NHANES Monitoring of Serum 25-Hydroxyvitamin D: A Roundtable Summary1–3

Elizabeth A. Yetley,4∗ Christine M. Pfeiffer,5 Rosemary L. Schleicher,5 Karen W. Whinney,6 David A. Lacher,7 Sylvia Christakos,8 John H. Eckfeldt,9 James C. Fleet,10 George Howard,11 Andrew N. Hoofnagle,12 Siu L. Hui,13 Gary L. Lensmeyer,14 Joseph Massaro,15 Munro Peacock,13 Bernard Rosner,16 Donald Wiebe,17 Regan L. Bailey,8 Paul M. Coates,4 Anne C. Looker,7 Christopher Sempos,4 Clifford L. Johnson,7 and Mary Frances Picciano4 for the Vitamin D Roundtable on the NHANES Monitoring of Serum 25(OH)D: Assay Challenges and Options for Resolving Them

Abstract
A roundtable to discuss monitoring of serum 25-hydroxyvitamin D [25(OH)D] in the NHANES was held in late July 2009. Topics included the following: 1) options for dealing with assay fluctuations in serum 25(OH)D in the NHANES conducted between 1988 and 2006; 2) approaches for transitioning between the RIA used in the NHANES between 1988 and 2006 to the liquid chromatography tandem MS (LC-MS/MS) measurement procedure to be used in NHANES 2007 and later; 3) approaches for integrating the recently available standard reference material for vitamin D in human serum (SRM 972) from the National Institute of Standards and Technology (NIST) into the NHANES; 4) questions regarding whether the C-3 epimer of 25-hydroxyvitamin D3 [3-epi-25(OH)D3] should be measured in NHANES 2007 and later; and 5) identification of research and educational needs. The roundtable experts agreed that the NHANES data needed to be adjusted to control for assay fluctuations and offered several options for addressing this issue. The experts suggested that the LC-MS/MS measurement procedure developed by NIST could serve as a higher order reference measurement procedure. They noted the need for a commutability study for the recently released NIST SRM 972 across a range of measurement procedures. They suggested that federal agencies and professional organizations work with manufacturers to improve the quality and comparability of measurement procedures across all laboratories. The experts noted the preliminary nature of the evidence of the 3-epi-25(OH)D3 but felt that it should be measured in 2007 NHANES and later.

Executive Summary
Starting in 1988–1994 and continuing through the present time, the NHANES has generated nationally representative population data on serum 25-hydroxyvitamin D [25(OH)D] concentrations. Recently, the National Center for Health Statistics (NCHS) of the CDC posted an analytical note on its Web page informing users that NHANES 1988–1994 and 2000–2006

1 Published in a supplement to The Journal of Nutrition. Presentations and discussions at the Roundtable “NHANES Monitoring of Serum 25(OH)D: Assay Challenges and Options for Resolving Them” were held in Rockville, MD, July 27–28, 2009. The roundtable was organized and cosponsored by the Division of Health and Nutrition Examination Surveys (National Center for Health Statistics, CDC) and the Office of Dietary Supplements (NIH). Funding was received from these 2 agencies and the Office of Research on Women’s Health of the NIH. The supplement coordinator is Elizabeth A. Yetley, Office of Dietary Supplements, the NIH. Supplement Coordinator disclosure: Elizabeth A. Yetley has been hired as a scientific consultant to the Office of Dietary Supplements/NIH. Her activities relate to the Office’s Vitamin D and NHANES activities and convening and chairing the roundtable meeting. She is also responsible for drafting the roundtable final report and obtaining necessary reviews and clearances from roundtable experts and government participants involved in the roundtable discussions. Guest Editor for this Supplement was Stephanie A. Atkinson. Guest Editor disclosure: no conflicts to disclose. Publication costs for this supplement were defrayed in part by the payment of page charges. This publication must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact. The opinions expressed in this publication are those of the authors and are not attributable to the sponsors or the publisher, Editor, or Editorial Board of The Journal of Nutrition. The Roundtable is dedicated to Dr. Mary Frances Picciano who was the driving force behind this effort prior to her death in August 2010.
∗To whom correspondence should be addressed. E-mail: beth@yetley.com.
serum 25(OH)D measures are likely affected by fluctuations in the assay performance (measurement procedure bias and imprecision) over time. A key question is whether differences in serum 25(OH)D concentrations across surveys are related to true population changes in vitamin D status or are a result of measurement procedure fluctuations unrelated to changes in vitamin D status.

A roundtable was convened in July, 2009, to discuss the scientific challenges involved in the measurement and evaluation of serum 25(OH)D across several decades of NHANES data and to discuss future challenges of switching from the DiaSorin RIA to a liquid chromatography tandem MS (LC-MS/MS) measurement procedure. Also discussed were issues of integrating the recently released standard reference materials (SRM) for serum 25(OH)D from the National Institute of Standards and Technology (NIST) and questions about the quantitative reporting of the C-3 epimer of 25-hydroxy vitamin D3 [3-epi-25(OH)D3] in 2007 and beyond.

**NHANES 2000–2006**

A DiaSorin RIA kit was used for measuring serum 25(OH)D concentrations during NHANES 2000–2006. The results of quality control (QC) pools showed problems of assay fluctuations in 2004 and 2006. During 2004, the serum 25(OH)D concentrations for the QC pools were increased compared with values in 2003 and 2005. The start of the upward drift coincided with a new lot of standards and tracer; it ended with a new lot of tracer and antisemur. The percent increase in the pools ranged from 4.1% for a low concentration pool to 10.2% for the medium and high concentration pools.

During 2006, the 25(OH)D concentrations for the high QC pool were lower compared with values in 2005. This downward drift coincided with a new lot of tracer and antisemur. The decrease in mean pool values before and during the drift ranged from 5.3 to 17.1% for the various QC pools.

The roundtable experts agreed that adjustments for the assay fluctuations are needed to avoid misuse of the serum 25(OH)D results. They identified several options for addressing the assay fluctuations. The experts agreed that the preferred long-term approach is to reanalyze a subset of serum samples across all the DiaSorin-based surveys (1988–1994 and 2000–2006) by using a LC-MS/MS measurement procedure that will be used for NHANES 2007 and beyond in conjunction with the use of the NIST SRM for serum 25(OH)D to facilitate traceability to a higher order measurement. However, this approach would take considerable time to complete, making it impractical for meeting current needs for estimates of vitamin D status of U.S. population groups.

The roundtable experts identified 2 short-term, preliminary options that could serve as surrogates for retesting the NHANES 2000–2006 blood samples by LC-MS/MS: 1) a statistical adjustment model that is based on the QC data but that better defines the boundaries for the periods of drift; and 2) a statistical approach that uses participant serum 25(OH)D and an ANCOVA model to control for short-time effects of assay fluctuations. The roundtable experts did not support an option of using pooled data from multiple years of NHANES 2000–2006 to average out the high values from 2004 and the low values from 2006.


After the release of the NHANES 2001–2002 serum 25(OH)D data, researchers noted that the serum 25(OH)D concentrations for NHANES 2001–2002 were lower than NHANES III (1988–1994). DiaSorin RIA kits had been used to measure serum 25(OH)D during both surveys. Subsequently, information from the manufacturer of the DiaSorin RIA kits indicated that between these 2 surveys, changes had been made: a new antibody with improved binding had been introduced and detergents were added to a buffer to reduce nonspecific binding.

To address whether these kit changes could be responsible for some or all of the differences between the NHANES 1988–1994 and the NHANES 2000 and later, the analytical laboratory in the National Center for Environmental Health (NCEH) at the CDC that was measuring serum 25(OH)D designed and conducted a crossover study. It reanalyzed a subset of 150 banked serum samples from the 1988–1994 NHANES using the 1998 reformulated DiaSorin RIA kit. Taking into account the results of the crossover study, as well as the drift in the 2004 QC data that occurred during the conduct of the crossover study, an adjustment equation using a Deming regression (which allows for error in both the dependent and independent variables) was generated: 1988–1994 NHANES 25(OH)Dcorrected 2004 RIA = (0.8429 X 1998–1994 NHANES 25(OH)D1988–1994 RIA) + 2.5762 nmol/L. The mean difference between the reformulated RIA assay used in NHANES 2000–2004 and the original RIA used in 1988–1994 before these values were adjusted to be similar to the reformulated RIA was ~12%. This equation is currently identified by NCHS as an interim adjustment pending a more thorough evaluation of the assay-related changes affecting comparisons of serum 25(OH)D data from the 1988–1994 NHANES with data from the NHANES 2000–2004. The assay-related differences between the DiaSorin RIA used in the 1988–1994 NHANES and the reformulated DiaSorin RIA used in the NHANES 2000 and later, which are unrelated to changes in vitamin D status, accounted for much of the difference in the serum 25(OH)D concentrations between the 1988–1994 NHANES and the 2000–2004 NHANES.

The roundtable experts discussed approaches for making NHANES 1988–1994 serum 25(OH)D data comparable with the NHANES 2000–2006. In addition to the preferred long-term approach described above in which a subset of the DiaSorin RIA analyses would be reanalyzed with the LC-MS/MS measurement procedure and use of the NIST SRM, a short-term option was suggested: use the regression equation from the 150-sample
crossover study after verifying whether a log or other transformation would improve the statistical characteristics of the Deming regression equation.

**Future use of the LC-MS/MS measurement procedure and the NIST SRM 972**

Given the difficulties with the DiaSorin RIA fluctuations and imprecision during NHANES III (1988–1994) and NHANES 2000–2006, the decision was made to replace the RIA with the more accurate, precise, and specific LC-MS/MS measurement procedure. Preliminary studies show that key performance characteristics (e.g. accuracy, specificity, and precision) are met by the use of the LC-MS/MS measurement procedure. In addition, the LC-MS/MS allows for separate quantification of serum 25-hydroxyvitamin D2 [25(OH)D2] and 25(OH)D3.

A protocol for a crossover study comparing results from the new and old measurement procedures was discussed. The roundtable experts agreed with the need for the crossover study and statistical adjustments to compare trends over time. They discussed the need for crossover study designs that take into account user data needs for accurate prevalence estimates at the tails of the 25(OH)D distribution curves.

