A Three-Parameter Statistical System for Radioimmunoassay Quality Control

Sara G. Sarno

Indiana University School of Medicine, Indianapolis, Indiana

We have implemented an efficient system of radioimmunoassay quality control using three statistical parameters: a Z-score plot for each control pool, a Z-sum plot that combines the Z-scores, and an average of normals graph. Concrete criteria were established for evaluation of assay validity. Use of these three parameters allows differentiation between random and systematic errors, detection of control sera biodegradation, and information as to sources of systematic errors. The system also allows increased reliability of results, fewer repeat assays, and rapid problem recognition and solution with a negligible increase in time spent on quality control procedures.

Every nuclear medicine in vitro laboratory should have a quality control program that meets a number of requirements: improved quality of the results, aid in detecting problems promptly, prevention of unnecessary repeat assays, and establishment of concrete rules for evaluating assay validity that could be used by any technologist.

In our busy laboratory (12,000 routine assays per year as well as research and technologist training), we have developed a quality control program based in part on the methods reported by Anderson et al. (*1*). It has resulted in early detection of deterioration of the control sera and therefore fewer repeat assays. In addition, it has given us a higher degree of confidence in the precision of our results and a concrete method of determining an assay's validity. We have modified and improved these methods and have found it satisfies the needs of our laboratory beyond what we had hoped.

Methods

For each test system, two or more control pools were obtained and assayed repeatedly until at least 30 values were obtained for each pool. The mean and standard deviation for each pool were calculated by the formula:

$$\bar{X} = \frac{\Sigma X}{N}$$

where $\bar{X}$ = the mean of the values; X = the sample value; and N = the number of samples.

and

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N - 1}}$$

where σ = the standard deviation.

These values are updated monthly to compensate for small shifts in the control pools. With these data, three types of graphs were constructed for each system.

A Z-score chart (similar to a Shewhart chart) (*2*) was constructed for each control pool used in the test system. The Z-score was calculated from the formula:

$$Z = \frac{X - \bar{X}}{\sigma}.$$

This simply converted the actual value of the control to its distance from the mean value in terms of its standard deviation. The Z-score chart was constructed on a linear graph paper with the vertical axis evenly divided and labeled −2, −1, 0, +1, and +2 where 0 was the mean and the remaining values were the standard deviations from the mean. The horizontal axis represented the day on which the test was run.

A Z-sum chart was constructed in the same fashion. The Z-sum was determined by the formula:

$$\text{Z-sum} = \frac{Z_1 + Z_2 + Z_3 + \ldots + Z_N}{\sqrt{N}}$$

where Z_1, Z_2, etc. are the Z-scores for the individual control pools

and

N = the number of control pools in the assay.

In addition, an average of normals chart (*3*) was constructed for each test system with the normal range on the vertical axis and the date of assay on the horizontal axis. The limits for average of normals values were indicated on the chart and were determined as follows:

1. A midpoint was calculated by averaging the values obtained from a large number (600–800) of normal patients using the normal range as strict limits.

For reprints contact: Sara G. Sarno, Division of Nuclear Medicine, UH P-16, Indiana University Medical Center, Indianapolis, IN 46223.

2. The standard deviation of the normal range was estimated by dividing the normal range by 4.
3. The standard deviation of the average of normals was determined for both 5 and 10 tests used in an average by dividing the standard deviation of the normal range by the square root of the number of tests used in an average.
4. The 95% confidence interval was determined for each situation by using the calculated midpoint ± 2 standard deviations of the average of normals.

Example: For T_3RIA, the normal range is 100–200 ng/dl.

Standard deviation of normal range =

$$\frac{(200 - 100)}{4} = 25.$$

Standard deviation of average of normals for

$$5 \text{ tests} = \frac{25}{\sqrt{5}} = 11.2.$$

Standard deviation of average of normals for

$$10 \text{ tests} = \frac{25}{\sqrt{10}} = 7.9.$$

95% confidence interval for 5 tests
$= 137.8 \pm 2(11.2) = 115.4\text{–}160.2.$
95% confidence interval for 10 tests
$= 137.8 \pm 2(7.9) = 122.0\text{–}153.6.$

The midpoint and standard deviation of the average of normals were calculated only once for a particular test system since the normal range was permanently established and the midpoint was determined by using a very large normal patient population. The assay values of all the patients that fall within the normal range were averaged and plotted on the chart each day.

The calculations can be facilitated on a daily basis by programming the computer not only to calculate the assay results, but also the Z-score of each control sample, the Z-sum, and the average of normal patients.

