

# Advances in standardization of laboratory measurement procedures: implications for measuring biomarkers of folate and vitamin B-12 status in NHANES<sup>1-4</sup>

Jay L Bock and John H Eckfeldt

## ABSTRACT

Population studies such as NHANES analyze large numbers of laboratory measurements and are often performed in different laboratories using different measurement procedures and over an extended period of time. Correct clinical and epidemiologic interpretations of the results depend on the accuracy of those measurements. Unfortunately, considerable variability has been observed among assays for folate, vitamin B-12, and related biomarkers. In the past few decades, the science of metrology has advanced considerably, with the development of improved primary reference measurement procedures and high-level reference materials, which can serve as the basis for accurate measurement. A rigorous approach has been established for making field methods traceable to the highest-level reference measurement procedures and reference materials. This article reviews some basic principles of metrology and describes their recent application to measurements of folate and vitamin B-12. *Am J Clin Nutr* doi: 10.3945/ajcn.111.013359.

## INTRODUCTION

As detailed elsewhere in this supplement issue, variability in laboratory measurement procedures has been a major challenge in assessing nutritional status for folate and vitamin B-12 (1-3). Similar problems have existed throughout laboratory medicine, but over the last few decades improved analytic technologies, along with international agreements on improved approaches to test standardization, have increased the accuracy of clinical laboratory tests. Modern concepts of measurement accuracy and standardization are unfamiliar to many health scientists, yet they are now recognized as essential for good patient care as well as for research and epidemiologic applications (4). A recent expert panel review of serum 25-hydroxyvitamin D measurements highlighted the importance of these concepts for nutrition applications (5). The purpose of the current article is to briefly review these concepts in the context of vitamin B-12 and folate assessment.

## MEASUREMENT STANDARDS, ORGANIZATIONS, AND GUIDELINES

Advances in measurement science, or *metrology*, depend on achieving international consensus on best practices. One fundamental requirement is a consistent set of measurement units. The International Bureau of Weights and Measures (BIPM) has maintained the primary measurement standards for the International System of Units (SI) since 1875. The SI continues to

evolve to provide the most precise and universal unit definitions. The important SI units for clinical laboratory measurements are those for time (second, defined relative to the cesium clock), length (meter, defined in terms of the speed of light in a vacuum and the second), mass (kilogram, currently still defined based on a material object), and amount of substance (mole, currently defined as the number of atoms in 0.012 kg of carbon 12). The conventional unit of volume for clinical laboratory measurements is the liter, which is defined as 1/1000th of a cubic meter.

New international agreements on metrology come mainly through the Geneva-based International Organization for Standardization (ISO), a federation of 163 national standards organizations. In 1984 the ISO worked with the BIPM and 2 other international organizations to codify a standard International Vocabulary of Metrology (VIM). The VIM is important because many common terms such as *quantity*, *accuracy*, *precision*, and so forth, have been used with a variety of meanings, causing confusion (6). The original VIM was directed mainly toward industrial-type measurements. It has recently been revised, with input from the International Federation of Clinical Chemistry and Laboratory Medicine, to include terms more relevant to clinical laboratory measurements. The current VIM third edition is freely available on the BIPM website (7). ISO Technical Committee 212 also produced a written standard (17511:2003), specifically directed to the clinical laboratory community that incorporated and extended the concepts of VIM (8). Definitions of terms below that are in quotations are from the VIM third edition unless otherwise indicated. Another important document produced by BIPM and other organizations, titled *Guide to the Expression of Uncertainty*

<sup>1</sup> From the Department of Pathology, Stony Brook University Medical Center, Stony Brook, NY (JLB), and the Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN (JHE).

<sup>2</sup> Presented at the conference "NHANES Monitoring of Biomarkers of Folate and Vitamin B-12 Status: a Roundtable Review," held in Rockville, MD, 15-16 July 2010.

