Primary Standardization of Assays for Anticonvulsant Drugs: Comparison of Accuracy and Precision

John F. Wilson,1* Ian D. Watson,2 John Williams,1 Pat A. Toseland,3 Alison H. Thomson,4 Grace Sweeney,5 Brian L. Smith,6 Lance N. Sandle,7 John D. Ramsey,8 and Nigel E. Capps9

Background: The accuracy and precision of methods for the measurement of the anticonvulsants phenytoin, phenobarbital, primidone, carbamazepine, ethosuximide, and valproate in human serum were assessed in 297 laboratories that were participants in the United Kingdom National External Quality Assessment Scheme (UKNEQAS).

Methods: We distributed lyophilized, serum-based materials containing low, medium, and high weighed-in concentrations of the drugs. The 297 participating laboratories received the materials on two occasions, 7 months apart. Expected concentrations were determined by gas chromatography or HPLC methods in five laboratories using serum-based NIST reference materials as calibrators.

Results: In general, bias was consistent across concentrations for a method but often differed in magnitude for different drugs. Bias ranged from −1.9% to 8.6% for phenytoin, −2.7% to 3.1% for phenobarbital, −2.7% to 0.5% for primidone, −8.6% to 0.3% for carbamazepine, −5.6% to 2.0% for ethosuximide, and −7.2% to 0.1% for valproate. Intralaboratory sources of imprecision significantly exceeded interlaboratory sources for many drug/method combinations. The mean CVs for intra- and interlaboratory errors for the different drugs were 6.3–7.8% and 3.3–4.2%, respectively.

Conclusions: For these long-established and relatively high-concentration analytes, the closed analytical platforms generally performed no better than open systems or chromatography, where use of calibrators prepared in house predominated. To improve the accuracy of measurements, work is required principally by the manufacturers of immunoassays to ensure minimal calibration error and to eliminate batch-to-batch variability of reagents. Individual laboratories should concentrate on minimizing dispensing errors.

© 2002 American Association for Clinical Chemistry

To ensure that laboratory measurements of drug concentrations in serum are of the highest accuracy, they should be traceable through an unbroken chain of comparisons of stated uncertainties to a national or international standard (1,2). In the anticonvulsant field, serum-based standard reference materials are available from NIST (Gaithersburg, MD). The cost of these materials is prohibitive for their use in routine analytical laboratories. A traceable link to such standards is possible, however, through participation by laboratories in an external quality assur-

1 Department of Pharmacology, Therapeutics & Toxicology, University of Wales College of Medicine, Cardiff CF14 4XN, United Kingdom.
2 Department of Clinical Biochemistry, University Hospital Aintree, Liverpool L9 7AL, United Kingdom.
3 18 London Road, Burgess Hill RH15 8QX, United Kingdom.
4 Pharmacy Department and Department of Medicine & Therapeutics, Western Infirmary, North Glasgow Hospitals University NHS Trust, Glasgow G11 6NT, United Kingdom.
5 Department of Bacteriology, Southern General University Hospitals NHS Trust, Glasgow G51 4TF, United Kingdom.
6 Department of Pathology, Maudsley Hospital, London SE5 8AZ, United Kingdom.
7 Department of Chemical Pathology, Trafford General Hospital, Manchester M41 5SL, United Kingdom.
8 TICTAC Communications Ltd., St. George’s Hospital Medical School, London SW17 0RE, United Kingdom.
9 Department of Pathology, Princess Royal Hospital NHS Trust, Telford TF6 5FF, United Kingdom.

All authors are members of the Steering Committee for the United Kingdom National External Quality Assessment Scheme for Drug Assays.

*Author for correspondence. Fax 44-29-2074-8316; e-mail wilsonjf@cardiff.ac.uk.

Received April 19, 2002; accepted August 14, 2002.
ance (EQA)/proficiency testing program (3). The value of a primary standard is first transferred to secondary material by a small group of reference laboratories. The secondary material is subsequently distributed by the EQA program to its members. We report such an exercise undertaken by the United Kingdom National External Quality Assessment Scheme (UKNEQAS) for Therapeutic Drug Assays, currently in use in ~28 countries, that provides a robust comparison of the accuracy and precision of current analytical techniques available for the routine clinical measurement of anticonvulsants.

Materials and Methods

Primary Standards

Two reference standards from NIST were used. SRM900 and SRM1599 are drug mixtures at three concentrations in a processed human serum base. Serum blanks are provided with both sets of standards. The certified drug concentrations and their uncertainties are listed in Table 1.

