Results of an international round robin for serum and whole-blood folate

ELAINE W. GUNTER,1* BARBARA A. BOWMAN,1 SAMUEL P. CAUDILL,1 DELLA B. TWITE,1 MYRON J. ADAMS,2 and ERIC J. SAMPSON1

Because of the increasing significance of folate nutriture to public health, a "round robin" interlaboratory comparison study was conducted to assess differences among methods. Twenty research laboratories participated in a 3-day analysis of six serum and six whole-blood pools. Overall means, SDs, and CVs derived from these results were compared within and across method types. Results reported for serum and whole-blood folate demonstrated overall CVs of 27.6% and 35.7%, respectively, across pools and two- to ninefold differences in concentrations between methods, with the greatest variation occurring at critical low folate concentrations. Although results for serum pools were less variable than those for whole-blood pools, substantial intermethod variation still occurred. The overall results underscore the urgent need for developing and validating reference methods for serum and whole-blood folate and for properly characterized reference materials. For evaluating study or clinical data, method-specific reference ranges (established with clinical confirmation of values for truly folate-deficient individuals) must be used.

INDEXING TERMS: intermethod comparison • radioassay • chromatography • microbiological assay • ion capture • chemiluminescence

In addition to its requirement as an essential nutrient for preventing megaloblastic anemia and its role as a cofactor in one-carbon transfers required for DNA replication, the water-soluble vitamin folate has assumed even greater epidemiological importance in recent years because of its relation to the prevention of neural tube defects [1, 2] and to homocysteine metabolism in prevention of cardiovascular disease [3, 4]. Serum or plasma folate concentrations are considered to reflect recent dietary intake, whereas erythrocyte folate concentrations are indicative of body stores [5].

Typical study designs rely on retrospective evaluation of banked specimens for folate analyses in serum or erythrocytes, necessitating the comparison of biochemical folate results between several large-scale studies (and introducing another variable, long-term stability of folate in stored specimens). However, results generated by various methods may not be comparable, as can be discerned by examining external proficiency testing survey results, such as those for serum folate from the College of American Pathologists (CAP) [6]. Such surveys, however, generally include only serum samples and do not address variability among analytical methodologies for erythrocyte folate. The UK National External Quality Assessment Scheme (NEQAS) is unusual in having an erythrocyte folate testing material but uses expired blood-bank materials rather than materials prepared as typical clinical specimens (personal communication, M. J. Lewis, NEQAS hematology coordinator, October 1995). Comparability of folate assays among European laboratories was an objective of the European Community FLAIR Concerted Action No. 10 Folate Intercomparison Study 1992–1993 [7]. Specimens in this study also were not typical clinical specimens but rather aliquots of Bio-Rad Labs. (Hercules, CA) Lyphochek lyophilized whole-blood control pools.

Additional major problems are the lack of accepted Reference or Definitive Methods for analysis in serum or whole blood, as well as Standard Reference Materials in these matrices. The UK National Institute of Biologics, Standard, and Controls (NIBSC), in response to requests of the International Committee for Standardization in Haematology (ICSH) Folate and Vitamin B12 Committee, is currently undertaking the preparation of lyophilized whole-blood pools as a tentative reference material; target values, however, are being determined as a

3 At the March 1996 meeting of the Ligand Survey Committee of CAP, however, the Committee agreed to consider the implementation of several erythrocyte folate challenges in 1997 (personal communication, David Witte, chairman).

---

Divisions of 1 Environmental Health Laboratory Sciences and 2 Birth Defects and Developmental Disabilities, National Center for Environmental Health, Centers for Disease Control and Prevention, Public Health Service, US Department of Health and Human Services, Atlanta, GA 30341-3724.

*Author for correspondence. Fax 770-488-4609; e-mail ewg1@chehsl1.em.cdc.gov.

Use of trade names is for the purpose of identification only and does not constitute endorsement by the Public Health Service or the US Department of Health and Human Services.

