants and Young Children, who are known to have high circulating concentrations of the C-3 epimer.

### List of participants (Food Standards Agency Workshop NO8029)

Expert Panel: Dr Jackie Berry, Manchester Royal Infirmary, Manchester; Graham Carter, DEQAS, London; Dr Christine Clewes (was at), MRC Human Nutrition Research, Cambridge; Professor David Fraser, University of Sydney, Australia; Dr Dominic Harrington, St Thomas' Hospital, London; Dr Tom Hill, University College, Cork, Ireland; Professor Bruce Hollis, Medical University of South Carolina, USA; Julia Jones, DEQAS, London; Dr Susan Knox, Glasgow Royal Infirmary, Glasgow; Dr Susan Lanham-New, University of Surrey, Guildford; Dr Helen Macdonald, University of Aberdeen, Aberdeen; Professor Hugh Makin, St Bartholomew's

and the Royal London School of Medicine and Dentistry, London; Dr Karen Phinney, NIST, USA; Dr Mary-Frances Picciano, Office National Institutes of Health, USA; Dr Sandra Rainbow, Northwick Park Hospital, Harrow; Dr Rosemary L. Schleicher, CDC, USA; Dr Inez Schoenmakers, MRC Human Nutrition Research, Cambridge; Dr Martin Shearer, Guy's and St Thomas' NHS Foundation Trust, London; Dr Dietmar Stockl, Ghent University, Belgium; Professor Linda Thienpont, Ghent University, Belgium; FSA: Dr Fiona Comrie (FSA Scotland); Dr Alison Tedstone; Heiko Stolte; Dr Elaine Stone; Gillian Swan; Ashwell Associates consortium: Dr Margaret Ashwell; A. d. I. H.; S. G.; Professor Christel Lamberg-Allardt and H. V., University of Helsinki, Finland; Professor Mike Wallace, Glasgow Royal Infirmary, Glasgow.

### Acknowledgements

No author has any conflicts of interest. A. d. I. H., M. A., S. G., A. M. W., C. L.-A. and H. V. were contracted by the FSA to organise and be rapporteurs of the Workshop. A. d. I. H. produced the first and subsequent versions of the manuscript, to which all other authors contributed substantially, in terms of drafting, critical reviewing and editing.

### References

1. Binkley N, Krueger D, Cowgill CS, *et al.* (2004) Assay variation confounds the diagnosis of hypovitaminosis D: a call for standardization. *J Clin Endocrinol Metab* **89**, 3152–3157.
2. Carter GD, Carter R, Jones J, *et al.* (2004) How accurate are assays for 25-hydroxyvitamin D? Data from the international vitamin D external quality assessment scheme. *Clin Chem* **50**, 2195–2197.
3. Hollis BW (2004) Editorial: the determination of circulating 25-hydroxyvitamin D: no easy task. *J Clin Endocrinol Metab* **89**, 3149–3151.
4. Scientific Advisory Committee on Nutrition (2007) *Update on Vitamin D*. London: TSO.
5. Wallace AM, Gibson S, de la Hunty A, *et al.* (2010) Measurement of 25-hydroxyvitamin D in the clinical laboratory; current procedures, performance characteristics and limitations. *Steroids* (epublication ahead of print version 24 February 2010).
6. Carter GD, Jones JC & Berry JL (2007) The anomalous behaviour of exogenous 25-hydroxyvitamin D in competitive binding assays. *J Steroid Biochem Mol Biol* **103**, 480–482.
7. Beastall G & Rainbow S (2008) Vitamin D reinvented: implications for clinical chemistry. *Clin Chem* **54**, 630–632.
8. Carter GD & Jones JC (2009) Use of a common standard improves the performance of liquid chromatography–tandem mass spectrometry methods for serum 25-hydroxyvitamin-D. *Ann Clin Biochem* **46**, 79–81.
9. Roth HJ, Schmidt G, Heinrich, Weber H, *et al.* (2008) Accuracy and clinical implications of seven 25-hydroxyvitamin D methods compared with liquid chromatography–tandem mass spectrometry as a reference. *Ann Clin Biochem* **45**, 153–159.
10. Phinney KW (2008) Development of a standard reference material for vitamin D in serum. *Am J Clin Nutr* **88**, 511S–512S.
11. Stöckl D, Sluss P & Thienpont L (2009) Specifications for trueness and precision of a reference measurement system for serum/plasma 25-hydroxyvitamin D analysis. *Clin Chim Acta* **408**, 8–13.
12. Wieringa G & Trainer P (2007) Commentary: harmonizing growth hormone measurements: learning lessons for the future. *J Clin Endocrinol Metab* **92**, 2874–2875.

13. Rosner W, Auchus R, Azziz R, *et al.* (2007) Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab* **92**, 405–413.
14. Costelloe SJ, Woolman E, Rainbow S, *et al.* (2009) Is high-throughput measurement of 25-hydroxyvitamin D<sub>3</sub> without 25-hydroxyvitamin D<sub>2</sub> appropriate for routine clinical use? *Ann Clin Biochem* **46**, 86–87.
15. JCGM (2008) *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*. Paris: International Bureau of Weights and Measures (BIPM).
16. ISO/IEC (2007) *Guide 99: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*. (VIM=Vocabulaire International de Métrologie). Geneva: International Organization for Standardization.