Secondary Standards

Three secondary standards, prepared by Cardiff Bioanalytical Services Ltd. (CBSL; Cardiff, United Kingdom), contained the same six analytes as the primary standards in low, medium, and high concentration mixtures. Lyophilized samples (2-mL) were prepared by adding weighed-in concentrations of drugs to 1.7 L of human serum (Scipac Ltd.). Blank samples containing no added drug were similarly prepared for analysis. The CV of dispensing by weighing was 0.09%.

Transfer of Primary Values to Secondary Standards

Five reference laboratories that used chromatographic methods were selected from among UKNEQAS participants. These laboratories were volunteers from the eight participants with the smallest root-mean-square difference in laboratory measurements vs their method subgroup consensus mean for the six study analytes in 48 routine EQA samples distributed during 1999.

After the main laboratory survey, the five reference laboratories were provided with frozen aliquots of the primary reference materials, which had been reconstituted by CBSL by addition of a weighed quantity of water. The reference laboratories measured all six drugs simultaneously in both primary and secondary materials by either HPLC or gas chromatographic methods. All laboratories used assay calibrators prepared in house.

RESULTS

Table 1. Drug concentrations and their specified uncertainties in certified standard reference materials from NIST.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenytoin</td>
<td>4.2 ± 0.1</td>
<td>16.7 ± 0.3</td>
<td>60.7 ± 0.9</td>
</tr>
<tr>
<td>Phenobarbital</td>
<td>5.3 ± 0.2</td>
<td>21.6 ± 0.2</td>
<td>103.6 ± 0.3</td>
</tr>
<tr>
<td>Primidone</td>
<td>3.6 ± 0.1</td>
<td>8.1 ± 0.2</td>
<td>18.6 ± 0.7</td>
</tr>
<tr>
<td>Ethosuximide</td>
<td>11.8 ± 0.4</td>
<td>75.9 ± 0.5</td>
<td>174.7 ± 0.6</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>2.9 ± 0.1</td>
<td>8.8 ± 0.1</td>
<td>19.4 ± 0.2</td>
</tr>
<tr>
<td>Valproate</td>
<td>14.5 ± 0.2</td>
<td>69.1 ± 0.7</td>
<td>142.5 ± 0.8</td>
</tr>
</tbody>
</table>

Laboratory Survey

The three secondary standard samples were distributed for measurement to UKNEQAS participants in August 1999 and again in February 2000. Laboratories reported the measured drug concentration, the analytical method, and the type of calibrator used for calculating each measurement.

Results

Reference Laboratory Data

The five reference laboratories provided duplicate or single measurements for all six drugs in both primary and secondary materials. The single measurements were for three different drugs from single laboratories. HPLC was used exclusively for phenytoin, phenobarbital, and carbamazepine, whereas gas chromatography was used alternatively for primidone, ethosuximide, and valproate by one, two, and three laboratories, respectively.

All reference laboratories reported zero or not detectable drug concentrations in the blank samples supplied with the primary standards and in blank samples prepared by CBSL. On average, the reference laboratories underestimated the primary values by 2.4–5.2%. Two laboratories reported significantly low values compared with the reference values (P < 0.05, t-test) for four or five analytes, two other laboratories reported significantly low values for one analyte each, and the fifth laboratory reported significantly high values for one analyte. However, simultaneous analysis of both primary and secondary standard data expressed as percentage differences from the sample means showed no significant interaction between a laboratory and standard set (P > 0.05, ANOVA). Thus, the pattern of between-laboratory differences in bias was comparable for both primary and secondary standards, allowing an unbiased transfer of the primary standard values to the secondary material.

Reference values for each drug were assigned to the secondary material by correcting the measured secondary standard values by the mean percentage bias in primary standard measurements calculated across all three concentrations of the primary standard. The percentage transformation eliminated significant between-sample
differences in bias (P = 0.81, two-way ANOVA). The use of a common concentration-independent correction was superior to separate corrections at the three drug concentrations. It reduced by a factor of 12 the variance of differences between assigned secondary reference and weighed-in values for the secondary reference samples (P >0.05, ANOVA). The drug concentrations assigned to the secondary standards and their precision are listed in Table 2.

