Letters to the Editor

Proficiency Testing in Forensic Toxicology

Sir:

Competency, the final criterion for all professional scientific activity, is constantly sought but seldom measured, and as such is deservedly a perennial issue. It is of paramount importance within the forensic sciences, where serious individual and social consequences often hang in the balance. Following a recently published report [*1*] and a subsequent exchange of letters [*2*,*3*] concerned with an assessment of analytical competency in forensic toxicology, a host of specific and general issues relative to the conduct of proficiency testing has been brought into open forum. Although much of importance has already been stated in these publications, a number of underlying professional principles, both philosophical and scientific, are still in critical need of discussion by all forensic toxicologists if appropriate and realistic proficiency testing is ever to become an accepted, routine part of the profession.

Under a contract issued by the National Institute on Drug Abuse (NIDA) a brief analytical proficiency study was conducted in 1974 by Eugene C. Dinovo and Louis A. Gottschalk (Department of Psychiatry and Human Behavior, California College of Medicine, University of California at Irvine). This study enlisted the cooperation of staff at forensic toxicology laboratories in nine major U.S. city-county jurisdictions. During the test period each laboratory received five simulated biological specimens, accounting for a total of 14 different drugs variously distributed in three 25-ml urine specimens and two 10-ml albumin preparations. In addition to instructions that ranged from a request for total quantitative identification to a simple statement that the sample be analyzed "in a routine way," each individual test sample was accompanied by varying amounts of information regarding the drugs to be found in that sample. Three reference laboratories were used to "check the manufacture of these test specimens" and to determine that "no gross errors in preparation had occurred." Unfortunately, the published account of the study does not give exact details of how these samples were manufactured, "checked," stored, and shipped.

Following the analysis of their data, Dinovo and Gottschalk concluded that (1) there is significant interlaboratory variation in the ability of these forensic toxicologists to measure and detect drugs; (2) there are startling differences in analytical accuracy and precision; (3) accuracy and precision improve markedly when the analyst has prior knowledge of the drugs present; and (4) there exists a real potential for artificial differences in mortality statistics when based on work from these laboratories.

In general, the data on which these conclusions were based indicated that for the acidic drugs and the volatile compounds the results from the nine test laboratories were adequate (even when the samples were presented as "general unknowns") and that the greatest scatter in the reported values and the greatest frequency of inappropriate negative results were associated with the basic drugs and the benzodiazepines. Although mean values and standard deviations for the reported analytical results are given, there is little discussion by Dinovo and Gottschalk as to how the gap was bridged between the analytical data and their conclusions of "significance." A discussion is critically needed because an in-depth analysis of the study results by Kelly and Sunshine [*2*] showed that statistically the test laboratory values and the reference laboratory results together represented a different population of values than the reputed, weighed-in concentrations. This observation suggests poor manufacture, storage, or shipment of the test specimens rather than problems in the analytical techniques used, in which case the analytical results would have been statistically scattered around the "true" values. This point, unfortunately, was blatantly ignored by Dinovo and Gottschalk [*3*].

Although the data presented in the report on this proficiency study are of interest and

deserve attention, they are not conclusive and certainly cannot serve as a basis for the doomsday conclusions drawn. Variation in measurements are a fundamental and basic phenomenon of all quantitative biological values. The parameters tested (variance, accuracy, and precision) tell little more than that which anyone familiar with analytical toxicology might already suspect. Quantitative variance between two laboratories testing the same specimen, or even a series of measurements within the same laboratory, would likely give rise to several sets of values. The proper questions to be addressed through proficiency testing are not whether variance exists, but what is the source of the observed variance and to what degree is it significant, if at all.

If a reasonable approach is to be taken to these problems, it must first be ensured that sources of possible variation are confined to the analytical procedures used or, at the very least, that variance arising from the preparation, storage, and shipment of the test specimens can be accounted for quantitatively. The group conducting this proficiency study gives no indication that they are aware of either of these principles. In short, there is no way to be certain that the reported variance results solely from the techniques employed at the separate test laboratories. The use of reference facilities only to "check for gross errors" in the manufacture of the test specimen undermines the fundamental utility of the survey. Unstated, but clearly implied in the important criteria outlined by Kelly and Sunshine [2] justifying the need for reference laboratories, is the futility of determining variance in a series of test measurements without first establishing a reference variance. To conclude, in the absence of this fact, that variability is "significant" is absurd.

