Quantitative Toxicology: Interlaboratory and Intermethod Evaluation in New York State

Steven N. Buhl, Patricia Kowalski, and Raymond E. Vanderlinde

The New York State Department of Health has conducted a proficiency evaluation program in quantitative toxicology since 1974. Serum samples containing a barbiturate and phenytoin, together with either glutethimide, procainamide, or theophylline, are sent to participating laboratories quarterly. Within the first two years of the program the percentage of laboratories able to quantitate 75% of the test samples to within 25% of the gravimetric values increased from 25 (1974–1975) to 40% (1975–1976). This improvement was partly due to licensure requirements, improved technology for sample preparation and analysis, and the availability of better quality-control practices. An obstacle to obtaining uniform accuracy is the lack of adequate calibration or testing materials. To overcome these obstacles, pure drugs are weighed into a bovine serum matrix, and the weights are confirmed by reference laboratories and used as the target values in the testing program. Comparison of the methods used by participants in this program for barbiturate and phenytoin yielded equations different from those found in other method evaluations.

Measurement of drugs and their metabolites in a patient’s serum is often the best approach to monitoring the patient’s physiologic response, compliance with prescriptions, or exposure to environmental hazards (1). Because an unexpected pharmacologic response (due to individual differences in absorption or metabolism of a given drug) or an overly toxic response to a presumed safe concentration of a drug can have serious medical and legal ramifications, accurate measurement of concentrations in serum is increasingly important (1).

Since 1974, the New York State Department of Health has conducted a state-mandated proficiency-evaluation program in quantitative toxicology, with strict requirements for analysis for barbiturates and other drugs. Twenty-five percent (12/47) and 40% (21/53) of the laboratories tested in the first two years were issued permits.

Laboratories that do not satisfactorily meet the requirements of this program are not licensed to perform analyses for toxic substances on samples from patients in New York State (outside New York City).

There are numerous studies comparing various methods of assaying drugs (2–14), but only few reports on laboratory proficiency (15–19); this is the first report on intermethod and interlaboratory performance of a large number of laboratories that incorporate such assays in their routine workload. We shall describe the basis, operation, and results of the New York State proficiency-evaluation program in quantitative toxicology; the problems encountered in preparing test samples; and some observations on currently used methods for measuring barbiturate, phenytoin, procainamide, theophylline, and glutethimide.

Materials and Methods

Drugs. Phenobarbital and secobarbital were purchased from J. H. Walker and Co., Mt. Vernon, N. Y., 10550. Glutethimide was purchased from CIBA, Summit, N. J. 07901; theophylline from Sigma Chemical Co., St. Louis, Mo. 63178; procainamide from Sterling-Winthrop Research Institute, Rensselaer, N. Y. 12144; and sodium salicylate from Allied Scientific, State College, Pa. 16801. Phenytoin (Dilantin, Parke-Davis) was obtained through USP-NF Reference Standards, Rockville, Md. 20852. Analysis of the drugs by gas chromatography–mass spectrometry revealed no contaminants.

Sample preparation and dispersement. Various amounts of a barbiturate, phenytoin, and one other drug were weighed and dissolved individually and then incorporated into 0.5 liter of bovine or human serum. The barbiturates and glutethimide were dissolved in methanol (<0.3 ml), phenytoin in sodium hydroxide (<0.5 ml, 1 mol/liter), and theophylline and procainamide-HCl in water. Sodium salicylate (added directly
to the serum) was included in about 30% of the samples from April 1974 to May 1976, but its analysis was not required.

The serum had been frozen immediately after withdrawal and preparation. For use it was thawed and Millipore filtered four times (prefilter, 1.2-, 0.8-, and 0.45-μm filters). It was stirred while the drugs were added and thereafter was kept only in a glass container. The pH of the serum was checked and, if necessary, adjusted to pH 7.5 before it was dispensed into 5-ml aliquots with a glass and Tygon-tubing dispenser (imprecision of bottling <0.01%). These were frozen, lyophilized, capped, and stored at 4 °C. Human serum was used in May and September 1974; bovine serum has been used since. The matrix was changed because the latter is more abundant and less hazardous to work with during sample preparation.