Received January 19, 1996; accepted May 16, 1996.
consensus of different methods (Susan Thorpe, NIBSC, personal communication).4

Serum and erythrocyte folate data from the Third National Health and Nutrition Examination Survey (NHANES III, 1988–1994) will be released in 1996 (Wright J, Gunter EW, Johnson CL, et al., ms. in preparation). These data will be useful for establishing US normative ranges and for assessing the prevalence of inadequate folate status in the US population. Because the folate analyses from NHANES III were performed with the Bio-Rad QuantaPhase II® radioassay (RA) kit [8, 9], which was produced in response to the need to correct the calibration of the original QuantaPhase kit [10], it is important to be able to compare these data with those generated by other types of analyses [11]. Currently, the major analytical approaches are microbiological assay, RA, chemiluminescence, magnetic separation, HPLC, and ion capture. A few laboratories use GC-MS for clinical research. Except for microbiological assay, HPLC, and GC-MS, most methods are based on competitive binding by a specific folate-binding protein.

Materials and Methods

To assess the potential variability among methods, the Centers for Disease Control and Prevention (CDC) invited clinicians and laboratorians actively involved in folate research, as well as manufacturers of frequently used assay kits, to participate in a round-robin interlaboratory comparison study. To prepare a series of serum and whole-blood pools with concentrations spanning the ranges of values seen in healthy as well as deficient subjects, we screened serum and EDTA-anticoagulated blood specimens collected by the Emory University Hospital Blood Collection Services under an agreement with CDC (including an omnibus informed consent and Human Subjects Review protocol). The screening analyses were performed in the NHANES laboratory at CDC by the Bio-Rad QuantaPhase II RA kit, used exactly as it was for NHANES III, with serum quality-control pools at four concentrations (“anemic,” low, medium, and high folate concentrations) and three additional concentrations of whole blood pools, all of which had been in use for several years.

SPECIMENS

Of the 20 normal, noninfectious, apparently healthy adult men and women donors screened, 12 were selected as having appropriate serum or whole blood folate concentrations. No effort was made to add external pteroylglutamic acid (PGA) or 5-methyltetrahydrofolate (5-MTHF) to the pools or to charcoal-treat any of the collected materials to contrive folate-deficient samples. We also did not specifically request or exclude donors who were taking folic acid supplements. Because multivitamin supplement use is prevalent in the US population, serum/plasma concentrations exceeding 45–68 nmol/L (20–30 ng/mL) are frequently observed clinically. Our aim was to prepare pools that would resemble samples typical of those analyzed in most clinical laboratories.

Blood was collected in 250-mL bottles by phlebotomy from the fasting donors as either EDTA-anticoagulated or nonanticoagulated whole blood. The nonanticoagulated blood was allowed to stand for 2 h, covered with aluminum foil, at ambient temperature to allow maximum serum yield. We determined hematocrits of the screening specimens to ensure that no potential donor was anemic. After serum was separated from the clotted blood by centrifugation, 1 mg of l-ascorbic acid (Sigma Chemical Co., St. Louis, MO) was added per milliliter of serum to enhance the stability of the folic acid during storage. Whole-blood materials were allowed to stand at least 90 min after collection at ambient temperature to allow maximum hydrolysis of the polyglutamyl folates by endogenous folate conjugase. Whole-blood pools were diluted 10-fold, i.e., 50 mL of whole blood with 450 mL of 10 g/L ascorbic acid diluent. This combination of ascorbic acid concentration and the 1:10 dilution to prepare the hemolysate was chosen because of its traditional use in microbiological folate methods [12]. Users of all other methods (e.g., RAs with a 1:11 initial dilution) were instructed to correct their calculations accordingly.