Accuracy, a second parameter measured in the proficiency test, can be defined as the degree to which a series of analytically determined values corresponds to some known, "true" value. In theory at least, it is a straightforward variable; however, in analytical human toxicology accuracy measurement is not easily defined. It can only be an assumption (but a critical one) that the weighed amount of drug added to a biological matrix is indeed present in the final concentration desired. There are no magic scales and pipets that can give an absolute value against which all other measurements can be compared, a fact that escaped Dinovo and Gottschalk entirely. Nowhere in their report do they describe standard analytical procedures used to establish these needed values.

Similar objections could be applied to their measurement of the third variable, precision; however, of the three it was most nearly assessed in its proper form. To measure the ability of the toxicologists at the test laboratories to reproduce a value in a series of analyses demands that value be constant within the test samples. Otherwise, the scatter of the data cannot be held as a true picture of the precision of the analytical methods employed. To judge precision by measurement of five different samples made up separately over a period of months, as implied in the report, is unacceptable. However, as with many other aspects of this proficiency study, essential details are omitted. There is no way of knowing whether the samples were made up individually, drawn from one pooled sample, or drawn from several pooled samples, either initially or over a period of time.

The assessment of the significance of error in any analytical activity is essential to the quality of performance and any subsequent improvement in that activity. It must be remembered that the determination of significance has little meaning outside the context of the final objective of the activity and responsibility of the analyst. Although the need to analyze biological samples with appropriate degrees of accuracy and precision is critical, this only partially defines the function of forensic toxicologists. The forensic toxicology laboratory is designed and intended to serve as an integral part of the scientific medicolegal investigation team. The primary responsibility of this team is to describe the circumstances and biomedical facts that contribute to an individual's demise. Seldom, if ever, does one part of this unit act independently or in complete isolation. Assessment of competency and proficiency in forensic toxicology must be developed from a consideration of practical as well as statistical determinations of significance.

It is practical significance that underlies and describes the real world in which forensic toxicologists must function. Unlike the neat mathematical formulas used to determine statistical significance, this broader variable combines many diverse and constantly changing aspects, ranging from scientific procedures to local politics. Whereas statistical significance relates to the "at the bench" measurable quantities in forensic toxicology, practical significance relates to all aspects, including the bench work; unlike its statistical brother, it always relates to the primary responsibility of forensic toxicology in medicolegal investigations. This does not deny the need and the duty of forensic toxicologists to strive for reductions in variability and increases in accuracy and precision. A hierarchy of importance does exist, however, between the statistical and the practical, with the latter serving as the true foundation of this analytical and public health science. State-of-the-art knowledge in physiology, toxicology, and pharmacology often denies the requirements for highly accurate measurements. For example, there could well be no practical significance in the forensic interpretation of a "true" urine-secobarbital concentration of 5 μg/ml and an experimentally determined concentration of 4 μg/ml. Statistical analysis would reveal an error of 20%, but it would be irresponsible for a forensic toxicologist to commit a large percentage of time, personnel, or expense to reduce this error, particularly at the sacrifice of other laboratory functions. To lose sight of practical significance in forensic science can rapidly threaten its purposes in spite of any scrupulous maintenance of tight analytical patterns. The importance of statistics as a tool in forensic toxicology should not be ignored, but the term "significant" must be an amalgam of the practical as well as the statistical. Forensic toxicology and analytical organic chemistry are not synonymous.

The apparent intention of NIDA to check the reliability of mortality statistics resulting from reports made by medical examiners and coroners is undeniably worthwhile. However, the careless fashion in which their contractors conducted the study and the lack of a firm scientific basis for their dogmatic conclusions cast doubt on Dinovo and Gottschalk's true purposes. In addition to the reported facts and conclusions, a close look at how the data were presented betrays prejudice and suggests an intent beyond their originally stated objectives. An illustration of their apparent bias is seen in their use (as opposed to their definition) of the term "false negative" in their original report [*1*]. The term as initially defined was intended to describe the situation in which a drug present in a spiked sample was not detected by the test laboratory. This broad definition is certainly legitimate, although it does not distinguish between the failure of a particular procedure designed to detect a given drug and the failure to detect a drug because it was not part of the analytical scheme employed. Later in the report, when referring to the results from specific samples, the authors continually state that "false negatives were reported." The implication, of course, is that these were all false negatives of the first type in which the drug was looked for, not found to be present, and then reported as such by the forensic toxicologist. This would have been possible only with the one test sample in which all the constituents had been identified to the laboratories prior to the analysis. Undoubtedly there were some failures in the screening procedures used and false negatives consistent with their original definition occurred, but for those specimens presented as general unknowns or with only drug class descriptions, the term "recorded" rather than "reported" would have been more accurate and certainly less prejudiced in its implications. But for the fact that it appears again and again, the phrase might be considered as a sin of poor usage rather than misdirected intentions.