**SRM for serum 25(OH)D.** The NIST released SRM 972 Vitamin D in Human Serum in June 2009. The SRM consists of frozen serum with 4 designated concentrations plus or minus the expanded uncertainty of vitamin D metabolites: 1) a normal human serum pool with a certified 25(OH)D3 concentration of 59.6 ± 2.1 nmol/L; 2) a blend of normal human serum and horse serum to obtain approximately one-half the level of 25(OH)D3 (30.8 ± 1.5 nmol/L); 3) a normal human serum spiked with an amount of 25(OH)D2 (64.1 ± 4.8 nmol/L) approximately equivalent to the concentrations of 25(OH)D3; and 4) a normal human serum spiked with 3-epi-25(OH)D3 (94.1 ± 2.9 nmol/L). The roundtable experts agreed that future releases of these reference materials should obtain serum pools from individuals with naturally high and low concentrations of serum 25(OH)D2 and 25(OH)D3 rather than using in vitro supplementation or dilution with nonhuman serum. The roundtable experts also agreed that the scientific rigor used to develop the NIST LC-MS/MS methodology for the SRM 972 project was sufficient to consider it a higher order reference measurement procedure.

**Commatability of the SRM 972 materials was not assessed prior to the release of the SRM. Therefore, NIST included a cautionary statement in the Certificate of Analysis indicating that the user should decide which levels will be useful for a particular measurement procedure. (Commatability means that the reference material behaves the same as a patient sample).** The roundtable experts supported NIST plans to conduct a commutability study in the near future and provided guidance on design issues.

NIST reported on the development of ethanol-based calibration solutions for 25(OH)D2 and 25(OH)D3 (SRM 2972) to help meet user needs for reliable reference materials. NIST also reported on its quality assurance program for measuring serum 25(OH)D.

**Integrating NIST reference materials into the NHANES analyses.** NCEH proposes to incorporate the 2 types of NIST reference materials into future NHANES analyses: SRM 972 with its 4 levels of serum-based reference values and the SRM 2972 solvent-based calibrator solutions for 25(OH)D2 and 25(OH)D3. The roundtable experts felt that including these reference materials should enhance the accuracy of comparisons across the serum 25(OH)D results both for future NHANES analyses and reanalyzing serum 25(OH)D samples from past NHANES. However, the experts expressed one note of caution: the traceability to high-level reference measurement procedures does not fully solve the issue of analytical specificity of the competitive protein binding assays.

**What about the serum 3-epi-25(OH)D3 epimer?** The planned conversion from the DiaSorin RIA to a LC-MS/MS methodology beginning with the NHANES blood samples collected in 2007 raised the issue of how to address the measurement of the 3-epi-25(OH)D3. Although the 3-epi-25(OH)D3 is not measured by using some immunoassays, it is usually measured along with 25(OH)D3 in LC-MS/MS measurement procedures. Failure to separate the epimer of 25(OH)D3 results in overestimating the concentrations of serum 25(OH)D3. Moreover, because the epimer is converted to a form of 1,25-dihydroxvitamin D that differs in biological activity from the hormone produced from 25(OH)D3, distinguishing between the amounts of 25(OH)D3 and 3-epi-25(OH)D3 is likely to have biological relevance.

The roundtable experts agreed that starting with the 2007 NHANES, the laboratory should separate and quantify the serum 3-epi-25(OH)D3 concentrations. This process would result in a more accurate measurement of the serum 25(OH)D3 concentrations. Moreover, the NHANES dietary and biomarker data could be used to examine whether concentrations of the serum epimer were associated with intakes of fortified foods and supplements.

**Research and educational needs**

The roundtable experts identified 3 areas requiring further follow-up from a research or educational perspective. The first is improving the standardization, harmonization, and accuracy of serum 25(OH)D results across laboratories. The roundtable experts felt that all laboratories—clinical, research, and public health—should produce serum 25(OH)D results that are accurate and traceable back to a high-level reference measurement procedure. The experts concluded that the primary responsibility for improvements in serum 25(OH)D measurement procedures should come from in vitro diagnostic device manufacturers. Federal agencies and expert panels should identify the characteristics of measurement procedures that laboratories need to control. These groups should inform laboratories and manufacturers of these measurement characteristics and also provide information on available guidelines on transferrability and traceability.

The second research need is to better understand the source(s) (e.g. in vivo or in vitro) and biological activity of the 3-epi-25(OH)D3. The 3rd research need is to better understand both the genetic and environmental factors that affect serum 25(OH)D3 concentrations. Also needed is a better understanding of quantitative dose-response relationships among intakes, serum 25(OH)D concentrations, and clinical outcomes.

**Background**

The NHANES started collecting data on serum 25(OH)D values in the 1988–1994 NHANES III and continued during the NHANES 2000–2006 (1–3). These nationally representative surveys provide summary statistics and prevalence estimates for assessing vitamin D status across a range of age, gender, and racial and ethnic groups.

Serum 25(OH)D is considered the best available biomarker of exposure to dietary and endogenous sources of vitamin D (4). Interpretation of serum 25(OH)D results from NHANES requires the ability to relate prevalence estimates to externally...
validated and generally accepted status cutpoints of inadequacy and safety as well as to determine whether time trends reflect changes in status over time. It is, therefore, essential that the NHANES values are accurate and their comparability with published data can be assessed. For nutritional biomarkers, this accuracy and comparability ideally derives from the ability to trace measurements to a higher order reference material, such as that provided by SRM and certified reference materials, and to higher order reference measurement procedures (5–7). This ability is particularly relevant for measures of serum 25(OH)D, because the reliability and consistency of different serum 25(OH)D analytical methodologies are highly variable (8–12) even within a laboratory over time (1,2,13).

Recently, the NCHS of the CDC posted an analytical note on their NHANES Web page informing users about 2 issues that should be addressed when analyzing and using serum 25(OH)D data from NHANES (2,14). The first issue was a caution regarding comparing serum 25(OH)D from the NHANES III (1988–1994) with the NHANES 2000–2006 because of a reformulation of the DiaSorin RIA kit that resulted in shifts in serum 25(OH)D assay results between the 2 time periods. The second was a caution that the data from the 2000–2006 NHANES are likely affected by fluctuations in the assay performance (measurement procedure bias and imprecision) over time.

The analytical note (2) also indicated the availability of a SRM for serum vitamin D (SRM 972) with certified values assigned by using LC/MS/MS candidate reference measurement procedures from the NIST (15). CDC stated its plans to incorporate the regular use of the SRM 972 into future measures of 25(OH)D in NHANES 2007 and beyond. CDC also plans to generate regression equations that will permit the adjustment of assay results across survey periods that used the DiaSorin RIA; 3) the challenges involved in switching to a LC-MS/MS assay in the surveys conducted during and after 2007; 4) issues related to integrating the recently released SRM 972 (15) into future NHANES analyses; and 5) questions of if and how to deal with the serum 3-epi-25(OH)D3. The 3-epi-25(OH)D3 differs from 25(OH)D3 in the configuration of the C3-hydroxy group [3α (epi) vs. 3β]. The roundtable also identified research and educational needs to support using currently available serum 25(OH)D data and to enhance the measurement of this biomarker in future surveys.

### Measurement of Serum 25(OH)D in the NHANES Surveys

The procedures used for measuring serum 25(OH)D concentrations in the NHANES and the covered age groups have evolved since the 1988–1994 NHANES (Table 1). The original DiaSorin RIA was used for the NHANES III (1988–1994) (1,2). The reformulated DiaSorin RIA was used for the NHANES from 2000 to 2006. Starting with the 2007 survey year, the LC-MS/MS measurement procedure will be used. The age of coverage in the 1988–1994 NHANES was ≥12 y and in the 2000–2002 NHANES was ≥6 y. Subsequently, the age coverage was expanded to all persons ≥1 y.

The analysis of serum 25(OH)D for all of the NHANES surveys was conducted in Atlanta, GA, at the NCEH at CDC. Detailed descriptions of the measurement procedures used by NCEH for assaying serum 25(OH)D in the NHANES are provided on the NHANES Web site (18–21). The NCEH quality assurance system has a number of components (Table 2). The NCEH laboratory follows the requirements for quality laboratory testing and is certified under the Clinical Laboratory Improvement Amendments program sponsored by the Centers for Medicare and Medicaid Services (22). The NCEH laboratory also participates in the proficiency testing programs of the Vitamin D External Quality Assessment Scheme (DEQAS) (23) and, since 2007, the College of American Pathologists (24). The laboratory is enrolled in the NIST Quality Assurance program for vitamin D in human serum (25). The roundtable experts noted that participation in these quality assurance programs is important. However, they also noted that these programs do not

### The Roundtable

The roundtable, NHANES Monitoring of Serum 25(OH)D: Assay Challenges and Options for Resolving Them, was held in Rockville, MD, on July 27–28, 2009. The Office of Dietary Supplements (ODS) at NIH and the Division of Health and Nutrition Examination Statistics at NCHS/CDC cosponsored the roundtable discussions. Additional funding was obtained from NIH’s Office of Research on Women’s Health. Eleven external scientists with expertise in laboratory sciences, biostatistics, or vitamin D metabolic and clinical issues joined government scientists with expertise in analyzing and using NHANES’ nutritional biomarkers. The roundtable experts and participating government scientists served as coauthors of this summary report of the roundtable discussions. This meeting was intended to serve as a dialogue between external experts and government scientists. Members of the roundtable were neither asked nor expected to make recommendations. The focus of the roundtable discussions was on identifying the key science-based issues affecting the accuracy of serum 25(OH)D measurements in NHANES and options for effectively addressing these issues within the NHANES surveys. The roundtable focused specifically on the measurement of serum 25(OH)D in NHANES. The experts were not asked to address the broader issues of measurement procedure differences among a range of methodologies currently in use in clinical and research laboratories or the biological basis for commonly used cutpoints of vitamin D status.

The roundtable focused on topics of particular relevance to the challenges related to the accuracy of serum 25(OH)D results across previous NHANES (1988–1994 and 2000–2006) and to the anticipated challenges in future NHANES (2007 and beyond). This focus involved: 1) a review of QC issues related to measuring serum 25(OH)D; 2) a review of approaches for harmonizing past assay results across survey periods that used the DiaSorin RIA; 3) the challenges involved in switching to a LC-MS/MS assay in the surveys conducted during and after 2007; 4) issues related to integrating the recently released SRM 972 (15) into future NHANES analyses; and 5) questions of if and how to deal with the serum 3-epi-25(OH)D3. The 3-epi-25(OH)D3 differs from 25(OH)D3 in the configuration of the C3-hydroxy group [3α (epi) vs. 3β]. The roundtable also identified research and educational needs to support using currently available serum 25(OH)D data and to enhance the measurement of this biomarker in future surveys.
replace the need for traceability of assay results to the high-level reference measurement procedure and reference materials that have only recently become available from NIST.