Four sets of test samples were sent to each laboratory during each testing year, which begins July 1. One sample from each of three or four lots prepared for each test set was sent to every laboratory in the program and simultaneously to one or more reference laboratories. After the laboratories reported their results, the target values were established. We do not know whether the reported values from any given laboratory represent single determinations or averages of multiple analyses.

**Definition of proficiency limits.** The limits for accuracy of each determination before July 1, 1976, were defined as the target value ±25%; after this date the limits were the target value ±30%. The target value was the gravimetric value of the drug after correction for the change in volume (8%) caused by lyophilization and reconstitution. As a safeguard against anomalies in sample preparation, when the measured values determined by both the reference and the participant laboratories did not agree with the gravimetric values, the target value was defined as the mean of all reference and participant laboratory values excluding outliers (<0.5 or >2 times the mean). If the mean of the participant laboratories' values, the gravimetric value, and the mean of the reference laboratories' values all disagreed, the gravimetric value was taken as the target value.

**Results**

Since the New York State quantitative toxicology proficiency program began, there has been a gradual change in the type of method used for barbiturates (Table 1). The immunologic EMIT (Enzyme Multiplied Immunoassay Technique; Syva Corp., Palo Alto, Calif. 94304) method is essentially replacing colorimetry and ultraviolet spectrophotometry. Radioimmunoassay was used by a single laboratory, for only one year. Use of gas chromatography has been generally stable. A similar change from colorimetric to immunologic methods has occurred for phenytoin.

**Barbiturates.** In all methods used to measure phenobarbital the amount of barbiturate was underestimated, as shown by the regression slopes of <1 (Table 2). Correlation with target values was especially poor for the colorimetric methods, as indicated by the low correlation coefficients.

The final discriminant for the efficacy of each method is the percentage of laboratory results that are within the defined limits of accuracy (Table 2). The poor proficiency rate for the colorimetric methods has led to their abandonment. We considered the possibility that this low rate was attributable to the concentrations being extended above 40 mg/liter. However, the participant laboratories' proficiency on the samples above 40 mg/liter was 80% for immunologic methods, 79% for gas chromatography, and 66% for ultraviolet spectrophotometry, compared to only 10% for colorimetry. Hence proficiency does not seem to have been affected for high amounts of drug in the sample. Neither glutethimide nor salicylate affected the accuracy or precision of the reported values.

Secobarbital was included instead of phenobarbital in one of three samples used during one test period. It was quantitated as accurately as phenobarbital by gas chromatography, ultraviolet spectrophotometry, and colorimetry. No immunologic assays were reported.

Some blanks containing no barbiturate were also included for barbiturate analysis. False-positive values were obtained for the ultraviolet spectrophotometric (18.2 ± 16.8 mg/liter; \( n = 4 \)) and colorimetric (13 ± 1.05 mg/liter; \( n = 3 \)) methods. No false positives were found by gas chromatography, and these samples were not tested by immunologic methods.

**Phenytoin (diphenylhydantoin).** Each method used to measure phenytoin underestimated the concentration (slopes <1, Table 2). Correlation with target values was poorer for the colorimetric methods than for either the gas-chromatographic or immunologic methods. The gas-chromatographic methods were the most nearly accurate and the colorimetric methods the least accurate. From 1974 to 1977 the percentage of laboratories using colorimetric methods has declined from 25% to

| Table 1. Percentage of Laboratories Using Various Methods of Measuring Barbiturates (from April Report of Each Test Year) |
|---|---|---|---|---|
| Year (no. of labs.) | UV spectrophotometric | Colorimetric | Immunologic | Gas-chromatographic |
| 1974 (41) | 37 | 29 | 0 | 34 |
| 1975 (38) | 37 | 21 | 0 | 42 |
| 1976 (36) | 22 | 3 | 22 | 55 |
| 1977 (39) | 10 | 3 | 0 | 44 |

* One laboratory used both gas chromatography and EMIT.
0%. Salicylates did not affect the accuracy or precision of the reported values.