<sup>3</sup> The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, the Centers for Disease Control and Prevention, the National Institutes of Health, the US Department of Health and Human Services, or the authors' affiliated institutions.

<sup>4</sup> Address correspondence to JL Bock, Director of Clinical Pathology, Stony Brook University Medical Center, State University of New York at Stony Brook, Stony Brook, NY 11794. E-mail: jay.bock@stonybrook.edu.

doi: 10.3945/ajcn.111.013359.

in *Measurement* (9), details procedures that should be used to measure and express the inherent error in a measurement. All measurements have some uncertainty, and an estimate of uncertainty should be available to assess the reliability of any measurement result.

National metrology institutes, such as the National Institute of Standards and Technology (NIST; formerly the National Bureau of Standards) in the United States, are generally charged with producing reference materials and providing the highest available level of reference measurement procedures. The ISO coordinates these standardization activities. In the United States, various governmental agencies such as the US Food and Drug Administration, and nongovernmental organizations such as the Clinical and Laboratory Standards Institute (CLSI), publish numerous consensus guidelines that have broad international reach, including some with direct relevance to metrology, such as CLSI C53-A on commutable reference materials (10). The Expert Committee on Biological Standardization of the World Health Organization (WHO) also provides international reference preparations for a variety of biological *measurands* (substances being measured), mainly complex biological substances (eg, proteins and hormones) that often have their measured values assigned an activity in International Units rather than a substance concentration in SI units (ie, moles per liter).

Much of the international effort to improve the accuracy of clinical laboratory results revolves around the concept of *traceability*, which is described in detail in ISO International Standard 17511:2003, and which has been reviewed in depth by Vesper and Thienpont (4). Approximately 10 y ago, the BIPM partnered with the International Federation of Clinical Chemistry and Laboratory Medicine and the International Laboratory Accreditation Cooperation to establish the Joint Committee for Traceability in Laboratory Medicine (JCTLM) for measurements made in clinical laboratories. This committee maintains a database of higher-order reference materials and reference measurement procedures that is also available on the BIPM website (<http://www.bipm.org/jctlm>). We describe traceability and related concepts below and discuss them in the context of folate and vitamin B-12 assessments.

## BASIC DEFINITIONS AND PROCEDURES

### Measurand

A measurand is defined as “the quantity intended to be measured.” Unfortunately, many measurement procedures do not measure only what the user intends, and exactly what the user intends to measure is at times ambiguous, especially in complex biological systems with many closely related analytes that have similar but not totally equivalent bioactivity on a mole-per-liter basis. Every measurement has inherent uncertainty related to the ambiguity in the definition of the measurand: this is known as *definitional uncertainty*.

### Accuracy

Accuracy of a measurement procedure is the closeness of the measured value to “a true value of the measurand.” In some clinical situations, absolute accuracy of a laboratory measurement may not be essential: for example, if only a single measurement procedure is used to monitor a given patient’s health condition and reference ranges for that particular measurement

procedure are used for clinical interpretation of results. However, most patients undergo measurement procedures at many different laboratories, often based on very different methodologic principles, and test results are often compared with cutoffs in the literature derived from very different measurement procedures. If different research studies or population surveys are to be compared, and in most clinical situations, accuracy of clinical laboratory results becomes essential.

The “true value” is an unattainable ideal, but it may be defined operationally based on appropriate reference materials (*see below*). It is also generally accepted that certain methods are preferable to others to determine the “true value” of a measurand in a sample. For example, competitive-binding immunoassay may be a satisfactory analytic methodology for many clinical applications, but it is based on a single measurement that has known susceptibility to cross-reactions, and it requires external calibration. Mass spectrometric methods, on the other hand, can generate a great deal of molecule-specific information that increases the confidence in a measurement result. These methods also can use internal standards labeled with stable isotopes (eg, deuterium or carbon 13) to correct for any loss of analyte in the various steps of the measurement process. The great progress in mass spectrometry instrumentation over the last several years has been a major benefit in improving clinical laboratory standardization.