At any given time, the NCEH laboratory uses 5 QC materials to monitor consistency in measuring different concentrations of serum 25(OH)D. Starting in NHANES III (1988–1994) and continuing through NHANES 2000–2006, these QC materials represented “total” serum 25(OH)D [i.e. the sum of 25(OH)D2 and 25(OH)D3]. Open-label pools representing high, medium, and low concentrations of serum 25(OH)D are used for bench QC, and a high and low pool are used as blind samples; the latter are labeled the same as unknown participant samples and, therefore, remain unrecognized by the analyst. The manufacturer of the DiaSorin RIA kits allows for QC pool limits that are wider [i.e. corresponding to analytical CV (CVA) of ~20%] than generally used for NHANES-related assays. Moreover, this manufacturer frequently changes QC pool lots. NCEH, therefore, generated, characterized, and used in-house QC pools to allow judgment of runs based on stricter limits and for longer intervals throughout the NHANES time periods. The typical CVA for the QC pools used by NCEH for the NHANES 2000–2006 was 10–13% (25–112.5 nmol/L concentration range). A Westgard-like multi-rule QC system that includes preset rules for determining when an individual measure is out of control (26) was used to monitor that measurement systems were performing adequately across a range of concentrations. These rules identified the systematic shifts in QC data described later in this publication. Twice during NHANES 2000–2006, the NCEH updated limits on its in-house QC pools to accommodate mean shifts when troubleshooting efforts were exhausted and no apparent reasons for the pool shifts were uncovered.

External review of the laboratory procedures and results are independently conducted by NCHS staff and contractors who are involved in the administration of NHANES. The participant data and the associated QC are matched. Administrative checks are performed including monitoring for correct data format, extreme values, correct demographics (age and gender), and evaluation of limits of detection. The QC is screened by using a

---

**TABLE 1** Measurement procedures used for serum 25(OH)D measurements in NHANES

<table>
<thead>
<tr>
<th>Survey</th>
<th>Age</th>
<th>Measurement procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988–1994</td>
<td>≥12 y</td>
<td>Original DiaSorin RIA</td>
</tr>
<tr>
<td>2000</td>
<td>≥6 y</td>
<td>Reformulated DiaSorin RIA</td>
</tr>
<tr>
<td>2001–2002</td>
<td>≥6 y</td>
<td>Reformulated DiaSorin RIA</td>
</tr>
<tr>
<td>2003–2004</td>
<td>≥1 y</td>
<td>Reformulated DiaSorin RIA</td>
</tr>
<tr>
<td>2005–2006</td>
<td>≥1 y</td>
<td>Reformulated DiaSorin RIA</td>
</tr>
<tr>
<td>2007–20081</td>
<td>≥1 y</td>
<td>LC-MS/MS</td>
</tr>
</tbody>
</table>

1 Table information from references 1–3.  2 Samples have not yet been analyzed.

---

**TABLE 2** Basic components of the NCEH Division of Laboratory Sciences quality assessment system for the serum 25(OH)D measurement procedure

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Written protocols</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Trained staff</td>
<td>MT (ASCP)</td>
<td>MT (ASCP)</td>
</tr>
<tr>
<td>Validated measurement procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recovery</td>
<td>Instruction manual2</td>
<td>Instruction manual3; DEQAS4</td>
</tr>
<tr>
<td>Precision</td>
<td>Instruction manual2</td>
<td>Instruction manual3; CDC5</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Instruction manual2</td>
<td>Instruction manual3; CDC6</td>
</tr>
<tr>
<td>Specificity</td>
<td>Instruction manual2</td>
<td>Instruction manual3; CDC7</td>
</tr>
<tr>
<td>Linearity</td>
<td>Instruction manual2</td>
<td>Instruction manual3; DEQAS8</td>
</tr>
<tr>
<td>QC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bench QC pools in every assay</td>
<td>3 Levels</td>
<td>3 Levels</td>
</tr>
<tr>
<td>Blind QC (1 in 20 samples)</td>
<td>2 Levels</td>
<td>2 Levels</td>
</tr>
<tr>
<td>QC rules</td>
<td>Westgard-type10</td>
<td>Westgard-type11</td>
</tr>
<tr>
<td>Proficiency testing</td>
<td>DEQAS</td>
<td>DEQAS</td>
</tr>
<tr>
<td>Calibration verification</td>
<td>Instruction manual2</td>
<td>Instruction manual3; CDC12</td>
</tr>
<tr>
<td>Equipment function checks</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

1 Same medical technologist (MT) certified by the American Society for Clinical Pathology (ASCP) was the DiaSorin RIA analyst for both periods.  2 Instruction manual from the INCSTAR (now DiaSorin RIA) kit manufacturer for the original assay (catalogue no. 68100).  3 Instruction manual from the DiaSorin RIA kit manufacturer for the reformulated assay (catalogue no. 68100E).  4 DEQAS (23) recovery experiments showed >80% recovery of 25(OH)D2 and 25(OH)D3, but results may be invalid due to effects of ethanol spiking on antibody binding (10).  5 CDC QC pool data published with laboratory documentation (18).  6 CDC QC pool data published with laboratory documentation (19–21).  7 CDC confirmed Diasorin estimate of the limit of detection of 1.5 ng/mL (3.75 nmol/L).  8 DEQAS results showed equivalent recognition of serum 25(OH)D2 and 25(OH)D3 (10); note, INCSTAR and DiaSorin instruction manuals (catalogue no. 68100 and 68100E) provide assessment of specificity for 10 vitamin D metabolites.  9 CDC verified linearity (parallelism) of Diasorin RIA.  10 Visual inspection.  11 Visual inspection from 2000 to 2002; automated SAS program (26) from 2003 to 2006.  12 CDC measured correlation between 2 Diasorin calibrator lot numbers: (lot no. 623520–5) = 0.9188 (lot no. 603520–5) + 1.024 nmol/L; R² = 0.999.  13 Equipment function checks included verification of acceptable performance of manual and robotic pipets, and centrifuge temperature and speed; in addition, assessment of gamma counter background, efficiency, and normalization to achieve acceptable performance parameters was periodically performed.
3 SD from the mean rule, and lot descriptive statistics and Shewhart charts are plotted for the QC data. Distribution statistics of participant data are evaluated for reasonableness. Distribution of participant values is compared by year for trending analysis. In addition, NCHS sends blinded split replicates to the NCEH laboratory and calculates the CV of the replicates.


The regular use of bench and blind QC pools during all NHANES provides a basis to monitor the consistency of assay results over time. In monitoring these results, researchers observed problems of assay fluctuations in 2004 and 2006 (2,14). In addition, an assay difference was observed between the served problems of assay fluctuations in 2004 and 2006 (2,14). Researchers observed changes in vitamin D status over time and against externally established cutoffs of adequacy and insufficiency. The roundtable participants discussed the issues of assay fluctuations and statistical adjustments to address them.

**Harmonizing serum 25(OH)D across the NHANES 2000–2006**

The NHANES Analytical Note for 25(OH)D analysis states that, “… data users should be aware that the 25(OH)D data from the 2000–2006 NHANES were most likely affected by drifts in the assay performance (method bias and imprecision) over time. These assay drifts are likely due to reagent and calibration lot changes in the reformulated DiaSorin assay” (2). The multiyear population-weighted means, selected percentiles, and percentages below 2 cutoff points for serum 25(OH)D for 2001 through 2006 suggest that compared with 2000–2002, the serum 25(OH)D values are consistently higher in the 2003–2004 survey period and consistently lower in the 2005–2006 survey period across the entire 25(OH)D concentration range (5th–95th percentiles) (Table 3). These consistent trends were also seen across age–gender groups (data not shown). These observations suggest a systematic bias. A key question, therefore, became whether these differences are related to true population changes in vitamin D status or are a result of assay drift unrelated to changes in vitamin D status. To evaluate this question more carefully, CDC staff from NCEH and NCHS examined the results of laboratory QC pools across these time periods.

The performance of the QC pools used during the time periods in question shows that during 2004, the serum 25(OH)D concentrations for the high QC pool (HS0303c) and medium (MS0302) QC pools were increased compared with values in 2005 (Fig. 1B). This downward drift started in January 2006 and coincided with a new lot of tracer and antiserum. By April 2006, however, the material for the high QC pool (HS0303c) was exhausted and a new material (HS05472a) was put in place. This material was characterized during March–April 2006 and performed on the same level from April 2006 to April 2007 when analysis of samples from NHANES 2006 was completed. The percent decrease in mean pool values before and during the drift ranged from 5.3% for the low bench pool (LS0301) to 17.1% for the low blind QC pool (Table 5). By using these data from the 5

![FIGURE 1 The NHANES 2003–2004 QC performance data for 25(OH)D (20) (A). The NHANES 2005–2006 QC performance data for 25(OH)D (21) (B). Each inflection point represents the mean of 2 separately processed samples measured in duplicate in each assay. At any given point in time, there are 3 QC performance pools—low, medium, and high. The vertical lines indicate the time when pools were changed. Different line patterns also indicate different pools. The name of each pool is given above its respective data line. Shifts of concern occur within a QC pool (same line pattern, same pool number, within the same vertical lines). Differences between QC pools are not meaningful. Unit conversion: 2.5 nmol/L = 1 ng/mL.](image-url)

### TABLE 3 Multi-year population estimates of serum 25(OH)D from NHANES

<table>
<thead>
<tr>
<th>Survey</th>
<th>Sample</th>
<th>Percentile</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>5th</td>
</tr>
<tr>
<td>2001–2002</td>
<td>7807</td>
<td>60 (58-61)</td>
<td>23</td>
</tr>
<tr>
<td>2003–2004</td>
<td>7399</td>
<td>63 (59-66)</td>
<td>24</td>
</tr>
</tbody>
</table>

1 Values are population-weighted means (95% CI).
2 Values are population-weighted mean 25(OH)D concentrations at the respective percentiles.
3 Values are population-weighted percent (95% CI).
4 Unit conversion: 2.5 nmol/L = 1 ng/mL.

