STATISTICAL ANALYSIS

The Horwitz Ratio (HorRat): A Useful Index of Method Performance with Respect to Precision

WILLIAM HORWITZ[1] and RICHARD ALBERT[2]
U.S. Food and Drug Administration

The Horwitz ratio (HorRat) is a normalized performance parameter indicating the acceptability of methods of analysis with respect to among-laboratory precision (reproducibility). It is the ratio of the observed relative standard deviation among laboratories calculated from the actual performance data, RSD_R (%), to the corresponding predicted relative standard deviation calculated from the Horwitz equation $PRSD_R$ (%) = $2C^{-0.15}$, where C is the concentration found or added, expressed as a mass fraction. It is more or less independent of analyte, matrix, method, and time of publication (as a surrogate for the state of the art of analytical chemistry). It is now one of the acceptability criteria for many of the recently adopted chemical methods of analysis of AOAC INTERNATIONAL, the European Union, and other European organizations dealing with food analysis (e.g., European Committee for Standardization and Nordic Analytical Committee). The origin and applications of the formula are described. Consistent deviations from the ratio on the low side (values <0.5) may indicate unreported averaging or excellent training and experience; consistent deviations on the high side (values >2) may indicate inhomogeneity of the test samples, need for further method optimization or training, operating below the limit of determination, or an unsatisfactory method.

The Horwitz ratio (HorRat) or value is a simple performance parameter that reflects the acceptability of a chemical method of analysis with respect to precision. [*Note*: Per Lea of the Norwegian Food Institute suggested the format "HorRat" based on its origin as the Horwitz ratio. Although it was initially developed from the among-laboratory relative standard deviation (RSD_R), it is also applied to within-laboratory relative standard deviation, RSD_r, with less reliability. $HorRat_R$ and $HorRat_r$ are used to differentiate these 2 possibilities.] HorRat provides an index of interlaboratory precision as a function of concentration over the entire field of practical analytical chemistry. It is the ratio of the RSD among laboratories (reproducibility), RSD_R, in percent, as calculated from an interlaboratory method trial, to the predicted RSD, $PRSD_R$, in percent as calculated by a simple exponential equation ($PRSD_R = 2C^{-0.15}$) from the mean concentration, *C*, found or added, expressed as a mass fraction. Surprisingly, this ratio is more or less independent of analyte, method, matrix, and date of publication (as a surrogate for the state of analytical technology). The within-laboratory RSD (repeatability), RSD_r, is typically one-half to two-thirds of the among-laboratory RSD.

When an analytical chemist performs an analysis by applying a method of analysis to a test portion of a material, he/she obtains a concentration value for the analyte. This value is merely an estimate of the true value. Repeated application of the procedure provides a population of estimates with a mean, which is usually taken as the "best estimate" of the actual analyte content of the material, supplemented by a dispersion of values about that mean. This dispersion of estimates is best represented by the standard deviation. Use of the mean and the standard deviation permits mathematical manipulation of the distribution of values.

Every time a variable is changed in performing or repeating the analysis, a new distribution of results is obtained, even if the same method is being followed. The basic and tightest distribution is that provided by the single analyst working with a single instrument. If the analyst uses a second instrument, a different distribution is obtained, although it probably will not differ by much. Each analyst using a specific instrument has a specific distribution, and the pool of analysts who may use any instrument in a single laboratory leads to an overall "within-laboratory" standard deviation characteristic of that laboratory. Another laboratory will have a similar but different distribution. If we sample sufficient laboratories, we can obtain an overall distribution that can be used to characterize any laboratory, any operator, and any instrument, in general. It will be somewhat larger than that of any single laboratory and any single operator utilizing any single instrument, but it will provide a satisfactory representation of expected within-laboratory variability for any laboratory conducting that analysis on a specific analyte and matrix, in the absence of more specific information. But sampling many laboratories performing that specific analysis

Received January 20, 2006. Accepted by AP March 2, 2006.
[1] Retired. Corresponding author's present address: 14800 Pennfield Circle, #409, Silver Spring, MD 20906-1579; e-mail: wxhor@aol.com
[2] Retired. Present address: 1 Lennon Ct, Unit #1, South Boston, MA 02127.

provides an additional important piece of information: how laboratories differ from each other in estimating the same analyte in the same material by the same method. This is the type of variability that is obtained when several laboratories analyze "identical" portions of the same material, as may be the case in a "dispute" situation. The equation for the between-laboratory precision, s_R, is simply the sum of the within-laboratory precision, s_r, and the "pure" between laboratory precision, s_L, expressed as variances:

$$S_R^2 = S_L^2 + S_r^2$$

The 2 variances S_R^2 and S_r^2 are obtained directly from an interlaboratory study and are given the special names of reproducibility variance and repeatability variance. Variances are important to the statistician because they are the terms used in mathematical equations describing distributions, whereas the standard deviations are useful to the chemist because they are in the same units as the reported results. The repeatability variance or the squared standard deviation is useful to the statistician because a group of them from the same population follows the Chi-square distribution just as a group of means from the same population typically follows a normal distribution.

Chemists, however, usually use the RSD for their assessment of the result because it often is either constant or linear with concentration over a reasonable range of concentration. The typical differences for a consecutive series of titrations may be a few tenths of a percent, and the differences among a series of absorbance or chromatographic peak measurements may be about 1%. But, as major changes in measurement conditions occur, as, for example, in going from a single analyst to multiple analysts in the same laboratory or to other analysts in a different laboratory, the RSD may double.

Another important factor that results in major changes in RSD is a concentration effect. As analytical chemistry explored lower and lower concentrations, RSD appeared to increase exponentially.