Blanks assayed for phenytoin yielded false-positive results by gas chromatography (5.5 ± 12.8 mg/liter, n = 16) and by colorimetry (0.1 mg/liter, n = 2). No false positives were reported by laboratories using ultraviolet spectrophotometry.

**Procaenamide.** The results for procaenamide showed good agreement and good correlation with target values for all methods except ultraviolet spectrophotometry (Table 2). However, the number of ultraviolet spectrophotometric reports was small, and the results included some unusually discordant values. The colorimetric methods were the most nearly accurate and the gas-chromatographic and ultraviolet spectrophotometric methods the least accurate. Salicylate did not affect the accuracy or precision of the reported values.

**Glutethimide.** Glutethimide was included in 18 samples and used in testing six times (six lots with three samples each). Results for it showed a discrepancy between the gravimetric and the reference and participant laboratories' values in three samples each from two of the six lots prepared. The gravimetric value was higher than the measured value for five of these samples and lower than the measured values for one. The former findings suggest incomplete dissolution of the drug or interaction of the matrix with the drug. Errors in weighing could account for values being either higher or lower than the gravimetric values. The target values used for these samples are the means of all reported values, excluding outliers.

Gas-chromatographic methods generally underestimated glutethimide (slope <1); however, they had the highest percentage (86%) of acceptable results. Although the ultraviolet spectrophotometric methods had a slope of about 1, only 67% of the results were acceptable. Of the colorimetric results, only 50% were acceptable. Salicylate did not affect the accuracy or precision of the reported values.

**Theophylline.** Theophylline was included in four samples mailed to each laboratory in the January 1977 mailing. The average values for each of the four samples were higher for ultraviolet spectrophotometry (19, 28, 47, and 86 mg/liter; 10 laboratories) than by high-performance liquid chromatography (17, 27, 44, and 83 mg/liter; six laboratories). Because the results from both methods and those of the reference laboratories were lower than or equal to the gravimetric values (19, 33, 47, and 93 mg/liter), the target values were adjusted by the procedure given in Materials and Methods.

Both methods gave excellent results for theophylline in the complex matrix, which included large amounts of a barbiturate (up to 110 mg/liter), which can interfere with the ultraviolet spectrophotometric methods, and phenytoin (up to 68 mg/liter). All of the reported chromatographic values and 95% of the reported ultraviolet spectrophotometric results were within the accuracy limits.

Regression equations from several recent studies (2, 6-14) of methods are compared in Table 3. Although some of the equations are similar, the agreement in general is not good. Differences in technical competence or methodology may account for some disagreement. In another interlaboratory testing program, Pippenger et
Table 3. Comparison of Regression Equations from Evaluations of EMIT vs. Other Methods Used In Barbiturate and Phenytoin Analyses *

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Barbiturate</th>
<th>Phenytoin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>y = gas chromatography</td>
<td>y = UV spectrophotometry</td>
</tr>
<tr>
<td>This report</td>
<td>y = 0.87x - 2.0</td>
<td>y = 1.08x - 1.9</td>
</tr>
<tr>
<td>Spiehler et al. (2)</td>
<td>y = 0.90x + 0.79</td>
<td>y = 0.90x - 1.47</td>
</tr>
<tr>
<td>Schumann et al. (10)</td>
<td>y = 1.16x - 3.02</td>
<td></td>
</tr>
<tr>
<td>Booker and Darcey (6)</td>
<td>y = 1.26x - 2.72</td>
<td></td>
</tr>
<tr>
<td>Finley et al. (7, 9)</td>
<td>y = 1.04x - 2.42</td>
<td></td>
</tr>
<tr>
<td>Brunk et al. (8)</td>
<td>y = 1.39x - 0.94</td>
<td></td>
</tr>
<tr>
<td>Haven (11)</td>
<td>y = 1.22x - 5.515</td>
<td>y = 0.93x - 0.52</td>
</tr>
<tr>
<td>Legaz and Raisys (12)</td>
<td>y = 0.87x + 2.7</td>
<td>y = 0.97x + 2.3</td>
</tr>
<tr>
<td>Belfield et al. (13)</td>
<td>y = 0.91x + 4.6 b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>y = 0.86x + 19.7 c</td>
<td></td>
</tr>
<tr>
<td>Turri (14)</td>
<td>y = 0.76x + 0.26</td>
<td></td>
</tr>
</tbody>
</table>