All pools were dispensed in 1.0-mL aliquots into 2.0-mL high-density polypropylene cryovials (Nalgé, Rochester, NY) labeled with appropriate pool identifiers (e.g., SFOL-01, RBCF-04). The aliquots were promptly frozen at −70 °C until shipment. Extra aliquots were retained by CDC. Each laboratory was sent a shipment on dry ice containing six vials of each of the 12 pools and instructed to perform two determinations per vial on two vials per day for 3 days, for a total of 12 measurements per pool. No effort was made to ensure that the participating laboratories did not know the identity of the pools; rather, the objective was to assess the degree of comparability among leading clinical research laboratories. Laboratorians were instructed to report their actual results and key information describing their assay method such as calibration material, throughput, and expected reference ranges. Although we had originally measured the hematocrit of each donor pool, participating laboratorians were requested to report results as whole-blood folate, in nanograms per milliliter.

Participants and Methods

Of the 20 participating laboratories, 11 were in the US and 9 were international. The participants included two manufacturers, five government laboratories, six academic laboratories, and seven clinical research facilities.

Eight laboratories used the Lactobacillus casei microbiological assay, some following the traditional assay set-up in culture tubes and others using microtiter plates and readers. Eight laboratories used RAs: Bio-Rad QuantaPhase II; Simultrac S® and Simultrac S No-Boil® (ICN/Becton Dickinson, Orangeburg, NY); and Dual Count Solid Phase No-Boil® (Diagnostic Products Corp., Los Angeles, CA). Two laboratories used chemiluminescence methods: Corning Magic-Lite® and ACS: 180® (Corning, E. Walpole, MA). Two newly developed
assays were also used, ion-capture IMx® (Abbott Labs., Abbott Park, IL) [13] and HPLC with fluorescence detection (unpublished observations, CM Pfeiffer and JF Gregory, 1995), by one laboratory each.

For calibration materials, folic acid was used in two of the microbiological assays, 5-MTHF in one chemiluminescence assay and the HPLC assay, and PGA in the remaining assays.

**STATISTICS**

For each laboratory, the mean, SD, and CV were computed for each pool. No attempt was made to separate out among-day, among-vial, or pure within-run analytical variance when computing SD and CV. To obtain an all-laboratory mean, SD, and CV, we first computed the mean and SD of the laboratory means for each pool. Results from any laboratory with a mean lower than this preliminary all-laboratory mean, minus three times the corresponding all-laboratory SD, were then excluded, and a final all-laboratory mean, SD, and CV were computed. Although the HPLC laboratory provided data from two separate runs, only one set of results was used so that accuracy and precision would be determined on the same basis as for the other laboratories.

**Results and Discussion**

Individual serum and whole-blood folate results reported by each laboratory, as well as the overall results, are presented in Tables 1 and 2. Results are grouped by similar method and are reported in nmol/L (1 nmol/L = 2.265 ng/mL). Laboratories 6, 7, and 8 are grouped to facilitate comparison because they all used the same version of the L. casei assay. Similarly, laboratories 9–12 used the Bio-Rad RA, and 13 and 14 used the Becton Dickinson boil RA.

For pool SFOL-01, two laboratories reported only ">45 nmol/L" and not a diluted specimen result for the highest-concentration sample; thus, their results for this sample were not included in the final calculations. Laboratory 14 reported serum results only, and laboratory 2 reported whole blood results only.

Mid-range folate concentrations appeared to have the best agreement. Variation at low and high concentrations tended to be greater than mid-range concentrations, as seen for high whole blood pool RBCF-01, low whole blood pool RBCF-02, high serum pool SFOL-04, and low serum pool SFOL-06.

Unlike proficiency-testing challenges, which are essentially a one-time snapshot of method performance, the objective of this round robin was to present performance averaged over 3 days.
with the average of 12 replicate measurements for each of the 12 pools. The variation among laboratories was especially striking for the whole-blood pools, which varied from threefold (pool RBCF-06) to ninefold (pool RBCF-01). Results for serum pools varied from twofold (pool SFOL-03) to ninefold (pool SFOL-06). Within pools, accuracy also differed markedly, as seen in the extreme results reported for pool RBCF-01.