Criteria for Assessment

I. If the value of the Z-sum is $< \pm 2$, accept the assay, except in the case where more than one control is $> \pm 2$ on the Z-score charts. This is probably (but not necessarily) a systematic error.
II. If the Z-sum is $> \pm 3$, reject the run. This is always a systematic error.
III. If the Z-sum falls between ± 2 and ± 3:
  A. Examine the Z-score charts:
    (1) The values should show a random distribution around the mean.
    (2) If only one control value is out of control limits, this is a random error and the results may be accepted.
    (3) If two control values are out of limits, it may be a systematic error. Examine the average of normals chart:
      (a) If the average of the normal patients in the assay shows no significant change from previous days or weeks, accept the run as this is probably a random error.
      (b) If the average of the normal patients and the mean of the control pools have shifted in the *same* direction, this is a systematic error; reject the run.
  B. Examine the Z-sum chart:
    (1) If the Z-sum shows a trend or shift, check each Z-score chart.
      (a) If only one control shows a shift, this is a random error.
      (b) If more than one control shows a shift, this is a probable systematic error and equipment, reagents, etc. should be checked.
      (c) A small shift in one control (< 1 standard deviation) is only significant if it persists for three or four days.
    (2) If the Z-sum is between ± 2 and ± 3, check all Z-scores and the average of normals. If they have all shifted in the same direction, this is a systematic error. However, the patient values can still be reported if the normal range and the average of normals shift are not significant.
    (3) If the Z-sum is $> \pm 2$ for two consecutive days, and even if all the controls are $< \pm 2$, this is a systematic error and the results should not be reported.
IV. Use of the average of normals chart:
  A. If there are seven tests or less in the average of the normal patients, the five test range must be used. Fewer than three tests in an average cannot be used, and fewer than five tests must be interpreted with caution.
  B. If there are eight tests or more in the average of the normal patients, use the ten test range for interpretation.

Discussion

Any quality control program should have improved quality of results as its prime consideration. The system we are using provides this by the use of the Z-sum. This factor enhances small changes in the control sera or in the analytic system. The system of using controls and their mean and standard deviation is not very reliable. Often no attention is paid to simultaneous shifts of the controls around the mean and if all of the control values are within ± 2 standard deviations, the assay is assumed to be valid and accepted without question.

However, we observed that in some assays, when all the controls were shifted in one direction, though remaining within control range, the resulting Z-sum was $> \pm 2$. Examination of the average of the normal patients sometimes showed a significant shift in the same direction. Acceptance of such an assay would result in pushing borderline normal patient values into the abnormal range and a systematic error would go undetected (Fig. 1).

On the other hand, random errors can be detected as opposed to systematic errors. A good example of a random error is the occasional pipetting error. This could affect one or even possibly two controls, and the result would be those controls affected falling out of range. Obviously, this is not a systematic

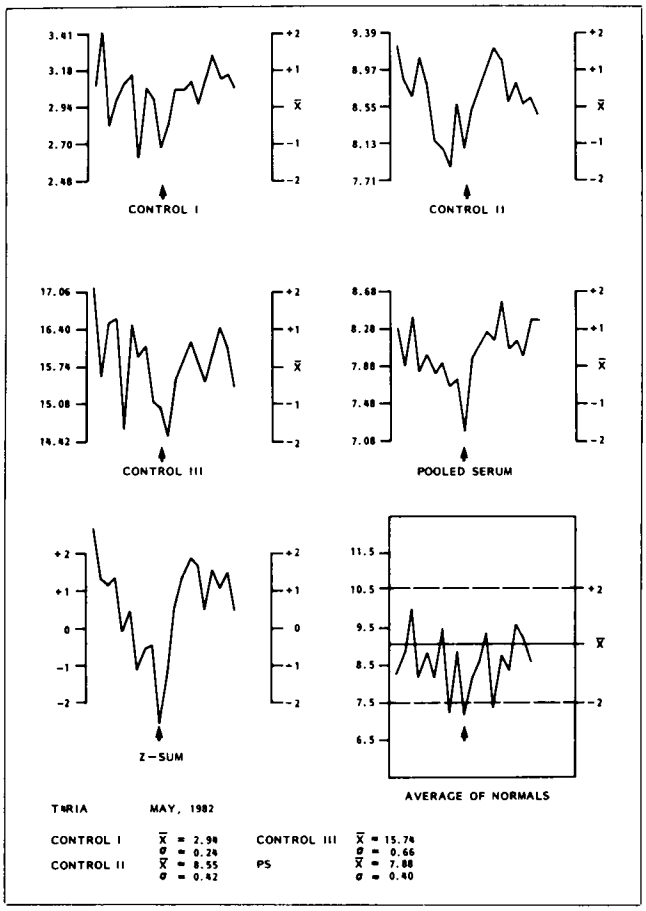

FIG. 1. An isolated case of systematic error showing downward shift (arrow). Note that all controls are in range, but Z-sum is > 2 s.d. Also, the average of normals shows a substantial shift in the *same* direction.

error since it only affects the specific control sample(s). Examination of the shape of the standard curve and the average of normals plot would solve this dilemma and the assay would be accepted (Fig. 2).