### Precision

Any measurement procedure should faithfully reproduce its own result, a property that may be referred to as *repeatability*, *reproducibility*, or *precision*. Usually error follows a Gaussian distribution, and SD (or CV) describes the measurement procedure’s precision. Note that the term *accuracy* in international metrology encompasses the concepts of both the trueness and precision of a measurement procedure. Thus, a measurement procedure with poor precision cannot be very accurate, even if an average of a number of measured values is extremely close to the true value of the measurand in the sample. Alternatively, a measurement procedure may be precise but inaccurate due to lack of trueness (ie, the procedure is biased). As an example mentioned in this conference, the Bio-Rad Quanta Phase II procedure for serum and red blood cell folate (Bio-Rad, Hercules, CA) was found to have acceptable precision but substantial negative bias compared with the microbiological assay (1).

### Specificity and interference

A measurement procedure must also possess specificity, which means freedom from interference. In metrological terminology, an interference is an *influence quantity*, defined as a “quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement” (7). CLSI Standard EP7-A2 describes common causes of interference and provides advice on assessing laboratory measurement procedures for interference (eg, hemolysis, icterus, and lipemia) (11). In some cases, the interference is due to the bulk composition of a specimen and not to one identifiable compound—clinical laboratories often refer to such interference as a *matrix effect*. For example, ethanol, saline, and pooled human serum represent 3 very different matrices, and routine measurement procedures that clinical laboratories use are likely to respond very

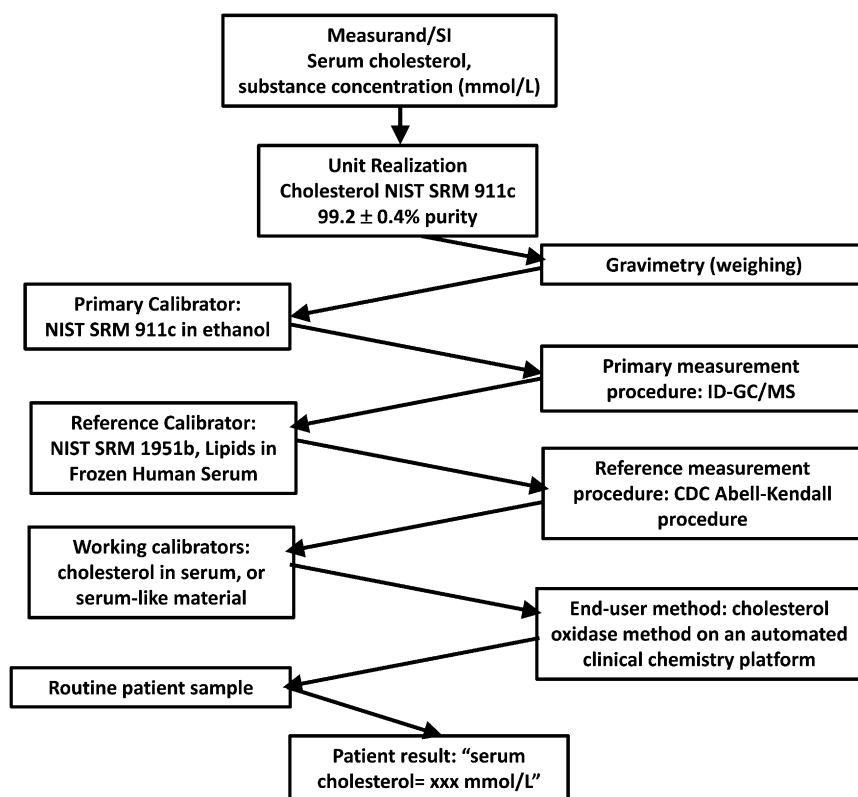
differently to the same concentration of an analyte in these different matrices. Much more subtle matrix differences (eg, animal serum or lyophilized human serum compared with fresh human serum) also have a demonstrated effect on some laboratory measurement procedures (12). Matrix effects give rise to concerns about the *commutability* of reference materials, which we discuss below in more detail because it is often ignored or underappreciated as a major hurdle to providing accurate clinical laboratory results.