As stated earlier, the higher-concentration pools were not achieved by adding purified PGA or 5-MTHF, but instead reflect naturally occurring folate exposure. Clearly, some characteristic of pool RBCF-01 affected the HPLC method quantification more than did any of the other five whole-blood pools. When results for this pool are excluded, the HPLC assay results (laboratory 20) for pools RBCF-02 though RBCF-06 are very similar to those for the Bio-Rad RA laboratories (labs 9–12). However, the serum results for laboratory 20 (HPLC) are only about half of those obtained by all other methods. This HPLC method is specific for 5-MTHF, which constitutes ~90% of circulating plasma folate, but the somewhat mixed results may indicate a possible loss of folate during extraction.

Similarly, the results for pool RBCF-01 by laboratory 16, which used the Diagnostic Products assay, were extremely high (917 nmol/L), but results for all other whole-blood pools by this method agreed more closely with the results by laboratory 13, which used the Becton Dickinson boil RA. Serum results from the Diagnostic Products assay, however, were more comparable with the ACS:180 results (laboratory 17) than with those from the Becton Dickinson assays.

Within-laboratory precision also varied considerably within and across methods, as seen in the summary entries for Tables 1 and 2. The best overall precision results, as reflected by the smallest CVs, were those of laboratories 6 (M), 12 (B-R), 13 (B-D), and 19 (IC). The ion-capture (IMx) assay is the newest addition to the folate assay methodology and, as can occur with many new products, the results in the hands of laboratorians using the newly adopted method may initially tend to be somewhat less precise. Overall, the microbiological assay results tended to have higher CVs than the RA group.

Traditionally, RA has been used clinically because of its relative ease and greater throughput compared with the L. casei assay. The RAs are calibrated, as are most of the microbiological assays, with PGA and take advantage of the fact that at pH 9.2 PGA and 5-MTHF have equal affinity for the folate-binding protein. The best microbiological vs RA comparison can be drawn from laboratories 6–8 (all of which used the same version of the L. casei assay) and laboratories 9–12 (all of which used the Bio-Rad RA). Within groups, each assay compares well with itself, but the two assays are not comparable across the range of serum and whole-blood concentrations examined in the survey. The Bio-Rad RA was initially marketed with calibrator concen-
trations to match the performance of the *L. casei* assay, but in response to questions raised by Levine [10], the QuantaPhase II assay was introduced in late 1993 with spectrophotometrically verified PGA calibrator concentrations, resulting in a 30% shift downward in measured folate concentrations. Except for pool RBCF-01, results for the other five erythrocyte pools by this RA are ~50% lower, on average, than those from laboratories 6–8 (all using the same *L. casei* assay). Serum results for the Bio-Rad RA and laboratories 6–8, however, agree much more closely.

The results for the lower-concentration serum pools, SFOL-02 and SFOL-06, were especially interesting. Too often, clinical diagnoses or population assessments are made from single measurements of serum or plasma folate. Such measurements are a problem, particularly because normal subjects and patients with definite deficiency may have a substantial overlap in values; accordingly, test results often include "borderline" or "indeterminate" ranges [14]. Clearly, it would be inappropriate to use one set of reference ranges for all methods. Examples of the variation in normal ranges cited by the some of the participating laboratorians, or suggested by the manufacturer of kits used, vary widely and are compared in Table 3.

In the FLAIR Folate Intercomparison, which preceded the recalibration of the Bio-Rad QuantaPhase RA, the authors concluded that most RAs tended to overestimate serum folate content because standards were improperly calibrated and PGA rather than 5-MTHF was used as the calibration material [7]. In our study, however, the RAs as a class clearly did not overestimate the serum folate. The chemiluminescence assays and the Diagnostic Products RA had the highest overall values for serum folate and demonstrated substantial bias with the lowest-concentration serum pools, SFOL-02 and SFOL-06.

Using three different calibrator source materials did not have a clear-cut effect but may have added to the overall variability. According to the manufacturer (personal communication, ICN/ Becton Dickinson, December 1995), the chemistries for the Abbott Magic-Lite (using PGA calibrators) and ACS:180 (using 5-MTHF) are similar; however, we found that, although they gave comparable results for the serum pools, their results for the erythrocyte pools were quite different. The manufacturers' suggested reference ranges for these assays vary slightly as well, which may be partially due to the different populations surveyed for this purpose or may reflect true assay differences. Laboratories 1 and 5 used folinic acid as calibrator material, but these laboratories are not necessarily more comparable with each other than with other laboratories using the *L. casei* assay.

