The National System of Scientific Measurement

J. S. Hunter

Summary. Mandated measurement methods are required by regulatory agencies and other government groups. These methods exist for measuring almost all physical, chemical, and biological phenomena. The methods have been culled from the literature, from the organizations that write voluntary standards, and some have been developed by the agencies. Few provide adequate estimates of precision, and fewer still provide any evaluation of interlaboratory bias. The societal costs of these poor measurements are large. Much needs to be done to meet the physical and statistical requirements for establishing and maintaining dependable measurements. Excepting those directly supported by the National Bureau of Standards, most of the nation's measurement systems are uncontrolled.

Measurements are an essential part of the communication between research scientist and development engineer, buyer and seller, regulator and regulated. Indeed, most measurements are taken for granted. The kilogram, meter, and liter are accepted with little question in Princeton or Berkeley, Peking or Cairo, as are units of time and temperature. In the United States the responsibility for this happy state of affairs rests with the National Bureau of Standards (NBS), which maintains programs to guarantee the usefulness of the International System of Units (the SI units) and of many other measures derived from these fundamental units. Traceability to the ultimate standards maintained by the NBS is conducted through a hierarchy of federal, state, and private laboratories employing working and field standards. The NBS provides state, county, and local officials with technical and operational guides that contain measurement specifications, standard tolerances, and model laws designed to support the measurement system. In many cases this is done in close collaboration with the National Conference on Weights and Measures, a forum for the exchange of measurement information staffed by the Office of Weights and Measures of the NBS.

The measurement system covering SI units and other closely allied measures can be declared to be in good health (*1*). Politically, few doubt the need for acceptable SI unit measurements, or object to paying the price necessary to ensure their validity. Statistically, these measurements may be declared "under control," with biases reduced to practical limits and with variances that are both stable and economically acceptable. The SI unit systems have performed so well and with so little fanfare that it may come as a surprise that many other measures in common use are lacking in dependability.

Today's Measurement Requirements

In recent years, our society has moved far beyond the need to maintain high-integrity measurement systems for everyday commercial and scientific use. The government has legislated a rapid growth of various regulatory agencies concerned with public health, safety, and the environment. Many of the laws enacted direct the establishment of measurement methods and protocols. For example, in the Clean Air Act of 1977 [Section 114(2)], we read:

> The Administrator may require the owner of any emission source to (A) establish and maintain such records, (B) make such reports, (C) install, use and maintain such monitoring equipment or methods, (D) sample such emissions (in accordance with such methods, at such locations, of such materials, and in such a manner as the Administrator shall prescribe. . . .)

The Federal Water Pollution Control Act contains a similar statement [Section 304(g)]. The Occupational Safety and Health Act [Section 6(b)7] directs that occupational health standards shall provide for monitoring and measuring employee exposure at such locations and intervals and in such a manner as may be necessary to protect employees. Most regulatory laws carry similar requirements.

There are thousands of government-mandated measurement methods. One reason is the number of pollutants that must be monitored. In addition to the criteria air pollutants (CO, SO_2, O_x, NO_x, and total suspended particulates), which can require different measurement protocols depending on whether the source is ambient, mobile, stationary, or a line source, there are aerosols, aldehydes, H_2S, and asbestos. In water there are approximately 200 toxic pollutants to be monitored, plus some 500 proposed hazardous substances—along with the usual measures of biological oxygen demand, turbidity, odor, and the common chemical elements. Of course, the many varieties of water (pretreated, treated, groundwater, sewage, seawater, and drinking water) can each require separate measurement methods. The Federal Insecticide, Fungicide, and Rodenticide Act requires that all pesticides be classified and registered and that information specifying their chemical and physical properties and degradability be produced. The Toxic Substances Control Act requires that new chemicals be screened and tested for toxicity before they are marketed. There are numerous mandated methods for measuring radioactive substances and for testing foodstuffs, drugs, and biological substances. We have reached the stage where there is a federally mandated method for measuring almost every physical, chemical, or biological phenomenon.