### Reference materials, traceability, and commutability

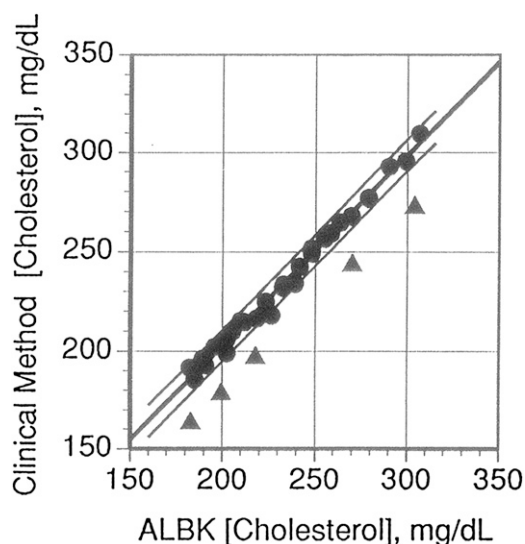
Reference materials are well-characterized materials that are used as calibrators to calibrate the measurement procedure or as trueness controls to check the accuracy of the measurement procedure. A useful reference material first and foremost requires a clear definition of the measurand. This is straightforward only in the case of well-defined, low-molecular-weight chemical substances such as cholesterol, sodium, and glucose. For these substances, primary reference materials or standards, which are substances of known and very high (>99%) purity, have been developed by national metrology organizations such as the NIST.

Relating routine measurements to the primary trueness standard is based, in the modern metrological approach, on a *traceability chain*. Traceability is a “property of a measurement result

whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty” (7). A traceability chain for serum cholesterol, which of the clinical measurands has one of the oldest and most rigorous reference systems, is shown in **Figure 1**. A very important aspect of the traceability chain is that reference materials must be *commutable*, which means that they will behave similarly to authentic clinical specimens across  $\geq 2$  measurement procedures. Lack of commutability has 2 major causes: matrix effects, as discussed above, and nonnative analytes being used to prepare the reference material (13). The consequences of noncommutability, in this case due to a matrix effect on cholesterol measurement caused by lyophilization of human serum, are shown in **Figure 2** (14). If the lyophilized reference material were to be used as a calibrator without understanding the noncommutability, the cholesterol measurements would be biased by  $\approx 10\%$ . Similarly, if a lyophilized material were used as an external quality assessment or a trueness control without considering noncommutability, false conclusions as to a clinical measurement procedure’s accuracy would be made. This situation actually happened for cholesterol in the early 1990s when many kinetic, cholesterol oxidase–based measurement procedures were falsely classified as providing significantly biased results. Similarly, reference materials for folate have been prepared by



**FIGURE 1.** A traceability chain for serum cholesterol. By using this scheme, a routine measurement of serum cholesterol in a clinical laboratory has a calibration that is ultimately tied to a solid material of high and known purity prepared by the National Institute of Standards and Technology (NIST). Reference materials are shown on the left of the diagram and measurement procedures on the right. Each reference material is used to calibrate the next measurement procedure down the chain (right-pointing arrows), and each measurement procedure is used to assign values to the next reference material (left-pointing arrows). Each step has some associated uncertainty, which ideally can be specified, and so the overall measurement uncertainty increases as one proceeds down the chain but can be estimated for the routine clinical result. For many more complex substances, including folate and vitamin B-12, a certified solid reference material is not currently available. Folate and vitamin B-12 measurements can, however, now be made traceable to serum-based NIST reference materials analogous to SRM (standard reference material) 1951b for cholesterol. SI, International System of Units; ID-GC/MS, isotope dilution-gas chromatography/mass spectrometry; CDC, Centers for Disease Control and Prevention.