Variability and Validity of Concentration Estimates

We are dealing with the variability and validity of chemical concentration estimates. The chemist does not know the true concentration of a target analyte. By repeating a suitable analytical procedure a number of times, the chemist seeks to bracket the true concentration and thus obtain a concentration estimate—a number expressed in the appropriate concentration units, such a parts per million or grams per milliliter. The repetitions must be done "independently," which means in such a way that a prior estimate does not influence or bias the current estimate being made. Eventually, the chemist is confronted with a set of *N* individual concentration estimates of the true analyte concentration. The tighter these estimates are, the better the situation is. The closer these *N* estimates are to one another—that is, the narrower the spread, the better is the precision. The typical measure of the spread of a set of values is the standard deviation, or its square, the variance: variance = (standard deviation)^2.

Although there may be collateral problems of homogeneity, stability, and sampling that eventually must be addressed by the chemist, the overriding concern is with concentration. The true concentration is estimated as the simple arithmetic average of the *N* concentration estimates. The bias of this estimate is the difference (positive or negative or zero) between this average and the true concentration. For very large *N*, this bias is the individual chemist's bias—also known as the "individual laboratory bias"—for the target analyte at the true concentration. The "within-laboratory variance" is simply the variance of the individual chemist's *N* concentration estimates. If there are several chemists involved in that determination, we can assume that one of them is representative of all, or we can have all of them perform the analysis on identical test samples and "pool" or average their results.

If *L* different laboratories each perform a single analysis of the same analyte at the same true concentration in the same matrix, these *L* estimates can be used to calculate an estimate for the reproducibility variance, RSD_R. For very large *L*, the variance of these *L* estimates is, by definition, the reproducibility variance. It is a hybrid of the within-laboratory variances and the variances of the *L* individual laboratory biases. If *L* is not very large—for example, only 8 to 15 laboratories contributing values—and if there are multiple estimates from each laboratory, statistical manipulation will permit one to estimate the crucial reproducibility variance which is a hybrid of the within-laboratory variances and the variance of the *L* individual laboratory biases. Note the interesting fact that in obtaining the reproducibility variance, we have transformed the individual unknown biases of the individual laboratories to a statistically manipulatable variance.

Although it had been observed that high concentrations had low RSD_R, and low concentrations had high values, this impression had not been developed quantitatively. This change in RSD_R with concentration could be represented empirically by what is now known as the Horwitz equation

$$RSD_R, \% = 2C^{-0.15}$$

or its equivalent,

$$S_R = 0.02C^{0.85}$$

where the concentration, C, is expressed as a dimensionless mass fraction (both the numerator and denominator are expressed in the same units). It was developed as a byproduct of an attempt to compare the interlaboratory reliability of the newer, simpler, and faster spectrophotometric and chromatographic procedures that were replacing the classical gravimetric and volumetric methods used in the analysis of pharmaceutical preparations (1). It became apparent that the precision of drug analysis as examined in multiple laboratories was a function of concentration only. The most

startling conclusion of the initial paper, which was presented at an "Annual University of Wisconsin Conference on Analysis" in 1977, was that the simpler, faster, and more elegant modern instrumental methods were no more accurate and precise than the lengthy, tedious, and complex classical volumetric and gravimetric procedures that they replaced. This simple, elegant concept was then extended to other areas of analytical chemistry, particularly those used by regulatory agencies, where the reliability of analytical methods is validated by interlaboratory testing before being placed into routine use. The value of the equation was recognized by the International Union of Pure and Applied Chemistry (IUPAC; 2) and by European analytical chemists, who incorporated the equation into legislation dealing with numerous regulatory limits of the European Union (EU). It is currently used as the primary screen for acceptance of methods by AOAC INTERNATIONAL, the independent professional organization that provides the validated methods of analysis used by the U.S. Food and Drug Administration (FDA) and other U.S. government agencies for enforcement purposes (3). The Analytical Methods Committee of the Royal Society of Chemistry (UK) published a summary description of the function and examples of its utility (4).

The basic observation was that the *among-laboratory* precision, expressed as RSD_R, doubled for every decrease of 2 orders of magnitude in concentration, C, expressed as a mass fraction. The original declaration (5) was "[W]e have found that the reliability of the results of our analytical operations can be summarized, in an oversimplified fashion to be sure, by plotting a mean coefficient of variation (CV) found, expressed as powers of 2, against the concentration measured, expressed in powers of 10, as shown in Figure 1." This statement was illustrated by 2 symmetrical expanding curves (±) of the RSD (or CV), expressed as powers of 2, beginning with pure materials at a concentration of 100% (where $C = 1$; RSD_R = 2%) placed where ordinarily $x = 0$ occurs, and extending through $C = 10^{-12}$ toward an unexpressed $x = 0$ ($C = 0$) to the right, in decreasing concentration multiples of 10. This curve (Figure 1) was first presented in a paper discussing quality control of trace constituents in cooperation with LaVerne Kamps, a pesticide residue specialist, and Kenneth Boyer, a trace element chemist (5). It was originally expressed in the form as presented by the statistician, Jung-Keun Lee:

$$RSD_R = 2^{(1 - 0.5 \log C)}$$

Thompson (6) transformed it into the equivalent compact forms given in Table 1. This curve, known as the Horwitz Horn or the Horwitz Horror, depending on the degree of belief in it, was characterized by Hall and Selinger as "one of the most intriguing relationships in modern analytical chemistry" (7). The Analytical Division of the Royal Society of Chemistry presented its Robert Boyle Medal to W. Horwitz in 2000 in recognition of this contribution to analytical chemistry, an honor previously given to the late I.M. Kolthoff, the internationally famous head of the Analytical Chemistry Department of the University of Minnesota.