The NBS estimates that the taking of measurements of all kinds costs 6 percent of the gross national product. It was estimated that in 1977 the federal government alone spent $690 million on the collection of data, and that approximately 43 percent of the data was generated by or for the environmental agencies (*2*). If the direct cost of making measurements is large, the indirect cost of making poor measurements must be huge. The question, "How well are these measurement systems performing?" can be answered in part by giving a few examples.

J. S. Hunter is a statistician and a professor of civil engineering at Princeton University, Princeton, New Jersey 08544.

0036-8075/80/1121-0869$01.50/0

Dependability of Data

Under the Federal Water Pollution Control Act, a National Pollutant Discharge Elimination System (NPDES) permit is required of anyone placing a pollutant into the nation's waters. To obtain a permit one must submit a report on the amount and chemical composition of the pollutant. The permit, when granted, almost always requires the use of one or more mandated measurement methods; may prescribe the number, location, and types of samples; and requires periodic reports. Some 70,000 NPDES permits have been processed. Complying with the NPDES requirements has influenced almost every industrial, commercial, and governmental analytical laboratory.

Measurement of biochemical oxygen demand and suspended solids is required in almost all NPDES wastewater treatment plants. The measurement methods employed for both are well established. (*3*). A recent investigation of these methods, involving approximately 150 treatment plants in Wisconsin, revealed that (*4*)

> 80% of the major municipal, 72% of the minor municipal, 60% of the major commercial laboratories with sufficient data to study, and 100% of the minor commercial labs generated unacceptable data for BOD [biological oxygen demand]. Twenty percent of the major municipal, 35% of the minor municipal, 70% of the major commercial and 33% of the minor commercial laboratories produced unacceptable data for suspended solids.

An NPDES quality assurance study conducted in New Jersey in March 1979 noted that of 77 participating laboratories, four did not even bother to return data and only 21 performed acceptably. Twelve laboratories inspected demonstrated "serious nonconformity to approved test procedures" (*5*).

For many years the Center for Disease Control has run an interlaboratory program to maintain control over measurements of lead in blood. Approximately 100 laboratories participate each month. A histogram of the results, developed from the April 1980 record (*6*), appears in Fig. 1. The best estimate of the blood lead concentration in the distributed sample was 41 micrograms per deciliter. The average reported by all participating laboratories was 44 μg/dl. In an ordinary sample of human blood, the nominal concentration lies between 15 and 20 μg/dl; 30 μg/dl would be considered noteworthy. Clearly, whatever the true amount of lead in a sample, the variability demonstrated in Fig. 1 guarantees numerous false alarms, or—perhaps more important when the true level is high—nonalarms. Since the interlaboratory program has been going on for some time, Fig. 1 represents the best the measurement system can do given present resources.

An important device for detecting and controlling interlaboratory bias is the Youden plot (*7, 8*). Two samples, each containing an unknown amount of some substance, are sent to several laboratories. Each laboratory analyzes its pair of samples and the two results are later recorded on an x-y plot, each point representing a laboratory's data. Figure 2 illustrates the results obtained from 50 laboratories participating in an early pararosaniline study (*9*). The data fall along a reasonably straight line, some consistently biased high while others are consistently low. Points unusually distant from the line reflect analysis problems more serious than those posed by bias.

The examples given thus far have been of measured characteristics of the environment and public health. The problems illustrated are not unique to these areas; similar problems occur in the more structured sciences. The quarterly *Journal of Physical and Chemical Reference Data*, published jointly by the Office of Standard Reference Data of the NBS, the American Chemical Society, and the American Institute of Physics, provides many data displays illustrating the remarkable variability of published chemical and physical constants or properties. Even though these quantitative determinations were made by conscientious scientists and each determination appears in reviewed scientific literature, the between-laboratory biases are often great (*10*).

From these illustrations it is clear that many measurements, both those in everyday use and those arising from sophisticated laboratory investigations, should be viewed with skepticism.