D$_{2004}$ + 1.7992. The SE was 0.0101 for the slope and 0.7063 for the intercept. A correlation coefficient of 0.9998 was obtained.

During 2006, the 25(OH)D concentrations for the high QC pool (HS0303c) were lower compared with values in 2005 (Fig. 1B). This downward drift started in January 2006 and coincided with a new lot of tracer and antiserum. By April 2006, however, the material for the high QC pool (HS0303c) was exhausted and a new material (HS05472a) was put in place. This material was characterized during March–April 2006 and performed on the same level from April 2006 to April 2007 when analysis of samples from NHANES 2006 was completed. The percent decrease in mean pool values before and during the drift ranged from 5.3% for the low bench pool (LS0301) to 17.1% for the low blind QC pool (Table 5).
As an additional criterion, the NCEH performance in the DEQAS proficiency testing program was reviewed for the time periods in question by comparing the NCEH DiaSorin RIA results with the DiaSorin-specific laboratory mean in DEQAS. During 2004, the NCEH analysis of DEQAS samples (mean of 53 nmol/L) performed similar to the DiaSorin-specific laboratory mean level (59 nmol/L). During 2006, the NCEH analysis performed low (mean of 39 nmol/L) compared with the DiaSorin-specific laboratory mean (48 nmol/L). These results indicate that the upward assay drift in 2004 was likely kit-related while the downward assay drift in 2006 could have been limited to the NCEH laboratory.


After the release of the NHANES 2001–2002 serum 25(OH)D data, researchers noted that the serum 25(OH)D concentrations for NHANES 2001–2002 were lower than NHANES III (1988–1994) (1,2). The question then arose as to whether the decreased serum 25(OH)D concentrations in the 2001–2002 surveys were due to assay changes and, therefore, not related to changes in vitamin D status in the population, or whether these changes were unrelated to the assay and were, therefore, likely due to real changes in vitamin D status in the population over this time.

After the discovery of this potential problem, NCEH contacted the manufacturer of the DiaSorin RIA kit and learned that a new antibody with improved binding had been introduced in 1998, a time period that fell between the analysis of the serum 25(OH)D in the NHANES III 1988–1994 and the analysis of the serum for NHANES 2000 and later. The manufacturer also reported adding detergents to a buffer to reduce nonspecific binding, resulting in improved precision (from ~20% to ~10% CV). The manufacturer indicated that these changes enhanced sensitivity of the assay from <7.5 to <3.75 nmol/L (<3 to <1.5 ng/mL).

To address whether these kit changes could be responsible for some or all of the differences between the NHANES 1988–1994 and the NHANES 2000 and later, NCEH designed and conducted a crossover study in 2004. In this study, NCEH reanalyzed a subset of 150 banked serum samples from NHANES III (1988–1994) that had been stored at −70°C using the 1998 reformulated DiaSorin RIA kit [appendix S1 in reference (1)]. The analysis was performed over a 2- to 3-mo period to better account for assay variability (20 samples/run, 1 run/wk). Samples were randomly selected by NCHS to cover the serum 25(OH)D concentration range of ~25–125 nmol/L (15–50 ng/mL). The randomization was based on the original NHANES III (1988–1994) serum 25(OH)D concentration values and did not consider age, sex, race/ethnicity, or any other factor.

After the crossover study was completed, drift was detected in the 2004 QC pools during the 2- to 3-mo period in which the crossover study was conducted. Therefore, the values for the 150 crossover study samples were adjusted to account for this drift. This adjustment was accomplished by applying a Deming regression equation obtained from 5 QC pools [Table 4; 25(OH)D2003 = 0.8924 \( \times \) 25(OH)D2004 + 1.7992 nmol/L] to the 25(OH)D results obtained with the reformulated assay in the crossover study to predict what the results would have been if they had been produced prior to the upward assay drift (i.e. to produce corrected 2004 RIA results). The corrected results were then regressed on the NHANES III values as obtained in 1988–1994 with the original assay for these 150 specimens; a correlation coefficient of 0.8966 was obtained. The Deming equation was the following: NHANES 1988–1994 25(OH)D correct 2004 RIA = 0.8429 \( \times \) NHANES 1988–1994 25(OH)D1988–1994 RIA + 2.5762 nmol/L. (Fig. 2A).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25(OH)D1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>Decrease</td>
</tr>
<tr>
<td></td>
<td>nmol/L</td>
<td>%</td>
</tr>
<tr>
<td>LS0301 Bench</td>
<td>49</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>34 ± 4.2</td>
<td>36 ± 3.6</td>
</tr>
<tr>
<td>MS0302 Bench</td>
<td>49</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>50 ± 4.3</td>
<td>55 ± 4.2</td>
</tr>
<tr>
<td>HS0303 Bench</td>
<td>49</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>106 ± 11</td>
<td>117 ± 10</td>
</tr>
<tr>
<td>Low blind</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>31 ± 4</td>
<td>32 ± 3.6</td>
</tr>
<tr>
<td>High blind</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>68 ± 5.6</td>
<td>74 ± 5.4</td>
</tr>
</tbody>
</table>

1 Values are mean ± SD. Unit conversion: 2.5 nmol/L = 1 ng/mL.

**TABLE 5 QC data for 2006 downward assay drift**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25(OH)D1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>Decrease</td>
</tr>
<tr>
<td></td>
<td>nmol/L</td>
<td>%</td>
</tr>
<tr>
<td>LS0301 Bench</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>34 ± 4.2</td>
<td>32 ± 4.9</td>
</tr>
<tr>
<td>MS0302 Bench</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>50 ± 4.3</td>
<td>44 ± 6.1</td>
</tr>
<tr>
<td>HS0303 Bench</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>106 ± 11</td>
<td>95 ± 13</td>
</tr>
<tr>
<td>Low blind</td>
<td>50</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>31 ± 4</td>
<td>26 ± 4.1</td>
</tr>
<tr>
<td>High blind</td>
<td>50</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>68 ± 5.6</td>
<td>59 ± 7.1</td>
</tr>
</tbody>
</table>

1 Values are mean ± SD of 25(OH)D. Unit conversion: 2.5 nmol/L = 1 ng/mL.
affecting comparisons of serum 25(OH)D data from NHANES 1988–1994 with data from the NHANES 2000–2004 (2). The mean serum 25(OH)D concentrations from the reformulated RIA assay used in NHANES 2000–2004 were 12% lower than the original RIA assay values from the 1988–1994 NHANES before these values were adjusted to be similar to the reformulated RIA (Fig. 2B). Looker et al. (1) found that these assay-related differences between the DiaSorin RIA used in the 1988–1994 NHANES and the reformulated DiaSorin RIA used in the NHANES 2000 and later, which are unrelated to changes in vitamin D status, accounted for much of the difference in the serum 25(OH)D concentrations between the 1988–1994 NHANES III and the 2000–2004 NHANES.

However, the fact that the assay drift in 2004 extended beyond the 2-mo comparison study period and, thus, could affect current population estimates of serum 25(OH)D status when 2004 data were included was not fully recognized prior to publishing population estimates for NHANES 2000–2004 (1). A follow-up letter to the editor by Looker et al. (14) described revised population estimates for NHANES 2000–2004 that were adjusted to account for the 2004 assay drift. When the additional adjustment was made for the 2004 assay drift, the means and percentiles for NHANES 2000–2004 were slightly lower (1–2 nmol/L) and estimates of prevalence of low serum 25(OH)D values were slightly higher (1–2 units) compared with the values that only included adjustments for the original to the reformulated DiaSorin RIA. The need for additional evaluation of the assay drift, including for 2006, was noted (2,14) and was a major reason for convening the roundtable.

**Roundtable discussion of harmonizing across all DiaSorin-based surveys**

The roundtable experts expressed concern that the user needs for longitudinal comparisons and prevalence estimates at the tails of the 25(OH)D distributions likely exceeded the performance capabilities of an immunoassay such as the DiaSorin RIA. However, the roundtable experts also realized that more precise high-throughput measurement procedures (e.g., LC-MS/MS) were not available at the time the NHANES III was conducted (1988–1994). The experts noted that NCEH’s plans to use an LC-MS/MS measurement procedure for analyzing serum 25(OH)D in NHANES 2007 and beyond should improve the precision and accuracy of results.

The roundtable experts were asked whether adjustments were needed in the NHANES data because of the assay fluctuation problems and, if so, what options were available for addressing this issue. The experts agreed that adjustments for the assay fluctuations are needed to avoid misuse of the serum 25(OH)D results. This need is most critical when the data are used for time trend evaluations and prevalence estimates. The need for quick adjustment approaches was underscored by the current needs of an ongoing DRI Committee (17) and the Dietary Guidelines Advisory Committee (16) for population-based means and prevalence estimates.

The roundtable experts identified several options for addressing how best to deal with assay fluctuations; one option was preferable but would require considerable time and therefore would not meet the immediate needs of the DRI or Dietary Guideline panels. The other options were preliminary but relatively quick surrogates for retesting.

The experts agreed that the preferred long-term approach is to reanalyze a subset of serum samples across all the DiaSorin-based surveys (1988–1994 and 2000–2006) by using the LC-MS/MS measurement procedure that will be used for NHANES 2007 and beyond. Traceability to the NIST LC-MS/MS measurement procedure should be enhanced by using the NIST SRM 972. The sampling plan for this study should be large enough to use the data to make prevalence estimates at the tails of the serum 25(OH)D distributions and to represent approximately every 3 mo of survey time. Although this approach would give the greatest confidence that appropriate adjustments for assay fluctuations could be made, it would also take considerable time to complete. Such a study was not considered practical to meet the short-term policy needs of several ongoing expert panels. However, this approach should be seriously considered even if shorter term options are used to meet current needs for information on the vitamin D status of U.S. population groups.

To meet the urgent policy-related needs, the roundtable experts discussed preliminary but relatively quick surrogates for retesting by LC-MS/MS. They identified 2 short-term, preliminary ap-
proaches for adjusting the NHANES 2000–2006 results: 1) a statistical adjustment model that is based on the QC pool data but that better defines the boundaries for the periods of drift; and 2) a statistical approach that uses participant serum 25(OH)D and an ANCOVA model to control for short time effects of assay fluctuations. The experts also commented on the interim approach identified in the NCHS Data Advisory to pool the data from the NHANES 2000–2006 surveys to smooth out the high 2004 and low 2006 data (2). To make NHANES 1988–1994 serum 25(OH)D data comparable with the NHANES 2000–2006, a short-term option was discussed: using the regression equation from the 150-sample crossover study after verifying that a log or other transformation would not improve the statistical characteristics of the Deming regression equation.

