Quality Control in the Office Laboratory*

JOSEPH H. RIDDICK, JR., M.D.
Associate Professor of Clinical Pathology, Medical College of Virginia, Health Sciences Division of Virginia Commonwealth University, Richmond

"The poorest of all laboratory work is that done in the private physician's office." That allegation was made by the Director of the New York City Department of Health Bureau of Laboratories in a statement to a Senate Committee of the Judiciary Committee on Antitrust and Monopoly on February 7, 1967 (1). Statements such as this typify the public relations problem facing directors of office laboratories today. The problem is twofold: First, how can the quality of the office laboratory be assured and documented? Second, how can this information be disseminated to legislators and the general public? The first is an easily solved technical problem toward which this paper is directed. The second and probably the most difficult is a media problem beyond the scope of a technical journal.

Laboratory quality assurance and documentation are best considered as a subset of the more general field of quality control, an industrial discipline with its roots in the industrial revolution. Industrial quality control is the technique, for instance, whereby the automotive industry may precisely design obsolescence into an automobile without impinging on the warranty period in order to maximize profits. Quality control in the drug industry assures not only that the amount of drug in a pill falls within acceptable tolerances but that the pill itself is not malformed. A malformed pill results in very bad public relations for the drug company, even if its function is not impaired. The pigeon, an animal with a very fine eye for detail, has been used successfully for assembly line quality assurance in removing faulty pills prior to packaging (2). Laboratory medicine also produces a product which can be subjected to quality control, the laboratory result.

Error Frequency Distributions. Any discussion of laboratory quality control should begin with a brief account of practical statistics, since statistics is the science whereby a decision can be made as to whether a laboratory result is acceptable or unacceptable. An introduction to these statistics should begin with the error frequency distribution.

If a unit of outdated blood bank blood is mixed well and dispensed into 100 test tubes, this set of tubes is called a "pool." The plasma glucose content of any one of these tubes is the same as that of any other, since all tubes came from the same bag of blood. If one attempts to assay each of those 100 tubes of blood for glucose, however, the assay results will not be identical for all tubes. Some will be above an average value and some below, as shown in the histogram in figure 1. A few results will deviate markedly from the average value but most will cluster around the average. The smooth curve (sometimes called Gaussian or bell-shaped) drawn through the steps of the histogram is called a frequency distribution plot, and it can be used to make a decision as to when a laboratory result deviates too far from the average value of the pool. If one calculates a so-called standard deviation (SD), one can define location of the cutoff point between acceptable and unacceptable results. The use of the standard deviation has two advantages over a single intuitive guess about where the line between good and bad data lies. First, use of the standard deviation (actually ± 2SD) as a decision point guarantees that 95% of the results will be in the acceptable range on the average. Second, standard deviation is used by most practitioners of laboratory medicine and is understood by them, thereby providing a common ground for discussion. It might be mentioned that the frequency distribution plot shown in figure 1 is a real life situation. That is, it is technically impossible to get
identical results for the plasma glucose in each of those 100 tubes. One can use more costly and difficult methods and the range on the horizontal axis will narrow, but the results cannot be made identical. This means that there is no "true" value for the glucose in the plasma. Analytical balances used for weighing the glucose also have error frequency distributions.

**Frequency Distributions of Normal Patient Laboratory Determinations.** Figure 2 shows the results obtained for blood urea nitrogen determined on each of 495 "healthy" patients plotted in the same fashion as the pool results in figure 1. As might be expected, most of the results cluster around an average value with a few markedly deviant values to the left and right of center. Notice that the frequency distribution plot is skewed with more deviant values to the right (occult disease?) than to the left. This is because there is no such thing as a negative blood urea nitrogen, so the curve is steeper on the left than on the right.

Now, suppose one took a large number of patients with renal disease, drew their blood, performed a blood urea nitrogen on them, and superimposed their frequency distribution plot on an idealized plot of the frequency distribution of urea nitrogen of "normal" patients (fig. 3). How might one define the upper limits of "normal" using this data? One might draw a vertical line at the lowest point between the normal peak and the sick peak and call this the upper limit of normal. It is obvious, however, that the two curves overlap and in the areas of overlap, some normal patients will be called sick when they are not, and some sick patients will be called normal when they are not. Anyone experienced with the SMA 12-60 has seen elevated uric acids in perfectly normal people who never develop gout. To further complicate things, a large laboratory error component will tend to broaden this range of ambiguity to an even greater extent.

**Effect of Methodology on Normal Ranges.** Figure 4 shows frequency distribution plots of two different assay methods for blood glucose—the glucose oxidase and the ferricyanide methods. The glucose oxidase method is a so-called "true glucose" method and produces comparatively low results, partly because it is subject to inhibition by some patients' blood. The classical ferricyanide glucose method, on the other hand, produces comparatively high results because it measures not only true glucose but also other reducing substances such as fructose and glyceraldehyde. From the point of view of the patient and his physician, the important thing is the normal range. The glucose oxidase blood glucose normal range is 10–20 mg percent lower than that of the ferricyanide method. If one used both methods...
simultaneously, one would widen the normal range as shown in the composite curve in figure 4 and a substantial increase in false negative blood glucose results would occur. This could easily happen if two different kits using different methodology were used interchangeably. The result would be a component of laboratory error which would widen considerably the range of ambiguity of the blood glucose data.