EQA LABORATORY SURVEY DATA

Drug measurements by 14 different methods were reported for some or all analytes by a total of 297 laboratories (60% in the United Kingdom, 30% in Europe, 10% outside Europe). Users of commercial immunoassays used the calibrators supplied by the manufacturers almost exclusively. Of 35 laboratories reporting measurements by HPLC, 26 prepared their own calibrators and 9 used calibrators from 8 different external suppliers.

Data selection. Data for each sample were first screened to reject outliers >3 SD from the sample mean by the robust method of Healy (4). Data were then selected for a drug where a laboratory had reported nonrejected measurements on both occasions for a sample by the same method and where there was a total of >20 measurements for the method. The numbers of measurements and those rejected or not meeting the selection criteria are displayed in Table 3. The most common reasons for excluding measurements were that the laboratory reported on only one of the two separate distributions of samples or because they used different analytical methods on the two occasions.

Accuracy. Drug measurements expressed as the percentage difference from the reference value were analyzed separately for each analyte by three-way ANOVA between method, drug concentration, and sample distribution. Significant differences between methods and concentrations were identified by the Student–Newman–Keuls test (P <0.05) and significant differences from the reference value by t-test (P <0.05).

No deterioration in the secondary standard was apparent over the 7 months of the study, there being no significant differences between sample distributions (P >0.05).

Significant between-method differences in accuracy were detected for all drugs. The mean (SE) percentage bias at the three different drug concentrations are displayed in Fig. 1, ordered by mean overall method bias. Methods not significantly different are indicated by solid lines at the bottom of each panel. Sixty-three percent of all drug/method combinations differed significantly from the reference value. The methods differing significantly are indicated in Fig. 1 by a star after the method name. Mean values for a method ranged from –8.6% (carbamazepine/Vitros assay) to +8.6% (phenytoin/Bayer turbidimetric assay) with a grand mean of –0.8% across the whole study. HPLC, when undertaken by the general population of scheme participants, significantly underestimated concentrations of four drugs by 2.3–2.8%.

Significant concentration-related differences in accuracy within a method were detected in 34% of cases (P <0.05). The Abbott TDx differed for phenytoin, phenobarbital, primidone, and ethosuximide, and the AxSYM analyzer differed for phenytoin, phenobarbital, and carbamazepine. The Vitros and the Bayer turbidimetric assays differed for phenytoin, and the CEDIA and Beckman turbidimetric assays differed for both phenobarbital and valproate. The Syva Emit differed for primidone.

Precision. The intra- and interlaboratory components of within-method variance were estimated by fitting a two-
way ANOVA between laboratory and concentration to the data for each method with the two sample distributions as replicates (5). Because the analysis was undertaken separately for each method, the variances calculated excluded differences in accuracy among methods identified in the previous section and were not additive to produce total variance. Significant differences ($P < 0.05$) between intra- and interlaboratory variances were identified by the Bartlett test followed by the Levy test (6).

Intralaboratory sources of variation exceeded interlaboratory sources in all cases except one (Fig. 2). The difference reached significance in 44% of drug/method combinations. These cases are indicated in Fig. 2 with a star after the method name. The mean intra-/interlaboratory variance ratio was 1.8. Between-method differences were significant for three drugs. The only comparisons that achieved statistical significance involved methods with variances at the extreme ends of the range. For...
phenytoin, the smallest intralaboratory variance, for the Abbott AxSYM, differed from the largest values, seen for the Bayer turbidimetric, Syva Emit, and Vitros methods. The two smallest interlaboratory variances for phenytoin, for the Abbott AxSYM and Bayer turbidimetric methods, differed significantly from up to five methods with the largest values. For carbamazepine, the intralaboratory variance of the Syva Emit was significantly greater than that for the two Abbott methods, whereas the interlaboratory variance of the CEDIA method was significantly greater than that for the Abbott AxSYM, Roche fluorescence polarization immunoassay (FPIA), and HPLC. For valproate, the only differences were in interlaboratory variance. The largest variance for the Beckman turbidimetric method differed from those for HPLC and the Abbott methods.

**Discussion**

Significant differences were detected in the accuracy and precision of assays for anticonvulsant drugs. In interpreting the data, it must be noted that the study is based entirely on measurements in lyophilized samples. Potential problems in the commutability of samples require consideration, and one cannot be sure that identical differences will occur when analyzing patient samples (7). However, analysis of mixtures of anticonvulsants is an issue that has to be addressed by manufacturers of assays used in clinical practice because polypharmacy is a
rational treatment option (8). Insignificant differences in method bias and precision have been identified between liquid and lyophilized samples (9), and because the biochemistry of the reconstituted lyophilized material from CBSL was not grossly abnormal, we therefore expect many of the observed differences to occur in patient materials.