A preliminary short-term approach to obtain estimates for the NHANES 2000–2006 is simply to pool the NHANES data for these years. With such pooling, the high values in 2004 and the low values in 2006 would give means and prevalence estimates similar to those that would be obtained with the statistical adjustment models based on QC pool data as done previously by NCEH. However, the consensus of the roundtable experts was that the option of pooling the high and low values was not desirable. It is the result of a chance occurrence that just happens to regress to the 6-y mean.

Another short-term approach for NHANES 2000–2006 is to refine the NCEH statistical adjustments based on QC pools. The performance of the QC pools should indicate where a correction of assay values is needed. However, the boundaries delimiting the periods of assay drift had considerable uncertainty. Approaches to reduce this uncertainty were discussed. For example, the possibility of linking adjustments to RIA kit lot numbers was considered. However, this approach may not be useful, because lot numbers for the assay kit do not change at the same time for all assay components. Assay fluctuations could be affected by several factors such as lot-specific degradation of reagents, different agent lots, or different calibrators. Therefore, kit lot numbers alone would not accurately reflect all of these potential sources of variability, because all do not occur at the same time.

Another short-term approach the roundtable considered is a statistical adjustment of the 2000–2006 NHANES data by using an ANCOVA approach. To adjust for time trends within the 2000–2006 time period, the participant data could be divided into 3-mo time periods, thus creating ~21 3-mo time periods with >1000 sample points in each period. Because the distribution of vitamin D levels is skewed and a log transformation will likely remove most of the skew, the data would first have to be log-transformed. A weighted linear regression of the following form would then be used:

\[ Y_{ij} = \text{intercept} + \text{time period} + \text{other covariates} + \text{error} \]

In this equation, \( Y \) is the log of the predicted value of serum 25(OH)D and \( i \) is a subscript to designate that \( Y_{ij} \) is the predicted value for each NHANES participant.

In this formula, the time period would be entered as a class variable and the other covariates would include age, sex, race/ethnic group, and BMI as well as latitude and season to capture sunlight effects. After the regression is run, an adjusted \( Y_{ij} \) would then be computed, adjusting the observed \( Y_{ij} \) to the \( Y_{ij} \) that would be obtained for a mean time period (i.e., a mean of the study effects over the 21 time periods). The adjusted \( Y_{ij} \) could then be transformed back to the raw scale for interpretability. A regression similar to that reported by Looker et al. (1) could then be performed, relating the adjusted 2000–2006 vitamin D level to the vitamin D level in NHANES III (1988–1994) after first transforming the data to the log scale. This approach has the advantage of being easily implemented with results available in a timely manner. It would adjust the percentile results and the mean values. The experts considered this approach to be more statistically valid than pooling across the 2000–2006 surveys.

The short-term statistical approach for making the 1988–1994 NHANES data comparable with the NHANES 2000–2006 raised one concern. The results of the earlier Deming regression on the 150-sample comparison study between the original DiaSorin RIA kit used in NHANES III (1988–1994) and the reformulated kit used in the NHANES 2000–2006 suggested that more scatter occurred at the high than at the low ends of the serum 25(OH)D values. This finding indicated that the statistical assumption of constant variance across the range of values was not met (Fig. 2A). Several roundtable experts suggested that a log-transformed scale for comparing distributions of serum 25(OH)D concentrations might resolve this problem. However, NCHS and ODS staff stated that they had evaluated a number of transformations, including log transformations, relative to these regression models and that these transformations did not seem to correct the problem. But, because of the concerns expressed by some roundtable experts, NCHS and NCEH staff agreed to reexamine the issue of log transformations of the serum 25(OH)D data used in the Deming regressions. Other roundtable experts suggested that log or other transformations might not solve the problem of nonconstant variance. The nonspecificity and imprecision of the RIA could be possible reasons for the nonconstant variance. If this is the case, log or other transformations might not resolve the problem.

Future Challenges

In addition to evaluating the assay drift problems encountered with the DiaSorin-based surveys, the roundtable experts were asked to comment on plans for analyzing serum 25(OH)D concentrations in future NHANES. Specifically, the issues of concern related to: 1) the need to harmonize serum 25(OH)D results from future surveys (2007 and later) that will be obtained by using an LC-MS/MS measurement procedure with earlier survey results obtained by using the DiaSorin RIA; 2) approaches for integrating the SRM 972 into the assay procedures; and 3) questions as to whether quantification of the C-3-epimer of 25(OH)D3 should be undertaken and, if so, how should it be reported and handled.


Given the difficulties with the DiaSorin RIA fluctuations and imprecision during NHANES III (1988–1994) and the NHANES 2000–2006, NCEH, NCHS, and ODS decided to replace the immunoassay with a potentially more accurate, precise, and specific chromatography-based measurement procedure: specifically, LC-MS/MS to measure serum 25(OH)D3 and 25(OH)D2. Several factors influenced this decision. The first concern was the multiple assay drifts that occurred during the NHANES 2000–2006 for which no effective corrective actions were available to the NCEH because the manufacturer controlled the formulation of the measurement procedure.

Second, in hindsight, the RIA did not meet desirable precision criteria, especially long-term precision goals. Ideally, the CV₃ for
a measurement procedure should be only a fraction of the biological variability within an individual (CV\textsubscript{i}) when measurements are taken at different time points (27,28). Short-term CV\textsubscript{i} was recently estimated from a convenience sample of persons from the NHANES 2000–2002 who had a second phlebotomy visit and for whom serum 25(OH)D was remeasured (29). Based on these within-person duplicate measurements of serum 25(OH)D, the CV\textsubscript{i} for 25(OH)D was estimated to be 11.3%. When general guidelines for deriving quality specification goals from CV\textsubscript{i} values (27,28) were used, CV\textsubscript{A} goals of 2.8% as optimal, 5.7% as desirable, and 8.5% as minimal were determined (29). By using the DiaSorin RIA, the NCEH laboratory obtained long-term representative CV\textsubscript{A} values for the low, medium, and high QC pools of 12.6, 9.7, and 11.8%, respectively. Thus, the goal of desirable, long-term analytical precision was not obtained with the DiaSorin RIA.

With the help of the newly available NIST SRM 972, the NCEH laboratory continues to revise its LC-MS/MS measurement procedure (30) to enhance performance characteristics for 25(OH)D2 and 25(OH)D3 results to be agreeable with the results obtained by the LC-MS/MS reference measurement procedure used by NIST in developing and characterizing their SRM 972. At the time of the Vitamin D roundtable discussions, the key performance characteristics of the assay (Table 6) were that they meet the target CV\textsubscript{A} goal of 8.5% and use a smaller sample volume (i.e. 100 \textmu}L) than was used in the published LC-MS/MS measurement procedure (30). NCEH is continuing to refine this measurement procedure to provide for separation of the serum 3-epi-25(OH)D3. The roundtable experts suggested ways to improve the epimer separation while keeping the assay run times short (e.g. 6–9 min) so that the assay would be suitable for the high throughput and rapid turnaround necessary for NHANES specimen analysis. The 1,25-dihydroxyvitamin D3 product of the epimer [3-epi-1,25-(OH)2D3] has reduced ability to bind to the vitamin D receptor and lower biological activity compared with the predominant nonepimer form of 1,25-dihydroxyvitamin D3 (31); thus, the 3-epi-25(OH)D3 may confound the assessment of vitamin D status.

As is customary when a new measurement procedure for a previously measured analyte is introduced into the NHANES, a bridging or crossover study comparing results from the new and old measurement procedures is conducted. The goal of such a study is to provide a sound basis for deriving adjustment equations to allow for time trend analyses that are reasonably free of measurement procedure bias. NCEH proposed a crossover study in which it planned to select 100 serum samples/survey year (25/(concentration quartile)) over the period of the NHANES 2000–2006 and 150 samples (38/concentration quartile) from each of the 2 periods of NHANES III (1988–1991 and 1992–1994). From these analyses, Deming regressions [e.g. (LC-MS/MS) = m \times (RIA) + b] could be derived for each time period.

Earlier, NCEH compared the serum 25(OH)D from the DiaSorin RIA against values using LC-MS/MS in 551 participant specimens, most of which were from NHANES 2005–2006 (30). In the NCEH comparison, the DiaSorin RIA gave results that had a mean value 9.5% lower than the LC-MS/MS results. It is not known if, or how much, these differences might be related to differences in detection of 3-epi-25(OH)D by the 2 comparison methods or to the timing of the study that predominantly occurred during the time when the QC pools indicated a downward shift in DiaSorin-based results.

The roundtable compared the NCEH results with the analyses described by Roth et al. (12) in which the authors evaluated the accuracy of routinely available methodologies (HPLC-UV, several immunoassays, and a competitive protein binding assay) against results obtained by using a reference measurement procedure (LC-MS/MS). First, the HPLC-UV and LC-MS/MS measurement procedures described in the Roth et al. (12) analysis showed good agreement and no bias. Roth et al. also showed that immunoassays generally gave lower results than chromatography-based measurement procedures for serum 25(OH)D. In both the Roth et al. and NCEH analyses, a high degree of scatter was shown in the Bland-Altman difference plots (% difference vs. mean) comparing immunoassay measurement procedures (y-axis) with LC-MS/MS (x-axis). The roundtable experts noted that the scatter primarily represents poor specificity of the competitive protein binding measurement procedures compared with chromatography-based measurement procedures. Based on the fluctuations of the DiaSorin RIA over 12 y of use by NCEH, a single comparison analysis of periodic NHANES specimens would not provide an adequate crossover study. Thus, NCEH proposed to develop multiple regression equations to account for the reformulation and method fluctuations, with the caveat that their ability to accurately predict measured values will be somewhat limited due to the less than ideal specificity of the immunoassay results.

The roundtable experts agreed with the need for the crossover study and statistical adjustments to compare trends over time. They emphasized that this study should be conducted and completed when or soon after the switch is made to the LC-MS/MS measurement procedure. The experts noted the need for log or other appropriate transformations to address the concentration-dependent relationships that are likely to occur based on past experiences with comparing serum 25(OH)D concentrations across 2 assays. They also noted that these adjustment equations will be more complicated than earlier experiences, because they will need to address both serum 25(OH)D2 and 25(OH)D3 rather than just total serum 25(OH)D.

