# A System for Rating the Performance of Laboratories in Analytical Check Series

C. H. PERRIN and E. M. GLUCKER,[1] Canada Packers Ltd., Toronto, Canada


## Abstract

The system described rates the accuracy of the analyses of laboratories participating in meat, oil, feed and fertilizer check series but it has a much broader application. It is based upon the assumption that the relative quality of such analytical work may be judged by comparing the normalized deviations of the reported analyses from the best estimates of their respective correct values. A computer program was written to carry out the required calculations and comparisons. We follow the concept of accuracy expressed so well by Eisenhart (1), Murphy (2), Youden (3) and recommended by ASTM Committee E-11 (4). Thus the accuracy of a measuring process is its freedom from total error. The total error is described by presenting estimates of the error of imprecision of the process and of its bias error relative to a suitable reference value. It is understood that the apparent accuracy of a measuring process is conditioned by the particular population of experimenters using it.


## Introduction

Chemists participate in analytical check series for a number of reasons. Primarily they wish to test their analytical skill and the reliability of the methods they normally use in their routine work. At times they hope to obtain a valid comparison of the reliability of various chemists, or instruments, or a comparison of the performance of new methods with that of established methods.

An assumption is made that all portions of the total check sample have essentially the same composition when analyzed so that differences between laboratories may be attributed solely to the analytical process. Since the analysts usually recognize the check samples as tests of their skill, the performance of the laboratories on a check series is expected to be better than routine.

The proposed system of rating laboratories assumes that a skilful analyst may be distinguished from a lucky analyst by the former's consistent ability to report analyses which closely approach the estimated true value which we call the ETV. For want of something better we normally must use the censored mean value of the analyses reported as our ETV. Our confidence in this estimate increases as the number of reported analyses grows and as the interlaboratory agreement improves. We gain further confidence in our ETV if the average values reported by different methods show good agreement.

## Discussion

The first step in rating the performance of the laboratories is to calculate the mean values for the analyses requested and the interlaboratory standard deviations around these mean values. Analyses which deviate by more than three standard deviations from the mean values then are excluded and the means and standard deviations from these censored means are calculated.

The next step is to compute the normalized deviations of the reported analyses from the censored means. First we subtract the ETV from the reported analysis, then we divide these deviations by the corresponding interlaboratory standard deviations. For example, if a laboratory reported 19.0% and the ETV were 20.0% with an interlaboratory standard deviation of 2.0%, we would find its normalized deviation was −0.500. Thus, the reported analysis was one half standard deviation below the censored mean of the laboratories.

Most check series request that several different analyses be reported and we rate the overall performance of a laboratory by the magnitude of its typical normalized deviation or TND. This is the quadratic mean of the individual normalized deviations and is obtained by extracting the square root of the average of the squared normalized deviations. The negative signs are eliminated in the process.

Finally, the laboratories are ranked in order of the magnitude of their TND. This is analogous to ranking the accuracy of shooting by keeping track of the distance by which marksmen miss the bull's-eye. The normalizing process has the effect of causing the targets to remain equal in size since it automatically allows more generous tolerances for the more difficult analytical tasks. Thus, the analyst gains nothing by failing to perform certain requested analyses. Occasionally all the data reported for a method of analysis are excluded from the calculation of the TND. This usually is due to lack of confidence in the ETV caused by insufficient data or by an extremely large interlaboratory standard deviation.

At this point in the development of the rating system the following information will be given in the report to the collaborators: censored mean for each analysis requested; censored standard deviation for each analysis requested; normalized deviation of each reported analysis; the TND for each laboratory based upon the methods which qualify for rating; and rank of each laboratory performing the minimum number of analyses to qualify for rating. (The minimum is to be chosen by the check series committee.)

The chemist should carefully study his individual normalized deviations to learn which analyses are contributing most to his typical deviation or error. He should also look for any trends in particular analyses which are repeated in successive check samples. This may reveal significant long-term systematic errors in some of his procedures.

A computer program is being developed which will summarize the performance of the laboratories for a year or other appropriate period. It is expected that the revelation of the existence of persistent trouble spots will lead to gratifying corrective action on the part of the collaborators. Knowledge of the character of the errors assists greatly in finding solutions to these problems.

When the ETV is the consensus of the participating laboratories we may say that an individual analysis X has a total error of X-ETV. This is made up of the laboratory's error of imprecision (as measured by its intralaboratory standard deviation) and the laboratory's systematic error or bias from the consensus of the group of laboratories. When, however, we are able to submit, as a check sample, a material of definitely known composition we may estimate a third error. This is the bias of the method used by "this group of laboratories on this sample." If the net effect of the three errors is zero by good fortune or if the random component happens to be zero together with zero values for the laboratory bias and the methodic bias, the reported value will agree with the known analysis of the reference material.


REFERENCES

1. Eisenhart, Churchill, J. Res. Nat. Bur. St., *C67*, 161–187 (1963).
2. Murphy, R. B., Res. St., *1*, 264–267 (1961).
3. Youden, W. J., Ibid. pp. 268–271.
4. ASTM Committee E-11, "Use of the Terms Precision and Accuracy as Applied to Measurement of a Property of a Material," E 177-61T.


[Received July 2, 1968]

[1] E. M. Glucker is Manager of Mathematics, W. R. Grace & Co., Washington Research Center, Clarksville, Md. 21029.