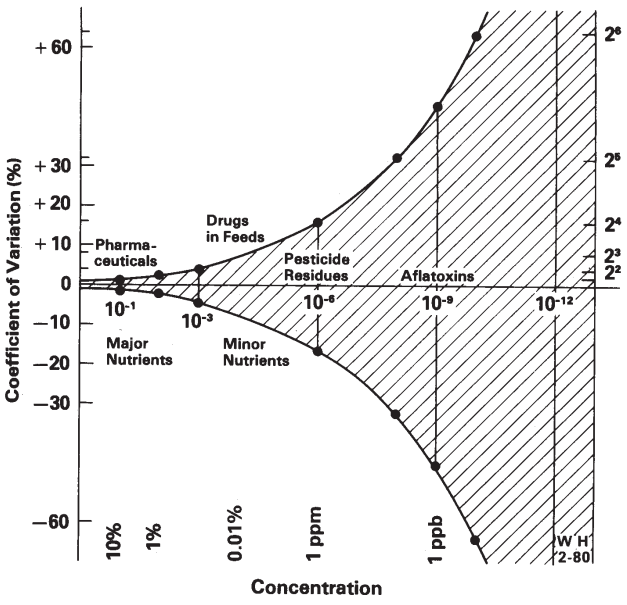

Figure 1. Original curve.

The Horwitz Equation

The equation can be expressed in several equivalent ways as shown in Table 1. Here the term "coefficient of variation" is assigned to the fractional expression and "relative standard deviation" to the percentage expression of the standard deviation divided by the mean, a useful distinction suggested by Taylor (8):

$$CV = RSD_R(\%)/100 = s_R/x_C$$

$$RSD_R\ (\%) = 100 \times CV$$

where CV = coefficient of variation (expressed as a fraction), s_R = standard deviation among-laboratories (expressed in the same units as the data), RSD_R = relative standard deviation among-laboratories (expressed as a percent), and x_C = mean or average concentration expressed as a mass fraction (the numerator and denominator are expressed in the same units; m/m or kg/kg = 1).

Table 1. Alternative expressions for the Horwitz equation

Expression, units	Mathematical	Spreadsheet
Standard deviation, s_R (m)	$= 0.02C^{0.85}$	= 0.02 * C^0.8495
Relative standard deviation, RSD_R, % (m/m)	$= 2C^{-0.15}$	= (s_R/C) * 100 = 2 * C^(–0.1505)
Coefficient of variation, CV, fraction (m/m)	$= 0.02C^{-0.15}$	= 0.02C^(–0.1505)

(*Note:* Analysts often do not distinguish between CV and RSD and will use the terms interchangeably, but the distinction is obvious from the position of the decimal point.)

The only variable in the equivalent expressions in Table 1 is concentration, which must be expressed as a mass fraction. If the concentration were expressed in moles, the equation would obviously be more complicated and it would no longer be independent of analyte because it would require the formula weight. Use of the concentration factor, g/mL, for aqueous solutions is acceptable, based on the assumption of 1 mL = 1 g.

These equivalent equations were developed by examining the results from interlaboratory collaborative studies, each usually containing data from about 8 laboratories analyzing at least 5 identical typical materials, usually in duplicate. The results, therefore, are expressed in terms of interlaboratory precision. The equations merely state that the interlaboratory precision is an exponential function of the concentration only. It is interesting to note that the (negative) exponential constant (0.1505) is (1/2) * $\log_{10}$ 2, but this fact has no theoretical significance. Lately, we have been using the equation with modified limits for single laboratory validation (SLV), an application initiated by EU regulations.

Which form of the equation should be used is a matter of personal preference or background. Statisticians typically express variability in terms of variance (the square of the standard deviation, s^2), which has theoretical significance with respect to statistical distributions. Chemists, on the other hand, usually use the RSD, which is often a constant or a linear function over a convenient concentration range. In fact, one can usually determine if data has been analyzed by a chemist or by a statistician from whether the variability is discussed in terms of RSD or in terms of variance.

The initial equation was based on the examination of the data from about 100 materials, primarily pharmaceutical dosage forms, examined by multiple laboratories in interlaboratory studies. The data base eventually encompassed over 100 000 materials (matrixes) and of the order of 10 000 interlaboratory studies before formal data gathering ceased. Practically every current published interlaboratory study in the food, drug, and agricultural area is interpreted in terms of the HorRat value. The accumulation of a formal database has been abandoned in view of the general acceptance of this interpretation of interlaboratory precision. Current efforts are devoted to examining clinical proficiency data.

The Horwitz Ratio or HorRat Value

The Horwitz ratio or HorRat value is the ratio of the RSD_R, in percent, calculated from the data, to the RSD predicted from the Horwitz equation, $PRSD_R$, thus:

$$\text{HorRat} = \frac{\text{RSD}_R}{\text{PRSD}_R}$$

Table 2. Some common concentrations and their equivalents as mass fractions

Concentration, common units	Mass fraction (spreadsheet notation)
100%, Pure substance	1.00 (E–0)
10%	0.10 (E–1)
1%	0.01 (E–2)
0.1%, mg/g	0.001 (E–3)
0.01%	0.0001 (E–4)
1 ppm, μg/g; mg/kg	0.000 001 (E–6)
1 ppb, ng/g; μg/kg	0.000 000 001 (E–9)

This equation transforms the RSD found to a fraction of the RSD expected from application of the exponential equation to the concentration. The precision found is, thereby, expressed as a fraction of the precision as calculated from the experimental concentration estimates. It equals 1 for exact correspondence. The precision is better than expected if the ratio is less than 1, and poorer if greater than 1. The empirical acceptable range is 0.5 to 2.0. This ratio has been expressed equivalently as "HORRAT," "Horrat," or, as suggested by Lea (9), in a form suggesting its origin, as "HorRat."

