

Use and Interpretation of Common Statistical Tests in Method Comparison Studies

James O. Westgard¹

Feature Article: Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. *Clin Chem* 1973;19:49–57.²

Marian Hunt and I published this paper to improve the understanding and application of commonly used statistics in method validation studies. We specifically wanted to clarify the misuse of the correlation coefficient and *t*-test statistics, and to point out important factors that affected the reliability of least squares statistics. This paper also provided some direction in making decisions on the acceptability of a method by using statistics to estimate the sizes of analytical errors. A subsequent paper (1) provided criteria for comparing those estimates of errors with quality standards that defined the amount of allowable error.

This paper was written at a time when I was very involved in the evaluation of new automated analytic systems. At that time little guidance was available as to how to analyze and interpret the data from method evaluation experiments. Decisions on acceptability of new methods seldom provided any rational explanation as to why a method was judged acceptable or not. The correlation coefficient was the statistic most often being used to justify decisions on acceptability, followed by the *t*-test, in which the decision was based on whether or not the *t*-value indicated a statistically significant difference (i.e., calculated *t*-value greater than the critical *t*-value) between methods. Neither of these approaches took into account the actual quality required for the application of laboratory tests.

Our approach was to experiment with different sets of data to see how the statistics responded to different analytical errors. Data sets were constructed to include known types and amounts of analytical errors. This investigation was a simple simulation study, made possible by new computer technology, a CompuCorp 344 Statistician desk calculator, which allowed me the

luxury of analyzing the data in my office rather than having to prepare punch-cards for use at the university's central mainframe computer. Today this study could be readily done on a personal computer using an Excel spreadsheet!

This early work provided me with many opportunities to make presentations at scientific meetings, and also led to the publication of an educational monograph (2) that became the basis for the American Association for Clinical Chemistry's longest running annual workshop, entitled Method Evaluation, presented by Carl Garber, Neill Carey, and David Koch. It was also a precursor to my work on statistical QC to assure the ongoing validation of method performance during routine operation. Few people recognize that the multirule QC procedure known as Westgard Rules (3) was the outcome of a similar simulation study (4). The control rules were the statistics, and the probability for rejection quantified the response of different rules to different errors.

Thirty-five years ago there was a great interest by analysts to understand method evaluation, how to do it right and how to correctly interpret the data. Unfortunately, there is still a need that must be readdressed every decade or so (5). Reports in our scientific journals today still have problems with statistics, and the information on use and interpretation is as timely today as it was then. As we concluded in that 1973 paper, "statistical tests can provide specific estimates of errors upon which judgments can be made, but they are not a substitute for judgments" (6).

References

1. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974;20:825–33.
2. Westgard JO, deVos DJ, Hunt MR, Quam EF, Carey RN, Garber CC. Method evaluation. Bellaire, TX: American Society for Medical Technology. 1978.
3. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 1981;27:493–501.
4. Westgard JO, Groth T, Aronsson T, Falk H, deVerdier C-H. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. *Clin Chem* 1977;23:1857–67.
5. Westgard JO. Points of care in using statistics in method comparison studies [Editorial]. *Clin Chem* 1998;44:2240–2.
6. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. *Clin Chem* 1973;19:49–57.

¹ University of Wisconsin CLS Program, Madison, WI.

² This paper has been cited more than 330 times since publication.

* Address correspondence to this author at: James O. Westgard, University of Wisconsin CLS Program, 1300 University Avenue, Madison, WI 53706. E-mail james@westgard.com.

Previously published online at DOI: 10.1373/clinchem.2007.094060