Before we too are accused of trying to sweep the "awful truth" under the rug, it should be stated that the Dinovo and Gottschalk survey certainly measured variability in something; further, there is no doubt that this variance is in part attributable to the analytical techniques used at the participant laboratories. Most forensic toxicologists are acutely aware that a continual struggle must be waged to maintain a high standard of analytical work and would be surprised by any finding that did not indicate some interlaboratory

analytical variability. The sad truth is, however, that the degree to which individual laboratory methods in this survey contributed to this variance is not apparent and any hope of recovering it is now lost. Although as a test of analytical competency the design of the study could rightly be considered a waste of time, the use of the data to conclude that there is "significant" doubt concerning the ability of the toxicologists at the test laboratories is an unjustified and very serious charge that should not go unchallenged.

There are no good arguments against the need and desirability of increasing and safeguarding the quality of forensic toxicology service to public health problems. Such a service, when managed efficiently, both at a practical and scientific level, not only augments the quality of medicolegal investigations but can, with certain qualifications, aid in the study of some aspects of drug use and misuse. Presumably this was Dinovo and Gottschalk's original intention. The real question raised by the report is not whether such safeguards and controls are needed, but who should design and oversee them. The answer to this question can also be inferred from the report; if the Dinovo and Gottschalk survey does nothing else it stresses by implication that programs designed to assess quality in forensic toxicology must be guided by those familiar with the discipline. To have less cannot but undermine the progress such programs should ensure.

For the past three years the Toxicology Section of the American Academy of Forensic Sciences has made fundamental advances towards clearly defining forensic toxicology in terms of professional qualification through a board certification program, careful evaluation of current analytical practices with a view to establishing "substantiated methods" appropriate to postmortem analytical toxicology, and finally mechanisms by which a realistic and meaningful proficiency testing program can be instituted. These three activities are integral parts of establishing and raising the standards of practice in forensic toxicology laboratories. It is essential that the Academy's efforts succeed if the profession is to be its own master. The board certification program is well established. The evaluation of methods is proceeding, but slowly, and proficiency testing has barely begun. The foregoing discussion illustrates how crucial it is that a proficiency testing program be started soon. It will be costly if done well and will undoubtedly require technical expertise and possibly management from outside the Academy Section, but this duty and responsibility must not be shirked if forensic toxicologists are to have their rightful professional claims recognized by the scientific community at large and the standards and objectives of medicolegal investigation advanced.

Kevin L. McCloskey, Ph.D.
Bryan S. Finkle, Ph.D.
Center for Human Toxicology
University of Utah
Salt Lake City, Utah 84112

References

[1] Dinovo, E. C. and Gottschalk, L. A., "Results of a Nine-Laboratory Survey of Forensic Toxicology Proficiency," *Clinical Chemistry,* Vol. 20, 1976, p. 846.
[2] Kelly, R. C. and Sunshine, J., "Proficiency Testing in Forensic Toxicology: Criteria for Experimental Design," *Clinical Chemistry,* Vol. 22, 1976, p. 1413.
[3] Dinovo, E. C. and Gottschalk, L. A., "More on Proficiency Testing in Forensic Toxicology," *Clinical Chemistry,* Vol. 22, 1976, p. 2056.

Discussion of "A Review of *Dental Evidence: A Handbook for Police*"

Sir:

I wish to make some comments in rebuttal to the unobjective, irresponsible, and mis-

leading review by Dr. L. J. Levine of my book, *Dental Evidence: A Handbook for Police,* which appeared in the October, 1976 issue of *Journal of Forensic Sciences.*

The foreword and the first chapter of the book clearly state that the purpose of *Dental Evidence* is *not* to present new cases, new research, or new methods or to impress its intended audience—police officers—with a flood of complicated dental terminology. Its purpose is to give to the police and to criminalists existing information in nontechnical language so that, after appropriate classroom and laboratory training by a dentist, designated evidence technicians or criminalists can do much of the preliminary stages of bite mark investigations that are now done by dentists. After all, similar procedures are being done every day in ordinary dental offices by dental assistants in states where this is permitted by law and are performed daily by police investigators, often using dental impression materials, in areas such as toolmark, tire print, or footprint evidence.

Irl Allen Gladfelter, Jr., D.D.S.
Consultant in forensic odontology
439 West Sixty-third St.
Kansas City, Mo. 64113