**FIGURE 2.** A cholesterol-oxidase-based clinical laboratory measurement procedure compared with the Abell-Kendall (ALBK) reference measurement procedure measured cholesterol concentrations using fresh human serum (●) and lyophilized human serum (▲), which in the past has often been used as a reference material for both calibration purposes and as a “trueness control” in proficiency surveys to evaluate a clinical measurement procedure’s accuracy. Reprinted with permission from reference 14.

supplementing a human serum pool with folic acid, whereas the main form of the vitamin normally in serum is 5-methyltetrahydrofolic acid (5-MTHF). This material could give very misleading results in comparing different clinical laboratory folate measurement procedures, depending on their specificity for different forms of the vitamin. The central importance of commutability of reference or proficiency testing materials in evaluating the accuracy of clinical laboratory testing cannot be overemphasized (13).

### ISSUES WITH COMPLEX MEASURANDS

As the complexity of the substance being measured increases, it becomes more difficult to establish metrological traceability. Proteins and other biomarkers can have many variant molecular forms in the circulation with differing biological activities. It may be known which forms are important clinically, and it may not be possible to prepare pure standards. Principles of traceability are still applicable in a more limited fashion (4, 8). *Harmonization* refers to the procedures to improve agreement among various laboratory methods when rigorous traceability to a primary standard cannot be achieved.

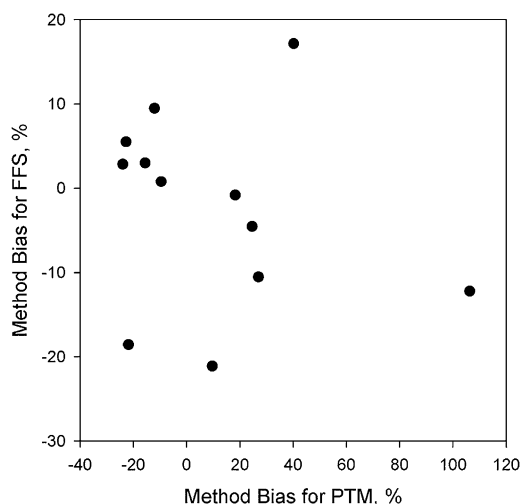
### MARKERS OF FOLATE AND VITAMIN B-12 STATUS

Folate and vitamin B-12, although simpler molecules than proteins, have characteristics that make standardization difficult. Each is a fairly complex molecule with multiple forms (“vitamers”) present in clinical samples, foods, food supplements, and reference materials. Their concentrations in serum are low, especially for vitamin B-12, and they both bind tightly to proteins in the circulation. For both substances, a microbiological measurement procedure has been the historical standard for defining accuracy, whereas competitive binding assays are now the predominant methods in clinical laboratories. These facts create

serious problems in defining the measurand and in establishing traceability, because the former is a functional measurement and the latter is a substance quantity measurement. Pure reference materials are not available. Another complicating factor is that red blood cells as well as serum and plasma have been a common matrix for folate assays.

The good news with respect to serum folate assay is that the important vitamers are well-characterized substances, and several stable-isotope-dilution mass spectrometry measurement procedures (for 5-MTHF, 5-formyltetrahydrofolic acid, and folic acid in serum) are now listed as reference methods in the JCTLM database (1, 15). Using these methodologies the NIST has assigned folate vitamers and total folate concentrations to a serum-based reference material, SRM 1955 (16), which has been shown to be commutable with human serum samples (17). The NIST certifies this material for several years, monitors its stability, and replaces it as necessary. For vitamin B-12, such methodology is more challenging because of the lower concentrations. Investigators have made progress in harmonization using WHO International Reference Reagent 03/178, with a vitamin B-12 value assigned by a consensus approach (18). The NIST is currently working on establishing a primary method, based on either isotope dilution mass spectrometry or neutron activation analysis, and has created the material for a candidate SRM, designated SRM 3951 (3, 19). Unfortunately, despite all the work in preparing standard reference materials and high-level reference measurement procedures, to our knowledge little has been done to investigate the commutability of this vitamin B-12 SRM with clinical measurement procedures, and without such commutability data its use as a calibrator or trueness control can be risky in our view.