Lacking a "gold standard" method or material, a definition of which results are "correct" is impossible, in that all of these methods have inherent errors. We examined weighted linear regressions of each laboratory's mean results against the means for the same method type, separately for the serum and whole-blood pools. For the eight microbiological assays, the slopes varied by 0.81–1.22 for the serum pools and by 0.77–1.32 for the whole-blood pools. For the eight RAs (collapsed together for greater statistical power), the respective slope variances were 0.81–1.28 and 0.75–1.15, and for the two chemiluminescence assays they were 0.91, 1.10 and 0.71, 1.29. (Results for the ion-capture and HPLC assays are not presented because they represent only one laboratory per method type.)

### Table 3. Variation in reference ranges used by some participating laboratories or suggested by manufacturers of assay kits.

<table>
<thead>
<tr>
<th>Lab/Co.*</th>
<th>Folate range</th>
<th>In serum, nmol/L</th>
<th>In erythrocytes, nmol/L RBC&lt;sup&gt;*&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-D S</td>
<td>&quot;Low&quot;</td>
<td>&lt;3.6</td>
<td>&lt;283</td>
</tr>
<tr>
<td></td>
<td>&quot;Indeterminate&quot;</td>
<td>3.6–5.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Normal&quot;</td>
<td>&gt;5.0</td>
<td>≥283</td>
</tr>
<tr>
<td></td>
<td>&quot;Low&quot;</td>
<td>&lt;3.4</td>
<td>≥272</td>
</tr>
<tr>
<td>Lab 15</td>
<td>&quot;Normal&quot;</td>
<td>3.4–38.3</td>
<td>≥272–1948</td>
</tr>
<tr>
<td>(B-D S-NB)</td>
<td>&quot;Low&quot;</td>
<td>&lt;227</td>
<td></td>
</tr>
<tr>
<td>DPC</td>
<td>&quot;Indeterminate&quot;</td>
<td>6.8–38.5</td>
<td>227–453</td>
</tr>
<tr>
<td></td>
<td>&quot;Normal&quot;</td>
<td>6.8–38.5</td>
<td>396–1586</td>
</tr>
<tr>
<td>ACS</td>
<td>&quot;Normal&quot;</td>
<td>5.9–39.2</td>
<td>328–1110</td>
</tr>
<tr>
<td>M-L</td>
<td>&quot;Normal&quot;</td>
<td>&gt;8.2</td>
<td>272–974</td>
</tr>
<tr>
<td>B-R</td>
<td>&quot;Deficient&quot;</td>
<td>&lt;3.4</td>
<td>&lt;136</td>
</tr>
<tr>
<td></td>
<td>&quot;Normal&quot;</td>
<td>3.4–46.7</td>
<td>215–1291</td>
</tr>
<tr>
<td>IC</td>
<td>&quot;Deficient&quot;</td>
<td>&lt;6.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;Normal&quot;</td>
<td>7.0–28.0</td>
<td>337–1461</td>
</tr>
<tr>
<td>M-Lab 1</td>
<td>&quot;Normal&quot;</td>
<td>&gt;6.8</td>
<td>&gt;317</td>
</tr>
<tr>
<td>M-Lab 2</td>
<td>&quot;Normal&quot;</td>
<td>5.00–50.0</td>
<td>500–1300</td>
</tr>
<tr>
<td>M-Labs 6–8</td>
<td>&quot;Deficient&quot;</td>
<td>&lt;4.5</td>
<td>&lt;227</td>
</tr>
<tr>
<td></td>
<td>&quot;Possible deficiency&quot;</td>
<td>4.5–6.1</td>
<td>227–340</td>
</tr>
<tr>
<td></td>
<td>&quot;Probable normal&quot;</td>
<td>&gt;6.1</td>
<td>&gt;340</td>
</tr>
<tr>
<td>M-Lab 5</td>
<td>&quot;Normal&quot;</td>
<td>4.5–22.7</td>
<td>317–770</td>
</tr>
</tbody>
</table>

<sup>*</sup> Codes as in Table 1.