There were significant differences in the magnitude of the bias at different drug concentrations for approximately one-third of drug/method combinations. In some cases, the pattern was of increased bias in the low-concentration sample, whereas in others, bias changed sign at different concentrations. These effects imply some nonlinearity in performance of the immunoassays concerned with respect to drug concentration and would prevent simple correction for bias across concentrations.

There were significant differences in accuracy among methods for all drugs. No one method performed consistently across all analytes. For example, the popular Abbott TDx FPIA gave high results for phenytoin but low results for carbamazepine and valproate. The chromatographic methods were generally members of the statistical group that was closest to the reference values. Significant differences from the reference values were nevertheless produced by chromatographic methods. These differences can be attributed to heterogeneity in the calibrators used by the laboratories performing chromatographic assays. Nevertheless, chromatographic methods were able to outperform many immunoassays in terms of accuracy, despite many laboratories choosing to prepare calibrators in house.

Within-method imprecision was partitioned into intra- and interlaboratory sources. Almost universally, intra-laboratory errors exceeded those from interlaboratory sources, which is in agreement with observations from the College of American Pathologists drug survey (10). The magnitude of intralaboratory imprecision was generally comparable among methods, although the Abbott AxSYM analyzer was shown to be statistically superior to some other methods in two cases. Interlaboratory variation within methods varied more between methods than intralaboratory variance. It was not notably less for analyzer/reagent groups from a single supplier compared with open analyzer systems or with chromatography, which might have been predicted to be more heterogeneous.

The scale of both inaccuracy and imprecision quantified by this study was surprising because it represents the current “state of the art” for a series of long-established assays working at relatively high micromolar concentration. Worse may be expected for drugs that are monitored in the nanomolar range or for which a commercial calibration material is not available. In comparison with a range of analytical goals for drug assays (9), the performance of these methods will not necessarily meet clinical requirements. The closed analytical platforms were not superior to open systems or to chromatography, where preparation of in-house calibrators predominated. It is therefore recommended that there should be attempts to improve performance by all concerned. Work is required principally from manufacturers of the available immunoassays to ensure minimal calibration error. That the scale of intralaboratory errors generally exceeds interlaboratory errors might seem to suggest that work by individual laboratories will be worthwhile either on internal quality control or to improve the accuracy of dispensing procedures. However, comparison of intralaboratory errors from simultaneous assay of replicates with those calculated from replicates measured 4 months apart (10) has demonstrated that long-term variance is more than twice that over the short term. In the present study, intralaboratory errors were those occurring over a 7-month period.

Factors that may contribute to this variability will include changes in batches of reagents and calibrators supplied by the reagent manufacturers to a laboratory, and drift and maintenance of instruments. It is possible, therefore, that a significant proportion of intralaboratory variance may be attributable to factors outside the control of the laboratory (10).

Laboratories have a responsibility to achieve the best consensus accuracy and minimum imprecision. Where they can influence such matters this should be done. However, our study indicates that their potential actions are limited and that variances are probably affected by changes in services received from suppliers. There is a need for the suppliers of drug assay products to consider their whole system variance over time and to take steps to minimize this. Setting agreed targets with users may be one way of deriving professionally agreeable targets. EQA schemes can stimulate these improvements and provide a cost-effective means to monitor progress by tracing results from routine laboratories to international standards.

We gratefully acknowledge provision of the NIST certified reference materials by the IFCC Scientific Division, the contributions of the members of the Joint Therapeutic Drug Monitoring Committee of the International Association of Therapeutic Drug Monitoring and Clinical Toxicology and the IFCC (J. Potter, J-J. Himberg, M. Wenk, and A.D. Fraser), provision of the secondary standard material by CBSL, and the analytical work undertaken by the five reference laboratories. The reference laboratories, in alphabetical order of contact person, were as follows: Dr. D. Berry (London, United Kingdom), Dr. P.M. Edelbroek (Heemstede, The Netherlands), Dr. H. Etter (Zurich, Switzerland), Dr. U. Juergens (Bielefeld, Germany), and Dr. A.H. Veefkind (Amersfoort, The Netherlands).

References
2. ISO/IEC 17025. General requirements for the competence of