Data of poor quality are a pollutant of clear thinking and rational decision-making. Scientific advancement suffers when one laboratory cannot duplicate the results of another. Biased data, and the relationships derived from such data, can have serious consequences in the writing of laws and regulations. Communications between the regulator and the regulated frequently break down over issues that derive directly from the measurement systems employed. And, unfortunately, a Gresham's law for data evolves whenever inaccurate data (frequently more cheaply acquired and more voluminous) overwhelm accurate data (*11*).

Physical Aspects of a Measurement Method

There appears to be little recognition of what is required to establish a measurement method or to maintain its associated measurement system (*12*). I begin therefore with a brief enumeration of some of the physical requirements that must be considered before a measurement method can be expected to produce useful information and follow with a similar listing of the statistical requirements, recognizing that the two are not necessarily mutually exclusive. In the following, η represents the true numerical value of some physical, chemical, or biological characteristic.

Physical characteristics of a measurement method include the following:

The metric. Should η be measured in parts per million or micrograms per cubic meter and should it be done in terms of η or of some transformation, such as $\ln \eta$?

The physical system of measurement. Is η to be measured directly or through some more readily measured surrogate response such as conductivity, color, or ion concentration?

The calibration curve. A calibration curve displays some chosen surrogate response, or instrument response, as a function of the true response. The construction of a calibration curve requires, in addition to the use of standard reference levels of η, carefully controlled and skillfully run laboratory trials.

The sample. The size of the sample, how it is secured, and its preparation must be prescribed.

Specificity. The surrogate response must be specific for η. For example, the release of iodine from potassium iodide is not specific to ozone but rather to the total oxidizing capacity of an air sample.

Interference. An instrument probe, or method for creating a surrogate response, can often disturb the physical characteristic being measured. Structural details of a sample may also influence the measured values.

Compatibility. Two different measurement methods are compatible when an x-y plot of their corresponding measures of various levels of η fall along a monotonic curve.

Traceability. Measurements should be directly related to national standards (*13*).

Complexity. Directions for reading a thermometer are simple. The description of the many steps necessary to determine the amount of SO_2 in an air sample requires 2.5 pages of closely packed description (*14*).

Statistical Aspects of a Measurement Method

Even when all the physical criteria of a measurement method are considered, repeated observations of the same level of η will produce different results. Thus a single observation y may be partitioned, in principle, into two components $y = \eta + \varepsilon$, where ε is an "error" or disturbance. The disturbance may have at least two components, one deterministic (a fixed bias), the other stochastic or random. They supply the reason for concern over the following statistical aspects of a measurement method.

Distribution of errors. If the metric for the recorded response has been well chosen, repeated observations on the same η should produce a reasonably symmetrical histogram.

Gaussian distribution. Random errors of observation frequently take a Gaussian form, and many statistics derived from observations have random aspects that are well characterized by this distribution. The distribution has two parameters, the mean η and the variance σ^2. In the absence of systematic bias, the mean η is considered to equal the true value of the characteristic being measured.

Bias. A fixed contribution to the measurement error is a bias.

Precision. The standard deviation σ is often used as an index of precision (or imprecision). One measure of precision is the reciprocal of the standard deviation, $1/\sigma$, or simple multiples such as $\pm 2\sigma$ or $\pm 3\sigma$.

Accuracy. Accuracy is a function of both precision and bias. Methods that give accurate measurements have good precision (small imprecision) and near zero bias. Methods that yield inaccurate data can have poor precision or unacceptable bias or both (*15*).

Coefficient of variation. The coefficient of variation is $\gamma = \sigma/\eta$, the standard deviation divided by the mean. It is a measure of variability made dimensionless and is often expressed as a percent, that is, as $\sigma/\eta \times 100$.