<sup>*</sup> Includes correction for hematocrit in final calculation: NOT whole-blood folate as used in round robin.
In summary, our comparisons of serum and whole-blood folate concentrations measured in 20 research and clinical laboratories by 7 different types of assays showed considerable intra- and intermethod variation. These results demonstrate that folate concentrations measured in one laboratory cannot be reliably compared with those assayed in another laboratory without an evaluation of interlaboratory differences. Hence, comparing folate data for different study populations obtained by use of different analytical methods is difficult. Excellent discussions of method comparisons and possible sources of variability have been previously presented by Shane et al. [15] and Brown et al. [16]. The reasons they cite for potential methodological differences are still valid but have become even more complex with the emergence of new analytical technologies. Using method-specific reference ranges in which "deficient" status results are clinically confirmed by other hematological markers is essential. Developing Definitive Methods and Certified Reference Materials is urgently needed for serum and erythrocyte folate analysis. In 1994, at the Sixth Conference for Federally Funded Human Nutrition Research Centers in Bethesda, MD, a strong argument was presented for a collaborative development of these methods and materials (Wayne Wolf, personal communication). Recent epidemiological evidence on the increasing importance of folate underscores the need for analytical accuracy in its measurement. Improved standardization is needed before researchers can collectively contribute to the evaluation of how well various folate cutoff values predict folate-preventable conditions such as megaloblastic anemia, hyperhomocysteinemia, and neural tube defects. Joint efforts toward this goal by the National Institute for Standards and Technology, CDC, US Department of Agriculture, National Committee for Clinical Laboratory Standards, Institute for Food Research/UK, NIBSC, and ICSH should be rigorously supported.

This work was supported in part by the National Center for Environmental Health and by the National Center for Health Statistics, CDC. We thank the following laboratories for their participation and support: Lynn Bailey, Jesse Gregory, and C.M. Pfeiffer, Department of Food Science and Human Nutrition, University of Florida, Gainesville, FL; Selwyn DeSouza, Bio-Rad Laboratories, Hercules, CA; Phillip Garry and Patricia Stauber, School of Medicine, University of New Mexico, Albuquerque, NM; Ralph Green, Michael Miller, and Lynn Mantufler, Cleveland Clinic Foundation, Cleveland, OH; Paul Finglas, Institute of Food Research, Norwich, UK; Victor Herbert and Spencer Shaw, Bronx VA Medical Center, Bronx, NY; Sean O'Brien and Brian Kelleher, St. James' Hospital, Dublin, Ireland; Robert Jacob, USDA Human Nutrition Research Center, The Presidio, San Francisco, CA; John Lindenbaum, Columbia University Medical Center, New York, NY; Klaus Pietrzik, Department of Pathophysiology of Human Nutrition, University of Bonn, Bonn, Germany; Jacob Selhub, Marie Nadeau, and Irwin Rosenberg, USDA Human Nutrition Research Center at Tufts University, Boston, MA; John Scott, Zana Kelley, and Eileen Hallinan, Department of Biochemistry, Trinity College, Dublin, Ireland; Nefertiti Sourial, St. Bartholomew's Hospital, London, UK; Tsunenobu Tamura and Howerde Sauberlich, Nutrition Sciences Department, University of Alabama–Birmingham, Birmingham, AL; Chris Thomas and Regine Steegers–Theunissen, University Hospital St. Radboud, Nijmegen, The Netherlands; Johan Ubbink, University of Pretoria, Pretoria, South Africa; and David Wilson, Abbott Laboratories, Abbott Park, IL.

We also thank the staff of the NHANES Laboratory, CDC, for their contributions in preparing and shipping these materials to the participating laboratories.

References