Moreover, this study, unlike the earlier crossover study to address the DiaSorin RIA kit reformulation, compares the results of a more precise and specific measurement procedure (LC-MS/MS) with the results of a less precise and specific measurement procedure (DiaSorin RIA). The experts were concerned that the earlier NCEH comparisons between the DiaSorin RIA and the first LC-MS/MS measurement procedure.
(30) showed a larger scatter of the DiaSorin results at the lower concentrations compared with the higher concentrations. This result is important given that the levels where policy decisions are being made are at the low end of the distribution curve. The experts noted that the scientific community generally knows from DEQAS reports that the DiaSorin RIA gives lower values than the LC-MS/MS measurement procedure (32), but the experts did not think that the concentration dependence of this relationship was generally recognized. This nonlinearity may indicate that a linear regression may not be the appropriate statistical approach to use and that other regression techniques such as piecewise or polynomial regression may need to be considered.

The roundtable experts had an extensive discussion on the design of the crossover study, such as the appropriate sampling approach and sample size. The experts liked the idea of annual or short-term regression equations, because they will address the assay fluctuation and reformulation problems that have occurred in the past as well as some problems that may not have been identified because they are below the detection level. The experts noted that some pooling of the data may be possible once the results of the annual/short-term regressions are available but that the first step should be to examine multiple regressions over relatively short time periods.

The roundtable experts also noted that the design of the crossover study should be guided by the areas of uncertainty, including, but not necessarily limited to, the assay fluctuation problems in 2004 and 2006 or greater uncertainties in specific segments of the serum 25(OH)D distribution curve. Most critical to NHANES users who want to make prevalence estimates are the uncertainties around the intercept and the slope of the regression lines for assay comparability across different time periods. Of particular concern is the great uncertainty around the intercept of a previous statistical adjustment equation; the experts felt that this issue could remain a concern for the planned crossover study. For example, the intercept (95% CI) for the Deming regression between the corrected reformulated DiaSorin assay and the original DiaSorin assay was 2.58 (−2.46 to 7.61) and the slope was 0.84 (0.73–0.93). In making sample size decisions, the user needs of NHANES data for time trend analyses and prevalence estimates at the distribution tails need to be kept in mind. The key question is: What 95% CI are most acceptable for the intercept and the slope from both a practical and statistical perspective? Once the 95% CI are determined, one can calculate the ideal sample size and sampling plan. Having a random sample of values across the whole serum 25(OH)D distribution range is important.

**SRM for serum 25(OH)D and its integration into the NHANES analyses**

**SRM 972 vitamin D in human serum.** The NIST SRM 972 (15), released in June 2009, “… is intended for use as an accuracy control in the critical evaluation of methods for determining the amount of substance concentration of vitamin D metabolites in human serum. This SRM can also be used as a quality assurance tool for assigning values to in-house control materials for these constituents.”

Upon initiating work on this project, NIST convened a panel of experts that included Neal Binkley of the University of Wisconsin, Bruce Hollis of the Medical University of South Carolina, Reinhold Vieth of the University of Toronto/Mount Sinai Hospital, Robert Heaney of Creighton University, and Donald Wiebe of the University of Wisconsin. NIST interacted with most of these experts on several occasions to solicit input on the desired characteristics of the SRM 972, including the need for multiple serum pools with varying concentrations of the analytes of interest. In developing SRM 972, NIST recognized that several potential analytical issues had previously been identified, including equivalency of measurement procedure response for serum 25(OH)D2 and 25(OH)D3, potential biases arising from the presence of 3-epimers of serum 25(OH)D in mass spectrometric measurement procedures, and the need to detect serum 25(OH)D at the relatively low concentrations traditionally used to define deficient and insufficient vitamin D status.

The roundtable discussion about the SRM 972 materials focused on 5 general issues: 1) the 4 materials contained in the SRM 972; 2) the analytical methodology used to derive the certified values for the reference materials; 3) commutability issues; 4) NIST plans for development of calibration materials for serum 25(OH)D; and 5) a quality assurance program.

As purchased, a set of the NIST SRM 972 consists of 4 levels of frozen serum with designated concentrations plus or minus the expanded uncertainty of vitamin D metabolites: 1) a normal human serum pool with a certified 25(OH)D3 concentration of 59.6 ± 2.1 nmol/L; 2) a blend of normal human serum and horse serum to obtain approximately one-half the level of 25(OH)D3 (30.8 ± 1.5 nmol/L); 3) a normal human serum spiked with an amount of 25(OH)D2 (64.1 ± 4.8 nmol/L) approximately equivalent to the concentrations of 25(OH)D3; and 4) a normal human serum spiked with 3-epi-25(OH)D3 (94.1 ± 2.9 nmol/L) (Table 7) (15). NIST used this process of diluting and spiking to achieve desired concentrations of serum 25(OH)D2 and 25(OH)D3 after it encountered difficulties in obtaining serum with naturally occurring low 25(OH)D3 concentrations and naturally occurring high 25(OH)D2 concentrations. The roundtable participants generally agreed that for future releases of these reference materials, obtaining serum pools from individuals with naturally high and low concentrations of serum 25(OH)D2 and 25(OH)D3 rather than using in vitro supplementation or dilution with nonhuman serum would be highly preferable.

The assignment by NIST of the certified values for SRM 972 was based upon the combination of results from iso-dilution LC-MS and LC-MS/MS results provided by NIST and results from LC-MS/MS provided by NCEH (15). As part of the value assignment process, NIST developed its own LC-MS and LC-MS/MS methodology (15,33). For the LC-MS/MS analysis, NIST performed a liquid/liquid extraction and used isotopically labeled internal standards for both 25(OH)D2 and 25(OH)D3. The sample extracts were analyzed by using 2 strategies, one focused on measuring 25(OH)D3 and 3-epi-25(OH)D3 and the other focused on measuring 25(OH)D2. The measurement procedures developed by NIST were optimized for value assignment of reference materials and should not be considered a high throughput method that can be applied to routine measurements. NIST tried different chromatographic columns, including chiral stationary phases, to achieve full chromatographic resolution of 25(OH)D3 and 3-epi-25(OH)D3.

NIST also conducted recovery studies by using a charcoal-stripped serum in which nonpolar materials were removed, such as vitamin D metabolites with very low 25(OH)D3 concentrations (obtained from DiaSorin), and a commercial serum with very low levels of 25(OH)D2 (15,33). NIST conducted spiking studies adding the internal standards at concentrations of ~50 nmol/L, either before or after the extraction process.

The recovery of 25(OH)D2 was 97% (0.4% CV, n = 5) and of 25(OH)D3 was 92% (0.9% CV, n = 5). NIST also examined different equilibration times to confirm that the internal standard...
was behaving the same as the analyte throughout the extraction process.

The roundtable experts agreed that the scientific rigor used to develop the NIST LC-MS/MS methodology for the SRM 972 project was sufficient to consider it a higher order reference measurement procedure. The roundtable experts noted that not all LC-MS/MS measurement procedures are the same, because extractions and chromatographic procedures vary widely. They noted that the NIST measurement procedure has a very long chromatographic run time and adequately maintains a high extraction efficiency. It also minimizes specificity and ion suppression issues that commonly account for differences among clinical and research laboratories using LC-MS/MS measurement procedures (34). Therefore, the roundtable experts suggested that the NIST measurement procedure is a high performance measurement procedure and that developing another higher order reference measurement procedure is not needed at this time. Since the July 2009 roundtable, an article describing the NIST LC-MS/MS measurement procedure has been accepted for publication (33). This is the first step toward nominating the NIST method as a reference measurement procedure. The second step is for NIST to submit their candidate reference method to the Joint Committee for Traceability in Laboratory Medicine, a process that has now been initiated. The Joint Committee for Traceability in Laboratory Medicine is the body that maintains a database of approved reference materials and reference measurement procedures (35).

As with many reference materials, commutability is often an issue of concern, particularly regarding the suitability of spiked or altered serum pools to serve as reference materials for immunoassays. Commutability means that the reference material behaves the same as a patient sample would (7,36,37). A reference material may not be commutable for several reasons, including nonhuman matrix, lyophilization, filtering, stripping, storage, or spiking. In addition, whether a reference material is going to be commutable is not always predictable. Commutability of the SRM 972 materials was not assessed prior to the release of the SRM and, therefore, NIST included a cautionary statement in the Certificate of Analysis indicating that the user should decide which levels will be useful for a particular measurement procedure (15). NIST plans to conduct a commutability study in the near future.

The roundtable experts strongly emphasized the need for a commutability study as soon as possible, similar to what was done with creatinine materials (38). Without the results of a commutability study, the use of the SRM 972 as a reference material, especially in clinical measurement procedures, requires caution (39). The experts felt that putting a warning in the certificate is inadequate, because individual users may not have the knowledge or skill to address this issue. At a minimum, instructions and guidelines to assess matrix effects (34,37,40) should be provided to users.

The roundtable experts provided some insights on the best strategies to effectively study the commutability of the SRM 972 (37–40). The study should be done by using individual patient samples, because pooling of serum from multiple donors can result in cross-reactivity that can interfere with certain measurement procedures. Clinical procedures are designed to measure individual patient samples, not lyophilized or otherwise altered serum pools. Confirming that the SRM 972 materials can give results that accurately reflect the measurand concentration as measured by the clinical procedure is necessary. The uncertainty is how the field measurement procedures can work with the SRM 972 material compared with an actual single donor specimen.

The commutability study should be designed to compare results across different laboratories and multiple measurement procedures, including MS, chromatography, and immunoassay. Mass spectrometric laboratories should be included in the commutability study, because their results are variable. The results from these laboratories should be compared against the high-level reference measurement procedure of NIST. Initial efforts could focus on the immunoassays, because this procedure currently raises the greatest concerns. Similar experiences with a study of creatinine measurement showed that manufacturer involvement in many aspects of improving measurement procedures is essential, including their involvement in commutability studies (27,38). A key issue is to make sure that the reference materials are in the scatter of the actual patient samples on a bivariate plot of the results by the field and by the reference measurement procedures. Commutability study results may show that some reference materials are commutable with some measurement procedures and not with others.