Sample. A sample may contain a single observation y or be a collection of n observations: y_i, $1 = 1, 2, \ldots, n$. Some aspect of randomization should attend the taking of observations comprising a sample; that is, it is essential to let the laws of chance enter upon taking the sample and recording the observation. Only then are the commonly used laws of probability, statistical estimation, and inference applicable.

Computed statistics. Using Roman letters for statistics and Greek letters for parameters, we note that the arithmetic average $\bar{y} = \Sigma y/n$, where n is the number of observations in the sample, is the statistic that best estimates η (in the absence of bias), and the statistic $s^2 = \Sigma(y - \bar{y})^2/(n - 1)$ best estimates σ^2, the variance of the observations. Parameters are fixed quantities, ideas, or mathematical concepts. Statistics, since they are formed from the data, are taken to be random variables, realizations, or numerical best guesses.

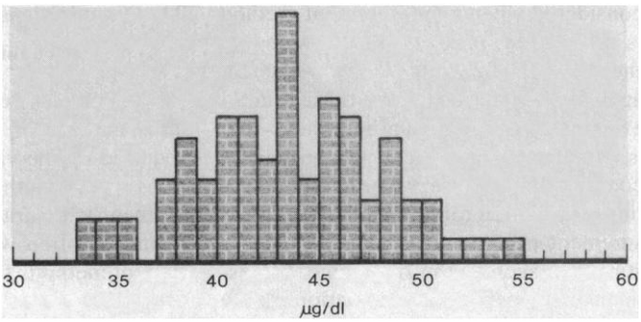

Fig. 1. Individual measurements of blood lead concentration by separate laboratories (*6*).

Interval estimates (confidence intervals). Statistics such as $\bar{y}$ and s^2 are point estimates of the parameters η and σ^2. Interval estimates for parameters are also possible; that is, the data may be used to create two additional statistics, the upper and lower bounds of a confidence interval. The interval is then accompanied by a probability statement: an insurance policy declaring that the interval contains the parameter with a certain confidence. Much care should be exercised in both the construction and interpretation of interval statements for parameters.

Independence. A statistical criterion commonly required in the application of statistics to measurement methods is independence. For most statistical methods, the assumption of statistically independent events is of great importance in establishing the probability of observations exceeding given boundaries, statistical tests of hypotheses, confidence interval statements, and in nonparametric procedures. The physical act of randomization can be used to provide statistical independence.

Statistical stability (stationarity). When the statistical parameters of the data-generating process are constant, the measurement process may be declared "stable" or under "statistical control."

Variance. Observations vary, and any statistic constructed from observations entrains this variability. Whenever the variance of a statistic is large, the width of the corresponding confidence interval for the parameter estimated by the statistic will similarly be large. The variance of many statistics (the average and least-squares estimates) is easy to determine. But there are other commonly constructed statistics for which the evaluation of the variance is not so simple. Consider the following statistic for the amount of SO_2 per microgram per cubic meter:

$$W = \frac{(A - A_0)(B_g)(D)(t + 273)(760)(10^3)}{(V)(P)(298)}$$

where A is sample adsorbance; A_0, reagent blank adsorbance; B_g, calibration factor; D, dilution factor; V, volume of air sampled (liters); P, barometric pressure (mm-Hg); and t, temperature of air sample (°C). Since each of the variables in this expression is a measurement, and hence has an associated variance (and possible covariance with the other measures), the variance of the statistic W must be determined (*16*).

Calibration. The statistical aspects of calibration require the prediction of a response η, considered difficult to measure, through the use of a more readily measured surrogate response ξ. The statistical problems associated with calibration are not simple, and very few calibrations in use fully incorporate appropriate statistical methods.

Ruggedness testing. A ruggedness test of a measurement method is performed by making small patterned changes in selected portions of the measurement protocol and recording the subsequent observations from which, on analysis, it is possible to determine the individual operations in the protocol that give rise to the greatest variance in the recorded responses (*17*).

Round-robin test. Once a measurement method has been proposed, a group of interested laboratories participate in a program to check the adequacy of the written test protocol across the different laboratories, individuals, and equipment. These round-robin studies quickly identify failures in the written protocols and provide preliminary estimates of between-laboratory bias (*18*).