Concentration

The concentration of the analyte, *C*, must be expressed as a mass fraction, where both the numerator and denominator are in the same mass units. In this case, the base unit is 1. Percentages are mass units with a base of 100. Some common, convenient concentrations transformed to mass fractions for use with the equation are given in Table 2.

The "E" notation is for use with computer spreadsheets, where the E stands for "exponent." Excel, for example, will recognize an E followed by a number as "10 to the power indicated by the number, positive or negative," as shown in Table 2. If a very small or very large number is written out for text, however, group the zeros in units of 3 for ease in counting to check the correctness of the number of zeros. Note in Table 2 for 1 ng/g, 8 zeros are required before the final 1, and 2 spaces are used as separators to make the number easily readable. The space is not necessary with 4 digits (0.0001). Use of spaces in Excel formulas, however, results in an error message requesting approval to change to an equivalent revised statement that omits them for calculation purposes. Avoid the expressions "parts per billion" and "parts per trillion," as they have different meanings to U.S. and European readers, although some EU documents published in their *Official Journal* use the expression "parts per billion" with the U.S. meaning.

Calculation Check

A convenient check for computer programs and programmable calculators is to insert a concentration of 1 μg/g (ppm) into the equation as $C = 10^{-6}$ (mass fraction =

0.000 001, or E-6), then $s_R = 0.16 * 10^{-6}$ and $RSD_R = 16\%$, as follows:

$$s_R = 0.02*C\text{^}0.8495 = (0.02 * 10^{-6}) \text{ ^ } (0.8495) =$$

$$0.02 * 7.998 * 10^{-6} = 1.60 * 10^{-7} = 0.16 * 10^{-6}$$

or

$$RSD_R\ (\%) = (s_R/C) * 100 = 0.02*C\text{^}0.8495 * C\text{^}(-1) * 10^2$$

$$= 0.02 * 10^2 * C \text{ ^ } (-0.1505) = 2*C\text{^}(-0.1505)$$

$$= 2 * (10^{-6}) \text{ ^ } (-0.1505)$$

$$= 2 * 8.00 = 16\%$$

The curve has a finite beginning with pure substances, where $C = 1$ as a mass fraction (corresponding to 100% as a percentage), and $RSD_R = 2\%$. Although the equation can be used with suitable adjustments with any concentration units, the use of mass fraction permits covering the entire range from pure materials to the ultratrace region below 10^{-6} (μg/g, ppm) smoothly without a shift in the base or viewpoint. The mass fraction is also related to the SI unit for concentration (amount of substance).

Origin

The basic data were supplied from numerous studies of drug dosage forms and food composition examinations published in the *Journal of AOAC INTERNATIONAL*. The analytes were originally determined by classical gravimetric and volumetric methods by the FDA, state regulatory agencies, and the regulated industry to control adulteration and misbranding of foods, drugs, and agricultural materials. The basic food composition standards and nutrient tables throughout the world were developed using such methods. Simultaneously, the State Chemists in the United States were perfecting the methods for the critical plant food elements, nitrogen, phosphorus, and potassium, that state laws required to be declared on labels of fertilizer containers, and for the animal feed constituents, moisture, ash, protein, fat, carbohydrates, fiber, and nutrient minerals. Most of the specific analytes were initially determined by analytical methods based upon stoichiometric reactions in the micro and macro concentration ranges of about 0.1 to 100% ($C = 0.001 – 1.0$). Not until vitamins, hormones, tissue and pesticide residues, and trace elements entered into the regulatory picture, together with the introduction of the spectrophotometer, did the analytical chemist venture routinely into the concentration region below about 0.1%. After the passage of the pesticide and food additive amendments to the Federal Food, Drug, and Cosmetic Act, with requirements for residue analysis, and with the application of chromatographic separations, electronic detectors, and instrumentation with computer control, did analysis in the mg/kg (E-6), μg/kg (E-9), and even ng/kg (E-12) became routine. Regardless of the concentration range, however, before effective regulation could be operative, enforcement officials had to determine how much allowance had to be made for typical variability between laboratories before concluding that goods were violative. The necessity for answering this question led to the formation of the Association of Official Agricultural Chemists (AOAC) in 1884 (10).

Range

The initial examination of data was confined to the higher concentration region of about 0.1 to 100%, about 3 orders of magnitude. The upper limit is the automatic restraint at a ceiling concentration of 100% or a mass fraction of 1.0. Although this restraint exists for the "true value," it does not exist for experimental estimates of the true value because one cannot achieve a practical average value of 100% unless some of the experimental measurements are greater than 100% to balance those values that are less than 100%. This fact is recognized in compendial specifications which often require that the parent drug form have a strength expressed in a form similar to 98–102%. The same argument applies at the other end of the concentration scale: one cannot have a true 0 concentration unless some of the signal observations, transformed to concentrations, are negative to balance the positive observations. Discarding negative concentration values, truncating or "censoring" those values below a limit of detection (LOD) or limit of determination results in a mean that is biased upward. Reporting values as half or some other fraction of the detection limit, or "less than the detection/determination limit" depreciates the contribution the actual value makes to the database. Therefore, the recommendation with regard to observations outside the conventional scale (greater than 100% or less than the LOD) is to record the actual value of the signal (positive, negative, or zero) transformed to a concentration to permit the operation of the law of averaging to provide a reasonable estimate of the "true value" (11).