NIST recently also developed ethanol-based calibration solutions for 25(OH)D2 and 25(OH)D3 to help meet user needs for reliable reference materials (41). The 2 calibration solutions were prepared gravimetrically, but NIST also made measurements by using LC-MS and tandem MS for the value assignments. The solutions are ~10 times more concentrated than the levels one would find in serum (e.g., ~600–800 nmol/L) so they can be diluted to make a calibration curve. NIST provided certified values for the 2 vitamin D metabolites in the calibration materials.

### TABLE 7 Certified concentration values for vitamin D metabolites in SRM 972

<table>
<thead>
<tr>
<th>Vial</th>
<th>25(OH)D3</th>
<th>25(OH)D2</th>
<th>C-3 epimer of 25(OH)D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Normal human serum</td>
<td>59.6 ± 2.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Level 2: 1 diluted 1:1 with horse serum</td>
<td>30.8 ± 1.5</td>
<td>4.14 ± 0.19</td>
<td>–</td>
</tr>
<tr>
<td>Level 3: Normal human serum + 25(OH)D2</td>
<td>46.2 ± 2.8</td>
<td>64.1 ± 4.8</td>
<td>–</td>
</tr>
<tr>
<td>Level 4: 1 + 25(OH)D2 + C-3 epimer of 25(OH)D3</td>
<td>82.3 ± 2.0</td>
<td>5.81 ± 0.52</td>
<td>94.1 ± 2.9</td>
</tr>
</tbody>
</table>

1 Values are certified concentration values plus or minus the expanded uncertainty. Expanded uncertainty is the uncertainty in the certified value, calculated according to the measurement procedure described in the International Organization for Standardization Guide, is expressed as an expanded uncertainty, U. The expanded uncertainty is calculated as $U = kuc$, where uc is intended to represent, at the level of 1 SD, the combined effect of between-laboratory and within-laboratory components of uncertainty. The coverage factor (k) is determined from the Student’s t-distribution corresponding to the appropriate associated degrees of freedom, and ~95% confidence for each analyte (15).

2 Complete information on the development, use, and interpretation of the SRM 972 materials can be found in the NIST Certificate of Analysis (15).

3 For 25(OH)D concentration values, molar concentration levels were calculated from mass concentration levels using the relative molecular masses. The equivalent factors for converting ng/mL to nmol/L are 2.4233 for 25(OH)D2 and 2.4959 for 25(OH)D3 and 3-epi-25(OH)D3 (15).

NIST plans to conduct a commutability study, because their results are variable. The results from these laboratories should be compared against the high-level reference measurement procedure of NIST. Initial efforts could focus on the immunoassays, because this procedure currently raises the greatest concerns. Similar experiences with a study of creatinine measurement showed that manufacturer involvement in many aspects of improving measurement procedures is essential, including their involvement in commutability studies (27,38). A key issue is to make sure that the reference materials are in the scatter of the actual patient samples on a bivariate plot of the results by the field and by the reference measurement procedures. Commutability study results may show that some reference materials are commutable with some measurement procedures and not with others.

NIST recently also developed ethanol-based calibration solutions for 25(OH)D2 and 25(OH)D3 to help meet user needs for reliable reference materials (41). The 2 calibration solutions were prepared gravimetrically, but NIST also made measurements by using LC-MS and tandem MS for the value assignments. The solutions are ~10 times more concentrated than the levels one would find in serum (e.g., ~600–800 nmol/L) so they can be diluted to make a calibration curve. NIST provided certified values for the 2 vitamin D metabolites in the calibration materials.
The roundtable experts supported the need for well-characterized calibration solutions, but they also raised several questions for further evaluation. For example, they asked about the calibrator currently being in ethanol rather than serum, which could be a concern, particularly for an immunoassay. NIST indicated that, if users would find serum-based calibrators useful, it would consider them for future releases.

NIST also reported on its plans to initiate a quality assurance program for measuring serum 25(OH)D. The intent of this program is not to replace the DEQAS and College of American Pathologists programs. Rather, the goal is to provide laboratories participating in the NIST quality assurance program with materials that have assigned values (known as “true values”) instead of just comparing the mean value of one laboratory with the mean of a group of laboratories using the same measurement procedure. NIST uses SRM 972 as a control material and provides participating laboratories with a material with known values. NIST will provide feedback to participating laboratories on how well their measurement procedures are performing. After participating laboratories have experience with this program, NIST hopes to convene a workshop to foster discussions among the laboratory staff and the NIST scientists.

Integrating NIST reference materials into the NHANES analyses. NCEH proposes to incorporate the 2 types of NIST reference materials into future NHANES analyses: SRM 972 with its 4 levels of serum-based reference values and the yet-to-be released SRM 2972 solvent-based calibration solutions for 25(OH)D2 and 25(OH)D3. NCEH is still refining an LC-MS/MS methodology for use in the NHANES, starting with the stored specimens from 2007. However, early comparisons of the NCEH results with the 4 levels of SRM 972 materials show that, for serum 25(OH)D3, the NCEH values are slightly biased on the high side (NCEH:NIST ratio for levels 1, 2, 3, and 4, respectively were 1.02, 1.03, 1.04, and 1.08). Larger differences, however, were observed for serum 25(OH)D2 (NCEH:NIST ratio for levels 1, 2, 3, and 4, respectively were 0.82, 1.30, 1.07, and 1.08). The 25(OH)D2 concentrations in the materials for levels 1 and 2 are less than the NCEH level of detection. NCEH continues to refine its calibration methodology to attempt to minimize these differences.

NCEH proposes to achieve traceability to the SRM materials in future NHANES by comparing and adjusting in-house calibrators to SRM 2972 solvent-based reference material. Serum-based SRM 972 will be used for calibration verification at regular intervals. The roundtable experts felt that the inclusion of these reference materials should enhance the accuracy of comparisons across the serum 25(OH)D results, both for future NHANES analyses and for reanalyzing serum 25(OH)D samples from past NHANES. However, the experts expressed one note of caution. The traceability to high-level reference measurement procedures does not fully solve the issue of analytical specificity of the competitive protein binding assays. The experts suggested that practices consistent with the current recommendations for improving serum creatinine measurement procedures would be helpful (27,38). For example, these would include continuation of NCEH efforts to identify analytical performance measures that incorporate considerations of biological variability and user reference values, participate in proficiency testing programs, and ensure analytical specificity. In addition, adequate communication to users as to how interpretation of 25(OH)D results is likely affected by changes in measurement procedures and traceability to reference materials and methods is needed.

What about the serum 3-epi-25(OH)D3?
The planned conversion from the DiaSorin RIA to an LC-MS/MS methodology beginning with the NHANES blood samples collected in 2007 raised the issue of how to address the measurement of the 3-epi-25(OH)D3. Epimers have identical chemical structures except for a single site of molecular asymmetry (in this case, C-3α- vs. C-3β-hydroxy). Therefore, they are expected to display very similar chromatography and give rise to the same ion pairs when LC-MS/MS analytical methodology is used (42).

Although the 3-epi-25(OH)D3 is not measured by using the DiaSorin RIA, it is generally measured along with 25(OH)D3 in MS/MS measurement procedures, because both the 3-epi-25(OH)D3 and the 25(OH)D3 are structurally similar and have the same mass. Failure to chromatographically separate the epimer of 25(OH)D3 could result in overestimating the actual levels of serum 25(OH)D3. Moreover, because the epimer is converted to a form of 1,25-dihydroxvitamin D that differs in biological activity from the hormone produced from 25(OH)D3 (31), distinguishing between the amounts of 25(OH)D3 and 3-epi-25(OH)D3 could have biological relevance. Although the biological activity of the 3-epi-25(OH)D3 is currently unclear, capturing information about population levels of the epimer may be helpful until more is known about its function. Measuring the C-3 epimer of 25(OH)D2 does not seem to be a concern, because it has not been detected in human serum (42).

The LC-MS/MS measurement procedure developed by NIST as of July 2009 does not separate the 3-epi-25(OH)D3 from 25(OH)D3. The roundtable experts were asked, therefore, whether the measurement procedure should be further refined to allow for this separation and, if so, how best to report results. Further refinements would delay the analysis of blood samples collected from 2007 to now. To address this question, the roundtable discussions focused on 2 issues: the usefulness of epimer separation and the analytical challenges related to separating and quantifying the epimer.

The current science relative to interpreting the epimer results is at a very early stage of development. The first questions are: What is the population prevalence of detectable levels of serum 3-epi 25(OH)D3 and what serum concentrations are likely? A publication by Singh et al. (42) reported that in their laboratory, ~23% of infants younger than 1 y had detectable levels of the epimer and that the epimer accounted for 8.7-61.1% of the total serum 25(OH)D concentration. No epimers were detected in other groups (children 1 y or older and adults with or without liver disease). However, NIST reported finding 3-epi-25(OH)D3 in levels 1, 2, and 3 of SRM 972. None of these materials had been spiked with the epimer; all had been collected from adult blood donors.

To confirm that this observation was not unique to the SRM 972 serum pools, NIST looked at a plasma pool (SRM 1950 Metabolites in Human Plasma) currently under development. This pool was prepared with samples from 100 healthy men and women with different race and ethnic backgrounds. NIST also observed the presence of 3-epi-25(OH)D3 in this pool. In addition, NIST looked at 20 different DEQAS samples and saw the epimer in all 20 samples in the range of 1–8%. NCEH checked its NHANES QC pools and also found small amounts of the epimer.

In addition, 2 members of the roundtable reported finding the 3-epi-25(OH)D3 in patient samples from their laboratories. One laboratory concluded that adult patients can have detectable amounts of this epimer with a range of 4–27% of the total 25(OH)D3 [25(OH)D3 + 3-epi-25(OH)D3]. Unlike Singh et al. (42), this laboratory found poor association between patient age and the 3-epi-25(OH)D3. The laboratory also found poor association between the concentrations of 3-epi-25(OH)D3 and...
25(OH)D3. Results from a second laboratory showed an epimer concentration of ~6% of the total serum 25(OH)D3 concentration [i.e. 25(OH)D3 + 3-epi-25(OH)D3] in adult patients. Thus, the roundtable experts concluded that the 3-epi-25(OH)D3 was likely common among U.S. population groups with a subset of individuals showing fairly high concentrations.