All these statistical criteria should be

considered when a measurement method is established. It is alarming how few of the methods mandated by government agencies have incorporated these criteria even informally. Unusual statistics—that is, peculiar numerical constructions composed from the recorded observations—are often found in published measurement protocols, and almost useless statistics (such as the requirement to estimate the coefficient of variation by using two or three observations) are encountered (*19*). Statistics such as the estimate of the geometric mean are frequently used carelessly (*20*). But most disturbing is the fact that the language of statistics is so poorly used. Since statistics, as a philosophy, has much in common with mathematics, one would think that as much care would be exercised in the use of statistical language descriptive of measurements as in the use of mathematical language. One could even argue that greater care is needed, since the language of statistics must deal with inductive inferences and measures of uncertainty.

The Organizations That Write Voluntary Standards

The burden of establishing the tremendous number of measurement methods required by the various laws and regulations swamped the resources of the government's laboratories and led to the adoption, often with little or no evaluation, of methods found in the literature—preferably those written by organizations that draft measurement standards which are determined by consensus and whose obeyance is voluntary. The use of voluntary measurement standards received the sanction of the Office of Management and Budget in Circular 43 FR-48-51 (3 January 1978): "Voluntary consensus standards will be adopted, whole or in part, and will be used by Federal agencies in lieu of developing and using house standards."

Approximately 200 nongovernment organizations are involved in developing voluntary standards and specifications. The American National Standards Institute forms a loose confederation of these groups. Each organization has committees that write and publish measurement standards and other specifications. The NBS maintains an *Index of U.S. Voluntary Standards*, edited by W. J. Slattery. The second supplement contains 5700 "voluntary engineering standards, specifications, test methods, codes and recommended practices published by 164 U.S. technical societies, professional organizations and trade associations."

A recent joint effort by these organizations and federal agencies resulted in the *National Handbook of Recommended Methods for Water-Data Acquisition* (*3*). This important document establishes, at least for the United States, almost all measurement methods for water characteristics.

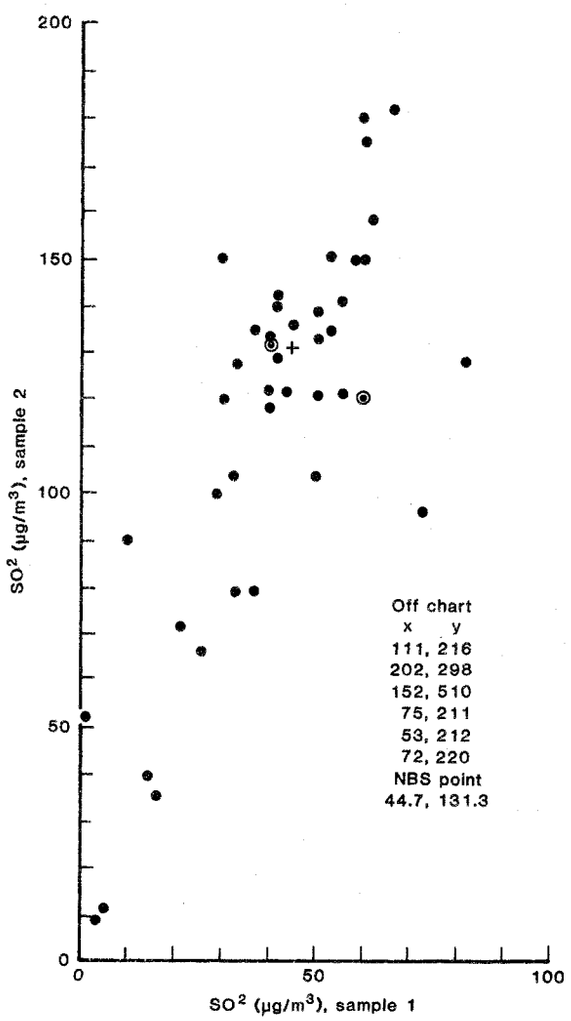

Fig. 2. Youden plot showing SO_2 measurements. Each point represents measurements of two standard samples made by a different laboratory.