Derivation

After a considerable amount of data had been accumulated and the multiplicative nature of the function became obvious, a practical ("heuristic") derivation of the Horwitz function was developed (12). It assumed that the fractional change in standard deviation in the interlaboratory environment, dS/S, was proportional to the fractional change in concentration, dC/C, thus:

$$dS/S = k\ dC/C$$

Integrating this function gives:

$$\log S = k \log C + \text{a constant of integration (Q)}$$

Inserting 2 known pairs of values will provide values for the 2 constants, k and Q:

For C = 0.01 and S = 0.0004 (i.e., RSD = 4%),

and C = 0.0001 and S = 0.000008 (i.e., RSD = 8%),

$$k = 0.8495 \text{ and } Q = 1.6991$$

Therefore,

$$\log S = 0.8495 \log C + 1.6991$$

$$\log S = C^{0.8495} + 1.6991 = C^{0.8495} + \log 50$$

$$S = (1/50)*C\text{^}0.85 = 0.02\ C\text{^}0.85$$

and in the form of relative standard deviation

$$RSD = S \times 100 / C = 2\ C^{-0.15} = 2*C\text{^}(-0.15)$$

Because practical units were not introduced in the derivation, all units must be the same, i.e., as mass fractions, where both the numerator and denominator are in the same units, so they cancel out.

Practical Example

Consider a pure metal that is analyzed frequently because it is the reference and quality control standard for that analysis. A laboratory may provide a series of analyses similar to the following as a result of its daily quality control checks: Day 1, 99.7% metal; Day 2, 100.2% metal; Day 3, 100.2% metal; Day 4, 99.6% metal; Day 5, 99.9% metal; Day 6, 100.3% metal; Day 7, 99.8% metal; Day 8, 99.9% metal; Day 9, 99.8% metal; and Day 10, 100.1% metal.

Average C = 99.95%, or as a mass fraction, 0.9995

Standard deviation (SD) = 0.237%, RSD = 0.24%

First note that these are *intralaboratory* values. The HorRat value was developed from *interlaboratory* data, and there exists a difference of almost a factor of 2 in relative variability between these 2 conditions because by definition laboratory-to-laboratory differences are not included when calculating within-laboratory variability. As a matter of fact, one of the assumptions of the statistical model used for developing the precision expressions for interlaboratory studies is that all of the laboratories operate with equal within-laboratory precision. The daily values are quite close together because the method is rather simple, consisting of dissolving the metal in acid, buffering the solution, diluting, introducing the solution into an atomic absorption (AA) or inductively coupled plasma (ICP) spectrophotometer for the measurement, and usually an automatic calculation.

To calculate the HorRat value, first calculate the "predicted RSD_R," $PRSD_R$, remembering to transform the concentration in percent to a mass fraction by dividing by 100 [C = 99.5(%)/100 = 0.9995(m/m)]:

$$PRSD_R = 2C^{-0.15} = 2 * C \text{ ^ } (-0.15)$$

$$= 2 * (0.9995) \text{ ^ } (-0.15) = 2 * 1.00 = 2.00\ (\%)$$

$$HorRat_r = RSD_r/PRSD_R = 0.24/2.00 = 0.12$$

This HorRat value is considerably better than predicted, primarily because this is intralaboratory data, indicated by the use of the lower case subscript "r." The chemical operations of weighing and dilution introduce negligible uncertainty into the operation. The test sample is pure and homogeneous. The calibration curve has practically no random deviations, and the calculations are performed automatically. Note, however, the variability spread of 0.6% in the 10 independent daily determinations that provides the estimate for the SD, and, in this case the RSD, of about 0.24%. [*Note:* Before the days of computers and calculators, quality control inspectors applied the rule that the SD could be approximated for 10 or fewer values by dividing the range by the square root of the number of values. Applying the rule here gives 0.6 (range) / $\sqrt{10}$ = 0.6 / 3.16 = 0.19 as a rough estimate of SD, deviating by about 20%.]

Laboratory-to-Laboratory Variability, s_L

The primary reason for the apparently low value is that the previous data are from a single laboratory, which does not include the biases introduced from the different environments, operations, instruments, and standards of various laboratories. Every laboratory operates under different conditions of temperature, light, humidity, power supply, atmospheric pressure, instrument adjustments, sources of reagents and supplies, and personnel operations and supervision. This variability in laboratory operations results in an irreducible laboratory-to-laboratory variability characterized by a standard deviation, s_L. For L very large, this is the standard deviation of the L average concentration estimates when each laboratory uses the average of an indefinitely large number of concentration estimates. For a single laboratory, $s_L = 0$ or is not defined because there is no other laboratory from which to be different.

Within-Laboratory Variability (Repeatability), s_r

For every procedure, every laboratory (and in fact every analyst and every instrument) has its own characteristic within-laboratory variability, s_i, where i is the laboratory or item number (i = 1, 2, 3,...l), but for the purpose of calculating method performance parameters it is necessary to assume that all laboratories perform with equal variability. (A Cochran or similar outlier test can serve to verify whether or not this assumption is correct.) This is a fairly good assumption, especially with standardized methods. In the review of method performance, it was found that the within-laboratory variability was in general fairly constant at about one-half to two-thirds the among-laboratory variability. Within-laboratory variability is called repeatability. The term "repeatability" is used both as a noun (as in the previous sentence) and as an adjective, i.e., repeatability standard deviation or repeatability variance.

When all the work is confined to a single laboratory, but utilizing different analysts and possibly different instruments and different times, the increased variability is characterized as "intermediate" precision, or sometimes as "within-laboratory reproducibility." Such variability is

between the limiting extremes of repeatability (within-laboratory) and reproducibility (among-laboratories).

The mean and SD for single laboratory parameters are easily calculated by the basic formulae for these parameters.

Between-Laboratory Variability (Reproducibility), s_R

When several sources contribute to an overall variability, the simple equations no longer apply because the statistical formulae for the individual sources of variability apply to variances, not to standard deviations. The between-laboratory variance is the sum of the within-laboratory variance and the "pure" between-laboratory variance. The interesting thing abut this "pure" between-laboratory variance is that when viewed from an individual laboratory perspective it is a bias—its difference from the "true" or accepted value. But when viewed from the perspective of the overall study, the composite of the individual differences can be handled as a "true" or "pure" between-laboratory variance. The sum of this composite variance and the pooled within-laboratory variance constitute the overall reproducibility variance.