* x = EMIT. b Simulated patient serum. c Patients' serum.

al. (15) found 88 combinations of extraction and gas-chromatographic methods being used for analysis of phenobarbital.

Discussion

Interlaboratory testing programs in quantitative toxicology have led to improved accuracy in the analysis of drugs. In the Antiepileptic Drug Proficiency Testing Program, laboratory proficiency improved from less than 50% in 1974 to 88% in 1977 (15, 16). In the College of American Pathologists toxicology program, CV's for analysis of both phenytoin and phenobarbital decreased (18). In the New York State program for quantitative toxicology, laboratory proficiency increased from 25 to 40% by the second year after initiation of a licensure program.

In any test program, occasional discrepancies are inevitable. In some test samples containing glutethimide and theophylline the gravimetric values did not agree with the measured values by either the reference laboratories or the participant laboratories. This problem of occasional nonagreement between the gravimetric and the measured values will continue until more is known about the interactions between the serum protein matrix or container and the added drugs (which were contaminant free) and the effect of these interactions on the various methods of measurement (1, 19, 20).

The results were analyzed according to general methods without regard to variations or modifications introduced by the participating laboratories. In another survey (16) more than 80 different gas-chromatographic methods were used. There are also numerous variations in colorimetric and ultraviolet spectrophotometric methods. Many of the variations are unique to particular laboratories.

Reported concentrations of barbiturate and phenytoin were generally low compared to the target values. This underestimation can result either from incomplete extraction of the drug in those methods requiring extraction or from interaction of the drugs with the serum components in such a manner that they cannot be extracted or measured immunologically (19-21). Low measurements by EMIT probably reflect the bovine matrix of our second-year samples (Syva Inc., personal communication). Bovine serum albumin is the protein conjugated with the drug used to generate the EMIT reagents; hence, nonproductive binding may have resulted in our samples with this matrix.

Although the immunologic methods do not quantitate or differentiate drugs as well as the gas-chromatographic methods, laboratories which use them can pass our quantitative test when the class or name of the drug is known.

Because patients' samples may contain metabolites of the drug(s) being measured and other concomitantly administered drugs and because about 10% of the overdose cases (22) ingest more than a single drug, our samples for future tests will include some complex mixtures of drugs, their metabolites, related compounds, other drugs that might reasonably be expected to be co-administered, as well as more blanks. Past experience indicates that these may pose problems for the ultraviolet spectrophotometric and fluorometric methods because of interfering compounds with overlapping absorption and emission spectra, and for the immunologic methods because of cross reactivities of drugs, metabolites, and related compounds (23, 24). Such problems were encountered in the past when glutethimide was included in samples with barbiturate, making quantitation impossible with either the radioimmunoassay or some colorimetric assays (24-26), and when theophylline and barbiturate were placed in the same sample, thereby creating potential interferences for the ultraviolet spectrophotometric analysis of each.

Although other studies have reported problems with gas chromatography of samples containing phenobarbital and glutethimide (16), we have not had any. High-performance liquid chromatography is convenient as well as specific for both quantitative and qualitative drug analysis (4, 5), and it has the potential for improving the accuracy of quantitative analysis.