The Measurement System

Once a measurement method is approved, attention must be focused on its global use, that is, its application in many laboratories by many individuals. We become concerned with the measurement *system*. For example, the importance of dependable measurement systems should be obvious to anyone contemplating national environmental monitoring. Weather maps, with their isoclines of pressure and temperature and displays of wind speed and precipitation, are widely read and immediately accepted. The reader takes for granted that 0°C is the same condition of temperature in Philadelphia as in Baltimore, and, should the temperature in the two cities be different, he is entitled to interpolate between them. He assumes (correctly) that the measurement systems for the responses displayed on the weather map are under control and are adequate for comparative purposes. Before similar maps can be usefully constructed for, say, the criteria air pollutants, their associated measurement systems will also have to be brought under control. A similar situation exists with the extensive records of the U.S. Geological Survey, which permit ready comparisons of stream flow over different locales during different years, whereas comparisons of the concentrations of most organic or complex inorganic chemicals in the same streams would be suspect. In the absence of dependable national measurement systems, one must ask how power plants can be uniformly compared to determine the efficacy of their scrubbing devices, or water treatment plants judged for their effectiveness.

Measurement System Requirements

A measurement system involves criteria for (i) choice of a sample and the number of samples; (ii) manipulation and preservation of the sample; (iii) control of analytical reagents; (iv) the actual use of the measurement method, including the calibration and maintenance of instruments; (v) the chain of custody of samples and numerical results; (vi) the methods of recording, manipulating, and reporting data; (vii) the training of personnel; and (viii)—of greatest importance—the control of interlaboratory bias. Most of these criteria can be satisfied by carefully preparing procedural handbooks giving instructions on sampling, the use of instruments, preparation of reagents, and so forth. Laboratory certification efforts place great emphasis on the availability and use of such documents. Many of the requirements essential to the control of a measurement system are primarily statistical in nature. However, major concern must lie with the continuing surveillance and control of the measurement system, and this requires ongoing cooperative interlaboratory studies designed to estimate, and then to reduce, interlaboratory biases.

No measurement system can be truly under control without measures of both its repeatability and reproducibility. Repeatability is a measure of the variability (imprecision) between measurements of a single response η within a single laboratory. Reproducibility measures the variability (interlaboratory bias) between measurements of the same response η across different laboratories. Reproducibility usually contributes far more to the total variability of a measurement system than repeatability (*21*).

It is sometimes argued that the reproducibility of measurements across laboratories need not be of great concern, since, given that a laboratory's bias is a constant, relative changes in a measured response are still valid when observed by that laboratory. This is true enough; but scientific acceptance implies that the results can be verified in other locales and by different individuals. To enhance scientific exchange, a mutually acceptable benchmark (the zero of the metric) must be in hand. Science is not that different from commerce: "How much is there?" still remains an important question. And, although a laboratory's bias need not disturb parochial thoughts and decisions, external interlaboratory comparisons would still be essential to ensure that any local bias remained a constant. A measurement method has little value without control over the associated measurement system, and this requires a measure of the system's "reproducible" variance.

When a measurement method is published by one of the voluntary standards-writing organizations or by some responsible government laboratory, it should at least be accompanied by an estimate of the precision of the method. In the *National Handbook of Recommended Methods for Water-Data Acquisition*, a review of the methods for determining inorganic constituents shows that 23 percent have no precision estimate and that only 10 percent have estimates that are based on reasonable interlaboratory studies. For the organic constituents in water, the handbook shows 13 percent without any precision statement and only 17 percent with statements of precision that suggest careful multilaboratory studies. The remaining statements of precision, or of coefficient of variation, are usually obtained from single laboratory reports, many based on few observations. When available, these estimates represent the best efforts of highly motivated individuals. Whether these estimates can be matched in the everyday world is doubtful.