The experimental data from the interlaboratory study must be disassembled by the statistical technique of analysis of variance to provide the reported values for s_r and s_R and the corresponding relative standard deviations. Detailed instructions are available in the Steiner portion of the AOAC Statistical Manual and in Wernimont and Spendley's "Use of Statistics to Develop and Evaluate Analytical Methods" (13). An Excel-based software program developed by and available (free) from Joanna Lynch of Cornell University (14) merely requires inputting the data and the program will provide the required output, including outlier removal.

The RSD between laboratories as a function of concentration is the primary reference value that determines the acceptability of methods of analysis. This was verified by examining the results from numerous collaborative studies compiled in a series of papers, classified by analytes, matrixes, or techniques, all of which supported a distribution of acceptable HorRat values ranging from 0.5 to 2.0 (i.e., 1.0 divided by and multiplied by 2). These are not absolute limits because transgressions are occasionally permitted in both directions. For example, it is often found in a series of materials of diminishing concentrations that the lowest level shows a HorRat value over 2, and all of the others are less than 2. Such a pattern permits the assignment of the lowest value as near or close to the limit of reliable measurement.

Table 3 is a compilation of the characterizations of method performance with respect to precision on a consistent basis, permitting comparisons that have been examined by Horwitz and his collaborators and published in the *Journal of AOAC INTERNATIONAL*, unless otherwise indicated.

As individual interlaboratory studies are located in the literature, primarily in the food and clinical chemistry field, they have been recalculated to HorRat values. In practically all cases of studies pronounced successful by the authors, the HorRat values correspond to the 0.5–2 limits. These isolated studies have not been accumulated into a database. Most

Table 3. References to databases used to validate the HorRat function

Product	Method, matrix, or analyte	*J. AOAC. Int.* reference
Drugs and pharmaceuticals	Chromatographic separation/ spectrophotometric measurement	**67**, 81–90 (1984)
	Gas-liquid chromatography	**67**, 648–652 (1984)
	Automated	**68**, 112–121 (1985)
	High-pressure liquid chromatography	**68**, 191–198 (1985)
	Miscellaneous	**68**, 830–838 (1985)
	Gravimetric and titrimetric	**71**, 619–635 (1988)
Foods	Dairy products	**72**, 784–806 (1989)
	Nutrition labeling/major nutrients	**73**, 661–680 (1990)
	Nutrition labeling/macro elements: Ca, Mg, P, K, Na, S	**76**, 227–239 (1993)
	Mycotoxins	**76**, 461–491 (1993)
	Pesticide residues	**83**, 399–406 (2000)
Pesticide formulations	Commercial products	**74**,718–744 (1991)
Standard reference materials (SRMs)	Certified reference materials	*ACS Monograph* 445 (1991)
Polychlorinated contaminants (biphenyls, dioxins, furans)	Various	**79**, 589–621 (1996)
Minerals	Geochemical analysis	*Fresenius J. Anal. Chem.* **351**, 507–513 (1995)
Derivation of curve		*Anal. Chem.* **69**, 789–790 (1997)
Fertilizers		Unpublished

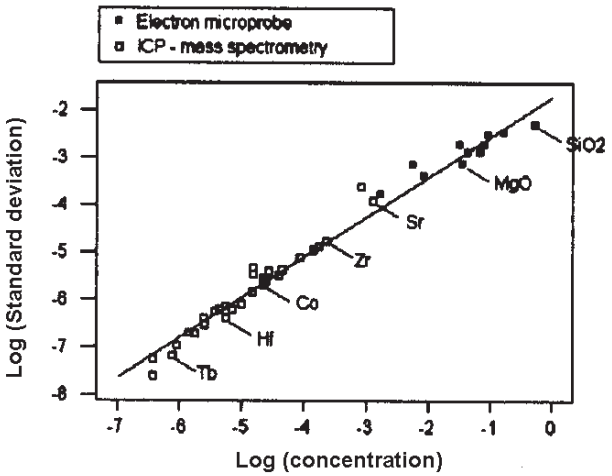

Figure 2. Conformity of geochemical data to linear form of the Horwitz equation. (Used with permission of the Royal Society of Chemistry.)

interlaboratory studies that are published in the food field are now interpreted by their authors in terms of HorRat values.

The Google search engine provided 1420 references to the use of "HorRat," not all of which are relevant to the present context. Some examples of unusual studies (in terms of the nature of the analytes or size) that utilized the HorRat value for interpretation include the New Zealand marine biotoxin program (15), Canadian mercury in coal (16), EU cacao butter equivalents in cacao butter (17), polymerase chain reaction (PCR) for transgenic plant material (18), and alkenone abundance in marine sediments (19). The HorRat value is particularly useful for determining the acceptability or failure of the study with respect to precision. But, precision/variability is merely one of the many tools that the chemist must consider in determining the suitability of a given analytical method.

Examples of the utility of the HorRat value under extreme conditions can be found in articles describing proficiency schemes for geochemistry (20, 21). In one case, a volcanic glass was analyzed for about 40 elements by laser ablation-ICP mass spectrometry (LA-ICP/MS) and electron probe over 6 orders of magnitude in concentrations. It concluded, "This test material, and analytical method employed, could hardly be more remote from the materials and methods that provided the original Horwitz data, especially as the mass of material analyzed in LA-ICP/MS is only a few micrograms. The data (Figure 2) conforms with the Horwitz function to a remarkable degree." (21).