A study that was part of the College of American Pathologists (CAP) Surveys Program shows some of the challenges in standardizing measurements of folate and vitamin B-12 (20). External quality assessment schemes, also called proficiency testing programs, such as the CAP Surveys, endeavor to assess the accuracy of clinical laboratory testing by sending identical specimens to many laboratories and comparing their results. Because the surveys require a large volume of specimen, and must test a range of possible measurement results, it is quite difficult and extremely expensive to distribute actual pooled human serum or plasma on a regular basis. Instead, traditional proficiency testing material (PTM) is generally made from animal plasma or outdated human plasma from blood banks, often with addition of preservatives, stabilizers, and other additives. The materials are also often supplemented with various analytes, possibly including nonnative forms of analyte. Hence, the commutability of PTM is generally very questionable. A 2003 CAP survey event, however, included a specimen from a fresh-frozen human serum (FFS) pool along with traditional PTM specimens. Investigators compared the results for 3 common measurement procedures related to anemia: serum ferritin, serum folate, and serum vitamin B-12. The dispersion of results for ferritin with FFS and PTM was essentially similar (the overall CV across methods was  $\approx 12\%$  in each case), and individual method biases for FFS and PTM were similar. For folate, however, the overall CV was 18% with FFS and twice as high (37%) with PTM. Furthermore, individual method biases between FFS and PTM did not correlate with one another (**Figure 3**). For vitamin B-12, CVs were similar for PTM and FFS (12% and 13%, respectively), and the individual method biases correlated only modestly ( $r = 0.55$ ,  $P = 0.05$ ; data not shown). Thus, the study showed that this PTM was not a commutable reference material for either vitamin B-12 or folate,



**FIGURE 3.** Method bias plot for folate. The scatterplot compares method bias for folate assays on 2 different materials: an artificial Proficiency Testing Material (PTM; *see* text) and a fresh-frozen human serum pool (FFS). Each point represents a commercial folate methodology used by  $\geq 10$  laboratories participating in a 2003 proficiency testing event of the College of American Pathologists. The method bias equals the mean result for the method minus the mean result for all methods. The method biases for the 2 materials showed no significant correlation ( $r = -0.224$ ,  $P = 0.48$ ), indicating that the PTM is not commutable with human serum for routine folate measurement procedures. Reprinted with permission from reference 20.

particularly the latter. The study did not investigate the cause of the noncommutability, which could relate to matrix effects, analyte factors, or both, as discussed above.

## CONCLUSIONS

In summary, scientists in the field of metrology have established consensus documents that define an approach to establishing consistent, accurate clinical laboratory test results for purposes of patient care and epidemiologic studies. The complexity of folate and vitamin B-12 make it difficult to achieve full realization of some of these principles. However, the modern metrological approach and new, sophisticated mass spectrometric techniques have improved the standardization of serum folate and vitamin B-12 assays, with expected benefits for both epidemiologic studies and patient care.

The authors' responsibilities were as follows—JLB: drafted the manuscript; and JLB and JHE: contributed to concept development, reviewed the manuscript, and had responsibility for the final content. Neither of the authors has a personal or financial conflict of interest.