The record indicates that, for most measurement methods, repeatability is poorly estimated, reproducibility is not estimated, and no continuing effort is made to stabilize or control the associated measurement system across the nation's laboratories. This statement is not limited merely to measures of the constituents of water, which, along with measures of radioactive phenomena, have a record of continuing improvement. It is true for almost all measures exclusive of the SI units.

The monitoring and control of between-laboratory biases can be accomplished by shipping samples containing a known or unknown quantity of substance (spiked samples, split samples, and reference samples are common), to each participating laboratory. Given an appropriately designed program involving many laboratories, with several samples sent to each laboratory and repeated measurements made on each sample, a wealth of information becomes available. Laboratory bias and the repeatability of individual laboratories can be immediately determined. A host of graphic techniques can be used to elucidate the data. Even though missing data and aberrant observations almost always exist and can add to the problems of analysis, a well-planned program invariably produces readily understandable information. Complications can occur when measurement protocols and equipment are used to provide measures of many physical characteristics simultaneously. The problems of multivariate calibration, and the between-laboratory control of multivariate responses, are largely unexplored.

Many of the problems of monitoring and controlling a measurement system and ensuring the quality of the information are statistical. In addition, considerable resources are required for the manufacture and distribution of samples to the collaborating laboratories. But the crucial requirements remain the continuing participation of the laboratories employing the measurement method, careful analysis and sharing of information, and some mechanism for improving the system further.

Laboratory Certification

Although the voluntary standards-writing organizations now have the major responsibility for establishing measurement methods, they have historically done very little to monitor the use of the methods. Generally, it is assumed that each laboratory will monitor its own performance. However, the burgeoning number of mandated measurement methods and the acknowledgment of the excessive interlaboratory biases have led to efforts to control the measurement system through laboratory certification.

A need for certifying clinical laboratories (of which there are approximately 15,000) was recognized when the Clinical Laboratory Improvement Act of 1966 authorized certification by the Communicable Disease Center. (The Medicare program of the Social Security Administration may also specify criteria for laboratory certification.) Unfortunately, very few federal funds have been allocated specifically for either accrediting or auditing the clinical laboratories that come under the Act's jurisdiction. The Act has largely eliminated egregious laboratory performances, but its effectiveness has only been modest in controlling the measurement systems in its zone of interest.

The Occupational Safety and Health Act (1970) made provision for the accreditation of laboratories (*14*). In 1974, the Safe Drinking Water Act defined a primary drinking water regulation as one that includes "quality control and testing procedures," and this led to the initiation of a National Quality Assurance Program for Laboratory Certification by the Environmental Protection Agency. Conversations at a national conference

on laboratory certification, sponsored by the American Public Health Association in 1977, resulted in the establishment of the American Association for Laboratory Accreditation.

On 25 February 1976 the Office of the Secretary of Commerce promulgated procedures for a National Voluntary Accreditation Program to fill "regulatory and non-regulatory *product* evaluation and certification needs." The program, inaugurated on 1 January 1979, is voluntary and includes cooperative efforts involving the private sector (*22*).

A review of the procedures and technical manuals associated with laboratory certification or accreditation programs shows them to be largely passive. They require that the equipment, personnel, and procedures be adequate. The control of within-laboratory measurement variability (repeatability) is usually met by a requirement to run a duplicate every ten or so samples. The laboratory is also required to participate in interlaboratory programs, that is, to report on occasional reference or unknown samples. Interlaboratory performance is, however, left largely to informal analysis. There is seldom any guarantee that the results of analyses by different laboratories will be distributed to the laboratories or made public. An aggressive effort to reduce interlaboratory variability is lacking.

The scientific fraternity and the public should keep in mind that it is the measurement that requires our confidence. Certification will lend respectability to a laboratory's performance, but it may say very little about the data. If measurements are to be used by legislators, regulators, the courts, the press, industry, and the public, then the data must be gathered under a system of active, frequently analyzed, and openly published interlaboratory control programs.