Questions then arose as to the source and biological function of the 3-epi-25(OH)D3. Whether the source is from fortified foods and dietary supplements or produced endogenously is unknown. The biological function of the epimeric form is very uncertain. In human intestinal cell culture assays, the 1α, 3α-epi-25(OH)2D3 had a reduced ability to bind to the vitamin D receptor and therefore a reduced capability of stimulating gene expression and calcium transport activity (45% as active) (31).

These questions remain: Does this in vitro finding occur in vivo? Is the 3-epi-25(OH)D3 converted to the C-3 epimeric form of 1, 25(OH)2D3 in vivo? And what is the biological function of the epimeric form in vivo?

The roundtable experts then discussed the analytical challenges in separating and quantifying the 3-epi-25(OH)D3 in the NHANES. Given the number of NHANES samples and the need for timely analysis once the new measurement procedure is implemented, a key requirement would be that the chromatographic run time for separating out the epimer would need to be reasonably short. The current NCEH LC-MS/MS measurement procedure that did not separate the epimer required 10 min/sample. Preliminary efforts to separate the epimer resulted in longer run times. However, the roundtable experts noted that measurement procedures with shorter run times have been developed and could be integrated into the NCEH LC-MS/MS measurement procedure. These adaptations would cut retention times while achieving good resolution and enhanced specificity of the serum 25(OH)D3 and the 3-epi-25(OH)D3.

Another issue is the current lack of stable isotopically labeled 3-epi-25(OH)D3 to be used as an internal standard. NCEH proposed to use the deuterium-labeled 25(OH)D3 as a surrogate internal standard for the 3-epi-25(OH)D3. It also proposed to use the assigned values in the 4 levels of SRM 972 for the 3-epi-25(OH)D3 concentrations (i.e. 3.5 nmol/L for level 1, 1.9 nmol/L for level 2, 2.6 nmol/L for level 3, and 94.1 nmol/L for level 4). The level 4 material could be diluted to get a wider range of reference material concentrations. NCEH also needed to determine the 3-epi-25(OH)D3 concentrations in its QC pools; spiking of these materials would likely be necessary to ensure that the pools reflected low, medium, and high levels of the epimer.

Once the analytical needs for the NHANES could be met, the roundtable participants agreed that starting with the 2007 NHANES, NCEH should separate and quantify the serum 3-epi-25(OH)D3 concentrations. This process would result in a more accurate measurement of the serum 25(OH)D3 concentrations. This result was deemed important, because the limited analysis of epimer data by NIST, NCEH, and several roundtable experts suggested that the amount of epimer was not trivial, with some individuals having very high levels. Thus, failure to separate the epimeric form from the 25(OH)D3 could result in an error for some but not all persons. Moreover, the NHANES dietary and biomarker data could be a useful database for examining whether intakes of certain fortified foods and supplements were associated with higher concentrations of the epimer.

**Research and Educational Needs**

The roundtable experts identified 3 areas requiring further follow-up from a research or educational perspective. The first is improving the standardization, harmonization, and accuracy of serum 25(OH)D results across laboratories. The second is research to enhance our understanding of the sources and biological importance of the serum 3-epi-25(OH)D3. The third is a better understanding of the factors that affect serum 25(OH)D concentrations and the interpretation of the biological importance of serum 25(OH)D concentration levels.

The NHANES have encountered problems across time regarding accurate and reproducible results for serum 25(OH)D. In addition, recent publications have identified differences in results among different measurement procedures and within laboratories over time (8–13,32). These challenges underscore the urgent need for improved measurement procedures and harmonization of results across available measurement procedures and among laboratories. The roundtable experts felt that all laboratories—clinical, research, and public health—should be able to produce serum 25(OH)D results that are accurate and traceable to a high-level reference measurement procedure. This harmonization is essential if the science is to move forward; all laboratories need to compare their results with commonly accepted cutpoints of status and published data relating serum 25(OH)D concentrations to intake levels and functional or clinical outcomes. Moreover, many laboratories need to be able to accurately monitor changes over time either in individual patients or groups. Because most of the published research results are based on 25(OH)D measurement procedures that likely give values that differ from the values obtained with reference materials and methods, the interpretation of previously established normal ranges, cutpoints of adequacy, as well as dose-response relationships between intakes, 25(OH)D concentrations, and clinical outcomes will need to be reevaluated.

Ultimately, however, the roundtable experts concluded that the primary responsibility for improvements in serum 25(OH)D measurement procedures should come from in vitro diagnostic device manufacturers rather than individual clinical and research laboratories. A recent example comes from a working group looking at improving creatinine measurement procedures (27). This group recommended that manufacturers play a key role in matching the calibrations of their measurement procedures to high level reference measurement procedures. The roundtable experts added one note of caution: the traceability to a high-level reference measurement procedure does not solve the problem of analytical specificity, because an individual patient can have an unknown influence quantity (e.g. interferent) in his or her serum. Again, the roundtable members suggested that experts should work with manufacturers to improve analytical specificity so what the measurement procedures are intended to measure is measured (27,38). The creatinine measurement working group also noted that government agencies such as NIST and NIH in collaboration with professional organizations should help support manufacturers in these efforts and in educating laboratories to demand high specificity and accuracy for the measurement procedures and equipment used in their laboratories.

The second identified research item is the need to better understand the source and biological activity of the 3-epi-25(OH)D3. Understanding the source of this epimer was considered to be of
high priority, particularly if it is coming from a food additive or dietary supplement ingredient source. However, the committee noted the current difficulty of measuring this epimer in foods, because the substance in foods would be the epimer of cholecalciferol, not the biomarker found in serum. Measurement procedure development would be needed before the epimeric form could be measured in foods. A faster alternative to evaluating whether foods are a likely source of the epimer might be to analyze the serum of persons consuming different types of fortified foods or supplements to see if a relationship exists between the consumption of specific foods and the presence of the 3-epi-25(OH)D3 in serum. The roundtable experts also noted the need for an understanding of the biological activity of the epimer in vivo.

The 3rd identified research need is to better understand the factors that affect serum 25(OH)D concentrations. Included in this concern was the need for more analyses of NHANES data on different life-stage groups and on the effects of other factors such as medication use and health status. Of particular interest was the possibility of using NHANES serum samples to evaluate associations between serum 25(OH)D concentrations and genetic and ethnic variability. A better understanding of quantitative dose-response relationships between intakes, 25(OH)D concentrations, and clinical outcomes was also identified as needed. This is particularly important as a better understanding of the comparability of 25(OH)D measurements is achieved through the use of reference materials and methods.

Summary and Conclusions
The roundtable discussion, NHANES Monitoring of Serum 25(OH)D: Assay Challenges and Options for Resolving Them, was held late July 2009. Participants included 11 scientists with expertise in laboratory sciences, vitamin D nutrition, and bio-statistics as well as 11 government scientists involved in generating and using the NHANES serum 25(OH)D data. The roundtable experts were asked to identify the key science-based issues affecting the accuracy of serum 25(OH)D measurements in NHANES and options for effectively addressing these issues. Topics included: 1) the challenges of dealing with measurement procedure fluctuations that occurred between the NHANES III (1988–1994) and the NHANES 2000 and during the NHANES 2000–2006; 2) approaches for transitioning between the immunoassay used in the NHANES III (1988–1994) and NHANES 2000–2006 surveys to the LC-MS/MS measurement procedure planned for use in the 2007 and later surveys; 3) approaches for integrating the recently available SRM from NIST (SRM 972) into the NHANES analyses of serum 25(OH)D; 4) whether and how to address measuring the serum 3-epi-25(OH)D3 in NHANES 2007 and later; and 5) identification of research and educational needs.

The roundtable experts agreed on the following key points. The NHANES data must be adjusted to control for assay fluctuations to ensure that reasonably accurate estimates of time trends and prevalences can be made. Failure to do so will result in misinterpretation and erroneous conclusions. The roundtable experts identified the preferred approach for determining adjustment equations as a reanalysis of a subsample of participant specimens from all DiaSorin-based surveys. To do so, the LC-MS/MS measurement procedure developed for the NHANES 2007 and beyond would be used in conjunction with the NIST SRM 972 and SRM 2972 to ensure traceability to the NIST reference measurement procedure. However, because interim estimates of vitamin D status are urgently needed for several ongoing public policy activities, the roundtable experts suggested that several short-term approaches to address assay fluctuations can serve as surrogates until results from the preferred reanalysis project become available.

The LC-MS/MS reference measurement procedure developed by NIST to certify the 25(OH)D concentrations in the 4 reference materials has undergone sufficient scientific rigor so that it can be used as a higher order reference measurement procedure by the scientific community. However, full confidence in the recently released SRM 972 materials and the in-process calibration reference materials (SRM 2972) depends on commutability studies that include a number of different clinical measurement procedures and laboratories.

Improving the quality of the laboratory measurements of serum 25(OH)D across all types of laboratories—clinical, public health, and research—is urgently needed. To be effective and efficient, manufacturers with technical support from the federal government and professional organizations have a critical role in these efforts.

Starting with the analysis of the 2007 NHANES when serum 25(OH)D will be analyzed by using LC-MS/MS, separating out and quantifying the serum 3-epi-25(OH)D3 would be useful. These results could help identify the prevalence of this epimer among U.S. population groups and evaluate possible relationships between epimer levels and intakes of fortified foods and dietary supplements that contain vitamin D. In addition, research on the source and the biological fate and function of epimers of 25(OH)D is needed.

Acknowledgments
We thank Dr. Anne Thurn and Claudia Faigen of the ODS, NIH, and Michael Bykowski and Mike Schultz of Consolidated Solutions and Innovations for their outstanding logistical, organizational, and follow-up support of the roundtable discussions. M.F.P. and C.L.J. conceived and sponsored the roundtable project. A working group consisting of E.A.Y., M.F.P., C.L.J., C.M.P., R.L.S., K.W.P., D.A.L., P.M.C., A.C.L., and C.S. planned and organized the July 2009 roundtable meeting. C.M.P., R.L.S., K.W.P., D.A.L., A.N.H., and G.L.L. prepared and presented data. E.A.Y. wrote the paper with considerable input from C.M.P., R.L.S., K.W.P., D.A.L., A.C.L., and C.S. E.A.Y., M.F.P., and C.L.J. had primary responsibility for the final paper content. All authors fully participated in the roundtable discussions. All authors read and approved the final manuscript.

Literature Cited