Limitations

The Horwitz curve does not apply to empirical analytes, i.e., those that are method-dependent, such as moisture, ash, fiber, and similar method-defined analytes, whose composition is ill defined and whose concentration estimate depends on the specific details of the method. The "true value" is whatever is found by the specific method. The results from application of such methods often exhibit very small variability within a single laboratory but high variability among different laboratories. Nor does it apply to indefinite analytes, such as enzymes, polymers, and many biomolecules, or to quality factors, physical properties, or physical methods such as color, density, viscosity, or drained weight, whose results are not ordinarily expressible as concentrations. Physical property measurements transformed to concentrations (e.g., alcohol determined by specific gravity or refractive index) often exhibit very tight measurements and correspondingly low HorRat values. Analytes consisting of homologous series, e.g., fatty acids in fats and polymer components, may show large reproducibilities, i.e., poor precision as compared with the historic norms exemplified by the Horwitz curve.

The general curve of found variability plotted against concentration appears to dip a bit at both ends—starting near 10% upward and at about 10 μg/g downward—where results are usually better than predicted.

At the high concentration end, the better-than-predicted values may be the result of a mathematical idiosyncrasy applied to paired values. High variability for moisture in foods in the 5–20% region can be transformed to low variability if reported as solids in the 80–95% region. Similarly low absolute variability of the composition-defining egg components, phosphorus and nitrogen, in sugared yolks can become high variability when reported on a sugar-free, moisture-free basis.

At the low concentration end, the better-than-predicted results for toxicologically important industrial contaminants, such as dioxins and polychlorinated multiring compounds in environmental substrates, can be explained on the basis of the standardized method of analysis used. The U.S. Environmental Protection Agency (EPA) and similar methods that are used for the analysis have been constructed with isotopic (^{13}C) internal standards and numerous internal control points, so that errors and deviations in laboratory operations signal out-of-control values before they become real data. However, commercial quotations for this analysis are of the order of $1000 per analytical value with a 2-week reporting time period. Concentrated training and continuous monitoring through proficiency exercises can minimize variability between laboratories as will be shown below.

Lower limit.—There exists a rationale for predicting the existence of a lower limit of quantification (LOQ) concentration as shown by Thompson and Lowthian (22). If we assume that the LOD is the blank value plus 3 times the standard deviation of the blank value, the point of intersection of the Horwitz equation and this limit is found by solving the equation (which assumes the blank value is 0):

$$C_{LD} / 3 = s_R = 0.02C^{0.85}$$

The point of intersection occurs at $C = 10^{-8.12} = 7.6 * 10^{-9}$, or about 8 ng/g (ppb). At this point, RSD_R from the Horwitz curve is about 33%. Therefore, interlaboratory studies that indicate an RSD_R greater than about 30% are unsatisfactory.

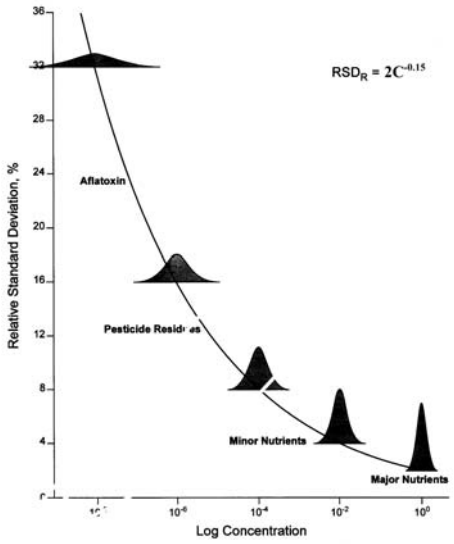

Figure 3. Interlaboratory coefficient of variation as a function of concentration.

In fact, as pointed out in the mycotoxins paper (23), as the concentration decreases below the level that generates this variability (about 10^{-8}), the number or percentage of false negatives and false positives reported in interlaboratory studies increases, both in theory and in actual practice. Therefore, method precision parameters above about 30% are invalid or, at least, call for special attention and treatment, including extending the number of concentration estimates, which serves to reduce the variability but at a cost of time and resources.

The reason for this is shown in Figure 3, where the curve from the Horwitz equation is inverted on the $\log_{10}$ x-axis; 100% is to the right and the concentration decreases to the left. Superimposed on the curve are the "theoretical" individual distributions at various concentration levels. Note how the distributions flatten out with decreasing concentration, attesting to the larger standard deviations and RSD. At the lowest concentration shown, the distribution overlaps the y-axis. The values in this left portion of this lowest distribution are the false negatives. Their relative and absolute numbers increases as the concentration decreases.

Performance at the High and Low Concentration Extremes

Both the upper and lower limits on precision can be circumvented if a large investment is made in training and practice prior to performing the interlaboratory study. This was demonstrated by the EU, which made such an investment in the development and performance of the methods for mycotoxins in foods that resulted in interlaboratory studies (24) with $HorRat_R$ values considerably less than 1, leading to the acceptance of a number of AOAC Official Methods. The lowest accepted level was 0.05 ng/g (50E-12) for aflatoxin M_1 in milk. This concentration value represents the lowest value at which a successful interlaboratory method performance study has ever been performed. The EU has indicated a desire to lower the validated level to 0.01 ng/g.

At the high concentration levels of solids and fat in milk, the Federal Milk Marketing Administration (FMMA) of the U.S. Department of Agriculture (USDA) in collaboration with the Dairy Department of Cornell University made a similar investment in standardization and instruction in the performance of the methods for solids and fat in milk (25). The resulting $HorRat_R$ values of 0.1–0.4 are equivalent to almost eliminating between-laboratory variability. In this case, large commercial transactions are based upon agreed values for milk components provided by different laboratories that justify the continuing large investment in training and quality control. As stated by Lynch et al. (25), "The success of the USDA FMMA program illustrates the ability of economics to impact on method testing performance."