In preparing our evaluations of methods, we used the
gravimetric values of drugs in the serum as the reference points. Such a comparison does not deal with the effects of various metabolites in clinical specimens of serum, nor is it established on a complete understanding of drug–matrix interaction. However, these defects are less disturbing than those encountered when comparing results from two analytical methods.

The differences in the regression equations (Table 3) found in the several evaluations probably result from lack of standardization of materials and methods in clinical toxicology. Any attempt at standardization must take into account several formidable problems: the many extraction methods, the propensity to modify published methods, and the disparity of results attributable to changes in personnel performing the tests.

The accuracy of any determination can be affected by numerous problems. Detectors on gas chromatographs have various sensitivities to particular drugs (27, 28), and different columns have application to specific drug analyses. The limitations of both must be considered. The use of external rather than internal standards, or the addition of the standard at the end rather than the beginning of the analysis, may result in incomplete extraction (27). Perhaps the best approach includes using several internal standards, incorporating them at the beginning of the extraction process, comparing detector responses for various drugs, and calculating the results from known detector responses and peak areas (27, 29).

Enzyme immunoassay is probably simpler to use than gas chromatography, but the potential for errors is just as great, particularly if the manufacturer's instructions are not followed exactly. The enzymatic response as a function of time is not constant for this technique. Hence, because the time and duration of the measuring interval varies among the spectrometers when determining the change in absorbance, it is possible that results will vary with different spectrometers (30).

Finding the amount of a drug when extraction is incomplete is very difficult in the ultraviolet spectrophotometric or colorimetric methods because internal standards are not generally used. Other problems in obtaining accurate results include nonspecificity of the colorimetric methods and introduction of short cuts in the ultraviolet spectrophotometric methods (such as using a single pH or absorbance reading for phenobarbital measurements).

Another problem shown by the low proficiency ratings up to 1976 may be that the manufacturers' reported values for commercially available standards (calibration materials) do not always agree with the measured value found in various reference laboratories for specific lots. For all these reasons, extreme caution should be used in interpreting and comparing evaluations of methods.

Although proficiency testing programs (external quality control) can lead to improvements in accuracy, improvement must begin with internal quality-control practices (16, 31–33). Several schemes for quality control have been published (16, 31, 33–35) or are marketed commercially. Their use as tools for improving the accuracy of results is to be encouraged.

We acknowledge the assistance of Drs. Peter Jatlow, C. W. Pippenger, Dennis P. Ritz, D. Tietelbaum, and M. K. Schwartz, who acted as our referees, and thank Mr. John Meola for constructive comments on this manuscript.

References


2. Spiehler, V., Sun, L., Miyada, D. S., et al., Radioimmunoassay, enzyme immunoassay, spectrophotometry, and gas–liquid chromato-
graphy compared for determination of phenobarbital and diphe

3. Evans, J. E., Simultaneous measurement of diphenylhydantoin and phenobarbital in serum by high-performance liquid chromatog-


5. Kabra, P. M., Gotelli, G., Stanhill, R., and Marton, L. I., Simul-
taneous measurement of phenobarbital, diphenylhydantoin, and

6. B. E. H., and Darcey, B. A., Enzymatic immunoassay vs. gas liquid chromatography for determination of phenobarbital and diphe

Chem. 22, 911 (1976).

8. Brunk, S. D., Hadjiioannou, T. P., Hadjiioannou, S. I., and Malm-


Pathol. 66, 823 (1976).

11. Haven, M., Phenyo tin and phenobarbital measurement by cen-

12. Legaz, M., and Raiays, V. A., Correlation of the "EMIT" antig
Biochem. 9, 35 (1976).


15. Pippenger, C. E., Penny, I. K., White, B. G., et al., Interlab-
oratory variability in determination of plasma antiepileptic drug concen-

Toxicol. 1, 118 (1977).


20. Ibid., p 526.