Responsibility

In its enabling legislation, the NBS was declared responsible for "the custody, maintenance and development of the national standards of measurement," but this role is now largely limited to the SI units and closely associated quantities. The major burden for establishing measurement methods has fallen upon the voluntary standards-writing organizations. Unfortunately, these organizations have little ability to control the associated measurement systems. To gain system control, the regulatory agencies are beginning to establish separate laboratory certification protocols, and we have seen establishment of several nongovernmental organizations committed to certifying laboratories.

The responsibility for and control of the nation's measurement systems is poorly centralized. It is possible that this diffusion is healthy for the development of viable measurements. Further, it is probably impossible to coalesce the nation's diverse measurement requirements into any single pattern. But clearly, the quantity of the scientific measurements now required by our measurement-intensive laws and regulations are piling up, while many of the desirable physical and statistical characteristics of good measurement methods and associated measurement systems are being given short shrift. The result is that the quality of many scientific measurements is suspect. The time appears ripe for a review of the adequacy of our present approach to scientific measurement.

References and Notes

1. R. C. Sang, *Final Summary Report: Study of the National Measurement System, 1972-75* (NBSIR 75-924, National Bureau of Standards, Washington, D.C., 1976).
2. S. S. Miller, *Environ. Sci. Technol.* **12**, 18 (1978).
3. *National Handbook of Recommended Methods for Water-Data Acquisition* (U.S. Geological Survey, Reston, Va., 1977).
4. S. Weber, *Evaluating the Accuracy of Biochemical Oxygen Demand and Suspended Solids Analyses Performed by Wisconsin Laboratories* (Technical Bulletin 114, Department of Natural Resources, Madison, Wis., 1979).
5. *Monthly Status Report of NPDES Quality Assurance Pilot Study in New Jersey* (Environmental Protection Agency, Washington, D.C., 1979).
6. *Blood Lead Proficiency Testing, April 1980* (Center for Disease Control, Atlanta, 1980).
7. W. J. Youden, *Ind. Qual. Control* **15**, 24 (1959).
8. _____ and E. H. Steiner, *Statistical Manual of the AOAC* (Association of Official Analytical Chemists, Arlington, Va., 1975).
9. J. Mandel, *Stand. News* **5**, 10 (1977).
10. W. A. Stockmayer, *National Needs for Critically Evaluated Physical and Chemical Data* (National Academy of Sciences-National Research Council, Washington, D.C., 1978).
11. G. E. Brown, Jr., *Hearings on Environmental Monitoring* (Congressional Research Service, Library of Congress, Washington, D.C., 1978), p. 6.
12. W. A. Lyon and J. S. Hunter, *Prog. Water Technol.* **13**, 237 (1980).
13. J. M. Cameron, *J. Qual. Technol.* **7**, 193 (1975).
14. *Fed. Regist.* **38**, 175 (1973).
15. C. Eisenhart, *J. Res. Natl. Bur. Stand.* **67c**, 161 (1963).
16. H. H. Ku, *ibid.* **70c**, 263 (1966).
17. W. J. Youden, *Mater. Res.* **1**, 862 (1961).
18. G. Wernimont, *Stand. News* **5**, 13 (1977).
19. J. Pratt, *Environmental Monitoring* (National Academy of Sciences-National Research Council, Washington, D.C., 1977), vol. 4, p. 96.
20. J. M. Landwehr, *Water Resour. Res.* **14**, 467 (1978).
21. Detailed definitions and explanations of repeatability and reproducibility appear in the *Annual Book of ASTM Standards* (Document ANSI/ASTM E 691-79, American Society for Testing and Materials, Philadelphia, 1980), part 41.
22. *First Annual Report, National Voluntary Laboratory Accreditation Program* (NVLAP-77, Department of Commerce, Washington, D.C., 1977).