In a busy laboratory, early detection of a problem in an assay system and clues to the cause of the problem are very important. The problem needs to be detected and solved promptly to avoid delay in reporting results and prevent the constantly mounting backlog of samples that results from assay reruns. Since control sera are biodegradable even when stored under optimum conditions, it is important to monitor the stability of such sera. Deterioration of the control pools can be disastrous if not detected promptly. Obtaining a new control pool and establishing new control ranges is a time-consuming task. In reality, the degradation of controls is a "random" error and does not affect either the standard curve or the patient results.

In the fall of 1982, we encountered such a situation. Not only did our three-level commercial control lot show definite signs of deterioration (9 months before expiration), but also our in-house control pool degraded rapidly. Within one week, we determined that biodegradation of controls was the problem. We obtained a new commercial control lot, prepared a new in-house pool, and began establishing new control ranges. We were able to continue releasing patient results by the use of the average of normals chart and careful monitoring of the standard curves during this period. The new control pools were evaluated and ready to take the place of the old lot by the time the old control pools were totally deteriorated (Fig. 3).

True systematic errors are those factors that affect the entire system, e.g., a pipette out of calibration, a deteriorated reagent, deteriorated standards, unstable incubation temperature, or a well-counter out of calibration. Such problems need to be resolved quickly to prevent repeat assays. An obvious example of this situation is Fig. 4. Within three days, we determined that the sampling pumps on the automated pipetting system required cleaning—two weeks before the regularly scheduled maintenance. As soon as they were cleaned, the Z-scores and Z-sum graphs fell back into their normal patterns.

Conclusions

This system of quality control has greatly enhanced our ability to use our laboratory efficiently. Use of the Z-score chart as opposed to the standard Shewhart chart has resulted in reduction of plotting time, and it is more accurate since a standard deviation factor is plotted rather than the actual control value. The Z-sum value, since it combines and enhances the relationship to the mean of all the controls used, allows us a "one-glance" method for quick acceptance of in-control assays. The average of normals chart gives us an additional

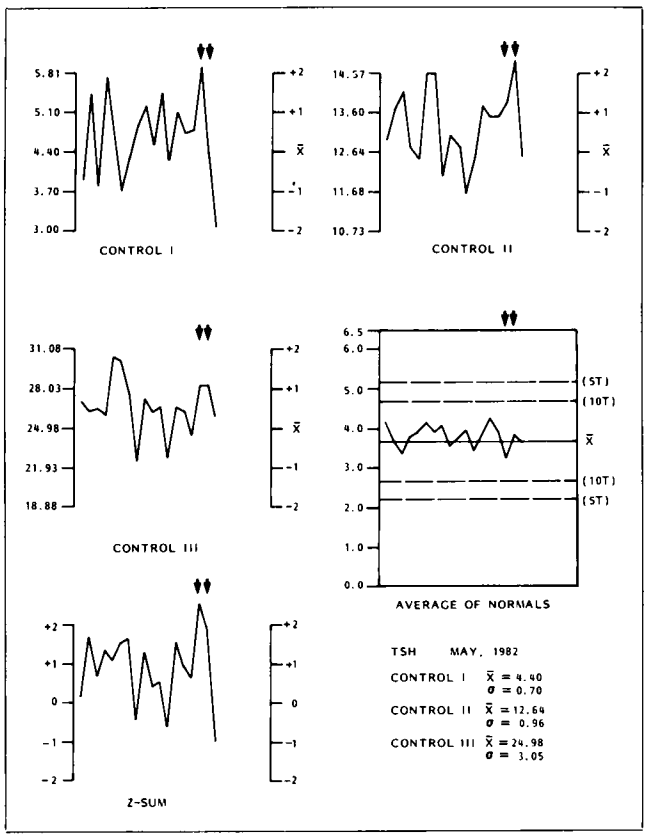

FIG. 2. Two examples of random error (arrows). Note that average of normals holds steady close to the mean.