## REFERENCES

1. Yetley EA, Pfeiffer CM, Phinney KW, et al. Biomarkers of folate status in NHANES: a roundtable summary. *Am J Clin Nutr* 2011;94(suppl):■■■■.
2. Yetley EA, Johnson CL. Folate and vitamin B-12 biomarkers in NHANES: history of their measurement and use. *Am J Clin Nutr* 2011;94(suppl):■■■■.
3. Yetley EA, Pfeiffer CM, Phinney KW, et al. Biomarkers of vitamin B-12 status in NHANES: a roundtable summary. *Am J Clin Nutr* 2011;94(suppl):■■■■.
4. Vesper HW, Thienpont LM. Traceability in laboratory medicine. *Clin Chem* 2009;55:1067–75.
5. Yetley EA, Pfeiffer CM, Schleicher RL, et al. NHANES monitoring of serum 25-hydroxyvitamin D: a roundtable summary. *J Nutr* 2010;140:2030S–45S.
6. De Bièvre P. The 2007 International Vocabulary of Metrology (VIM), JCGM 200:2008 [ISO/IEC Guide 99]: Meeting the need for intercontinentally understood concepts and their associated intercontinentally agreed terms. *Clin Biochem* 2009;42:246–8.
7. Joint Committee for Guides in Metrology (JCGM). International vocabulary of metrology: basic and general concepts and associated terms (VIM). Sèvres, France: JCGM, 2008. Available from: <http://www.bipm.org/en/publications/guides/vim.html> (cited 11 January 2011).
8. International Organization for Standardization (ISO). International Standard ISO 17511. In vitro diagnostic medical devices—measurement of quantities in biological samples—metrological traceability of values assigned to calibrators and control materials. Geneva, Switzerland: ISO, 2003.
9. Joint Committee for Guides in Metrology (JCGM). Evaluation of measurement data: guide to the expression of uncertainty in measurement. Sèvres, France: JCGM, 2008. Available from: <http://www.bipm.org/en/publications/guides/gum.html> (cited 11 January 2011).
10. Clinical and Laboratory Standards Institute (CLSI). Characterization and qualification of commutable reference materials for laboratory medicine: approved guideline. CLSI document C53-A. Wayne, PA: CLSI, 2010.
11. Clinical and Laboratory Standards Institute (CLSI). Interference testing in clinical chemistry: approved guideline. 2nd ed. CLSI document EP7-A2. Wayne, PA: CLSI, 2005.
12. Vesper HW, Miller WG, Myers GL. Reference materials and commutability. *Clin Biochem Rev* 2007;28:139–47.
13. Miller WG, Myers GL, Rej R. Why commutability matters. *Clin Chem* 2006;52:553–4.
14. Eckfeldt JH, Copeland KR. Accuracy verification and identification of matrix effects. The College of American Pathologists' Protocol. *Arch Pathol Lab Med* 1993;117:381–6.
15. Pfeiffer CM, Fazili Z, McCoy L, Zhang M, Gunter EW. Determination of folate vitamers in human serum by stable-isotope-dilution tandem mass spectrometry and comparison with radioassay and microbiologic assay. *Clin Chem* 2004;50:423–32.
16. Satterfield MB, Sniegowski LT, Sharpless KE, et al. Development of a new standard reference material: SRM 1955 (homocysteine and folate in human serum). *Anal Bioanal Chem* 2006;385:612–22.
17. Ihara H, Watanabe T, Hashizume N, et al. Commutability of National Institute of Standards and Technology standard reference material 1955 homocysteine and folate in frozen human serum for total folate with automated assays. *Ann Clin Biochem* 2010;47:541–8.
18. Thorpe SJ. The development and role of international biological reference materials in the diagnosis of anaemia. *Biologicals* 2010;38:449–58.
19. National Institute of Standards and Technology. Available from: <http://www.nist.gov/mml/analytical/organic/vitsb6andb12inserum.cfm> (cited 11 January 2011).
20. Bock JL, Endres DB, Elin RJ, Wang E, Rosenzweig B, Klee GG. Comparison of fresh frozen serum to traditional proficiency testing material in a College of American Pathologists survey for ferritin, folate, and vitamin B12. *Arch Pathol Lab Med* 2005;129:323–7.