**Effect of Methodology on “Biologic Variation.”**

Figure 5 is a time-course plot of the blood sodium level of the author drawn over a period of a month. The upper curve (JHR) represents the human blood; the two lower curves (Control 1 and Control 2) are two commercial pools assayed simultaneously with the human blood. Had there been no laboratory error, the lower plots (Control 1 and Control 2) would be perfectly straight lines parallel to the horizontal axis. Note that the two curves are sawtooth instead, describing day-to-day laboratory error. Because of within-day random error, or perhaps because of pipetting error they are not exactly parallel to one another. The most striking feature of figure 5 is that the human serum drawn on the same day shows the same pattern of variation. The implication is that a substantial component of so-called “biologic variation” for serum sodium actually is laboratory error. Carefully controlled studies such as this, in fact, indicate that the flame photometer is incapable of measuring the subtle changes of biologic variation and that, in fact, all variations measured in normal human blood sodium are actually laboratory error not biologic variation. Many other blood components, potassium, for example, show similar effects of day-to-day laboratory error. If one could reduce the day-to-day error component in measuring blood sodium, one might be able to pick up more subtle changes, the normal range would undoubtedly shrink and fewer false negative and false positive results would occur. The cost of increased accuracy and precision, however, is prohibitive at present.

**Setting of Control Limits and the Effect on Patient Care.**

Figure 6 is a graphic example of the effect of control limits set too wide in the laboratory. Bilirubin is traditionally a poor test from an accuracy standpoint because of the difficulty of maintaining adequate standards which will not deteriorate. The error shown in figure 6, however, is caused by an improperly calibrated reference serum used to calibrate the SMA 12–60 bilirubin. On day 20, the laboratory used the last of a particular manufacturer’s reference serum and began to use a new lot from a different manufacturer. The effect was an abrupt increase in every normal patient's bilirubin of 0.4 mg percent and an increase in the bilirubin of the patients in the moderately elevated range of 0.7 mg percent. Obviously the normal range was widened by this change. Tighter control limits would have detected this change sooner. Furthermore, an alert visual scan by a physician of the patient results for the day would have proven to be an extremely useful form of quality control. It was concluded that one could not rely upon the manufacturer’s brochure provided with the lots of reference serum.

**Some Simple Techniques for Quality Control.**

One of the oldest forms of laboratory quality control is the “repeat.” If one doubts the validity of the first result, send another one and compare the two.
is mixed well and divided into two parts. A phony name is attached to one specimen and the second member is submitted to the laboratory sometime later. This technique is designed to foil the laboratorian who claims that “biologic variation” caused the difference in the two results. A rule of thumb for the split sample is that if the results of the two split samples vary by more than 10% the results are suspect. This figure varies from test to test. The split sample is also useful in those tests for which reference standards are not yet available such as urinalysis, bacteriology and even some coagulation studies.

Another more complex method of quality control is the mixed specimen. This works best with blood chemistries. If you mix known amounts of each of two patient specimens, the results of all components (excluding enzymes) will be proportional to the original concentrations and volumes mixed. The calculation for the predicted results is fairly straightforward.

Most people rapidly tire of the exercise of preparing their own quality control samples and purchase these from some national program such as the Proficiency Evaluation Program (PEP) for the Physician’s Office Laboratory.

**Proficiency Evaluation Program for the Physician’s Office Laboratory.** Beginning in April, 1973, the College of American Pathologists and the American Society of Internal Medicine are jointly sponsoring a quality control program designed specifically for the physician’s office laboratory. The data obtained are treated in an entirely confidential manner and meet the demands of public health agencies, legislators, professional associations and patients for a third party evaluation of the office laboratory. Quarterly kits are mailed out to subscribers containing seven vials of test unknowns and a blood smear. These vials contain material for evaluating hemoglobin, red cell counts, hematocrit, glucose, bilirubin, cholesterol, urea, uric acid, urine specific gravity, urine protein, urine reducing substance, urine bilirubin, urine hemoglobin, urine pH, prothrombin time, white cell count, urine bacteriology and peripheral smear. One hundred tests per year are performed on a quarterly basis and mailed into the testing agency for evaluation. Four weeks after receipt, results from all subscribing laboratories are pooled, processed by computer, and a computer print out is mailed to all subscribers giving name of constituent, method used, subscriber’s result pass or
fail code, average of all laboratories for comparison, a good performance range and an acceptable performance range. In addition, a quarterly bulletin is sent to all subscribers, describing findings and including a personalized certificate for wall-mounting.

The advantages of such a system are apparent. First, the director gets a confidential warning that his laboratory is having difficulty, thus giving him the opportunity to correct the difficulty. Comparison of results with peers is always informative as to the wide interlaboratory differences. The effort expended is minimal and does not disrupt the normal function of the laboratory and the cost is far less than manual preparation of samples and data collection.

REFERENCES