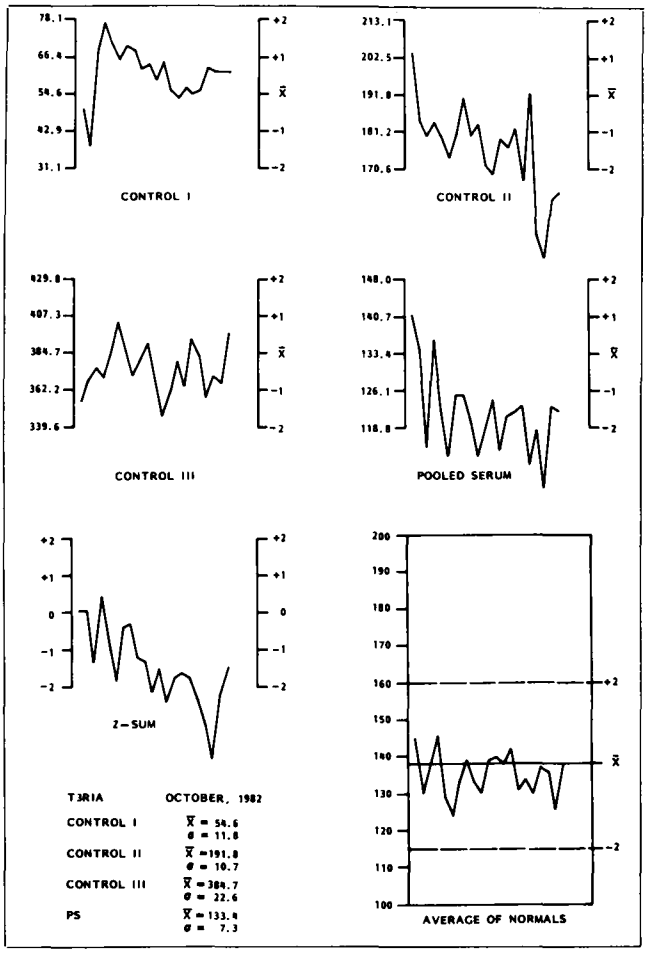

FIG. 3. Deterioration of control pools. Note that average of normals holds steady indicating that this is a random error affecting only controls. Standard curves also appeared normal.

factor with which to evaluate assay results that fall in a questionable area.

On the surface, it would appear that the use of this system would result in a substantial increase in time in calculations and plotting. However, with experience, the time consumed in this chore becomes almost incidental. Indeed, this is offset many times over by the reduction in assay repeats and rapid problem solution. With implementation of a computer program to do these calculations automatically, this time factor would become almost negligible.

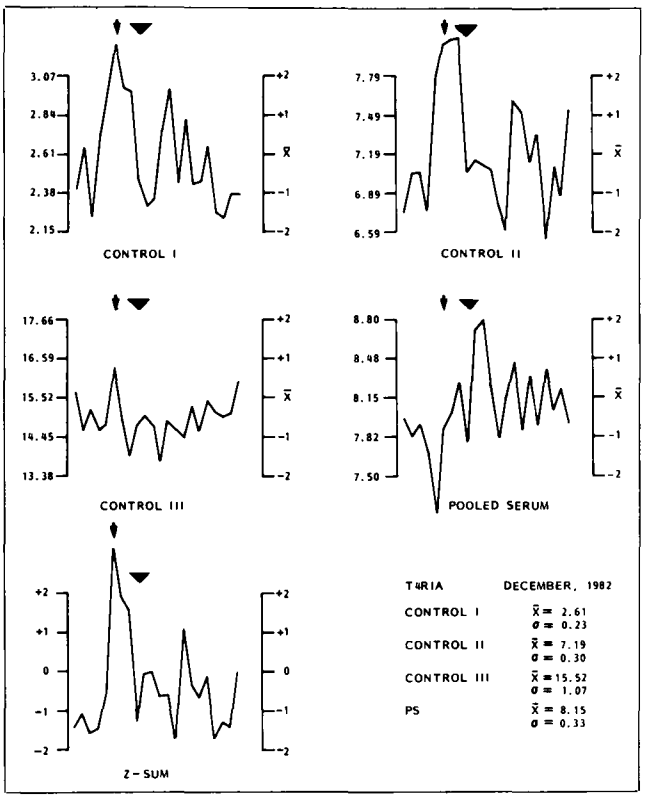

FIG. 4. Systematic error due to equipment failure. Pumps on automatic pipetting system needed cleaning (arrows). After cleaning (arrowheads) graphs returned to normal pattern.

Above all, this system has given us increased confidence in the precision of our results.

Acknowledgments

I wish to thank Mr. Ira D. Miller for helping me with the research material and Dr. Aslam R. Siddiqui for his editorial assistance in preparing this manuscript.

References

1. Anderson YJ, Troxler RG, Wease DF, et al. Three-pool analysis for quality control: a method for daily differentiation of analytic error from random variability. *Am J Med Technol* 1981;47:321-26.

2. Westgard JO, Barry PL, Hunt MR, et al. A multirule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 1981;27:493-501.

3. Hoffman RG, Waid ME. The "average of normals" method of quality control. *Am J Clin Pathol* 1965;43:134-41.