The exceptional results from the EU mycotoxin studies apparently formed the basis for the conclusions drawn by Thompson (26) that assign a constant RSD_R of 22% to all concentrations below about 0.12 ppm. This recommendation has been incorporated into several recent EU directives. It is doubtful if those conclusions can be transferred to other areas of analysis that have not been the beneficiaries of the large investment in initial method development and training, and subsequent continuous mandatory proficiency and quality assurance programs. The mycotoxin data "estimated by a robust procedure, shows a slope of unity on the plot and corresponds to a relationship $\sigma_R = 0.22C$. . . This line intersects the Horwitz function at a concentration of $10^{-6.92}$, about 1.2×10^{-7} or 120 ppb" (26).

Thompson drew 2 conclusions from his examination of this data: As reported by Horwitz, estimates of the reproducibility standard deviation, σ_R, from mycotoxin and low-level pesticide residue data were consistently lower than the basic function, designated σ_H, and on the basis of the EU mycotoxin studies concluded that a trend exists toward better precision. Further, the following function was suggested "as a contemporary model for reproducibility standard deviation":

$$\sigma_R = 0.22C \text{ if } C < 1.2 \times 10^{-7}$$

$$\sigma_R = 0.02C^{-0.8495} \text{ if } 1.2 \times 10^{-7} \le C \le 0.138$$

$$\sigma_R = 0.01C^{0.5} \text{ if } C > 0.138$$

Without further experimental confirmation, similar conclusions have been incorporated into several recent EU Directives dealing with other analytes. Horwitz et al. pointed out, as a result of their mycotoxin studies (23), that as the concentration decreases as the detection limit is approached (at about 10 ppb), the number of false negatives increases. This effect is real, because the observation was based upon formulated test samples utilized in collaborative studies. Mycotoxin studies are unique, however, in not utilizing a blank because workers in this field have stated informally that mycotoxin-free natural products do not exist. Other definitions of LOD and LOQ use the standard deviation of the blank signal multiplied by a constant such as 3 (for LOD) or

10 (for LOQ). The EU mycotoxin studies are also unique in having been performed under exceptionally well-controlled conditions where considerable effort was utilized in method development and optimization, dedicated resources were applied toward method training and performance, and the method performance trials were conducted under advantageous circumstances.

Outliers

A factor that complicates the interpretation of analytical results is the almost universal presence of outliers in sets of data. Outliers are values that do not appear to belong to the distribution of the bulk of the data. There are 2 schools of thought with respect to how to handle outliers. One extreme believes that because outliers appeared in the data they are part of the group, unless an explanation can be found as to the high deviation. At the other extreme are practical chemists who know how easy it is to make an unrecognized and untraceable error during laboratory operations, so they have no qualms in dismissing such values as the result of an unintentional operational blunder. The former group may de-emphasize the influence of outliers by utilizing "robust" statistics, which assigns the equivalent of an importance factor to values in accordance with their closeness to the mean or central point of the distribution. Robust approaches suffer from the lack of recognized robust statistics: the weighting factors employed are arbitrary and do not necessarily reflect the true physical situation. The latter group removes values whose probability of appearance is beyond an arbitrarily assigned probability, such as 5, 2, or 1%, assuming a normal distribution. Reexamination or recalculation of these extreme values occasionally reveals an error in setting up an equation, formula weight, or a calibration; a transposition of digits; or the use of a wrong value. The number of incorrect calculations of formula weights and concentration dilutions that occur in the descriptions of the preparation of reagent and standard reference solutions submitted to AOAC INTERNATIONAL for publication is surprising. For this reason, the preparation of all reagent and standard solutions always should be described in terms of actual measured weights and volumes (g, mg, mL), although they may also be designated in terms of moles or millimoles, if these values have some theoretical significance, such as in reaction ratios. But, in general, it is impossible to reconstruct the events leading to the reporting of most values that turn out to be labeled as outliers.

Evidence of Outlier Production

To determine if a value is an outlier, a "true" or reference value must be known for the system under examination. Such a value would be available for pure elements, pure compounds, formulated mixtures, or certified reference materials. Almost by definition, values deviating substantially from the accepted reference values can be considered as outliers. Gladney et al. compiled the values found in the literature of the reported results of the examination of National Institute of Science and Technology (NIST) standard reference materials (SRMs; 27). Their review revealed that chemists produce many more outlying values than was considered acceptable. As stated by Horwitz and Albert (28) in their examination of this data:

"This compilation of reported literature values from research, method development and quality control provides, in our opinion, the best performance that the analytical chemistry profession has to offer because it displays uncensored values that are opened to public exposure and potential criticism. Here we have an assigned 'true value,' and experience shows that analysts obtain better precision when they know the answer than when they are analyzing unknowns. Therefore the variability exhibited by this population of analytical values can be subjected to various outlier treatments. The treatment that provides the consistently 'best' estimate of the mean and uncertainty can be accepted as a reasonable way to remove outliers. 'Best' in this sense is providing a mean value closest to the true value, but yet not removing an excessive number of outliers. These are somewhat conflicting requirements so we can never say their application will be the 'best' outlier treatment. In this region of low concentrations the inherent variability is so large that it would take the statistical analysis of an unreasonable number of values (i.e., thousands) to determine if one distribution is significantly different from another, e.g., if a population is normally or non-normally distributed in order to apply the 'proper' statistical outlier-removal procedure."

The data from 11 biologically related SRMs for 29 elements where at least 8 quantitative values were available per analyte/matrix combination (117 total) were examined by 3 different outlier assumptions (28):

(*1*) The IUPAC-1987 procedure (Grubbs and multiple Grubbs with rejection at the 1% probability level), recycled until values were no longer removed or automatically stopped when 22.2% (2/9) of the values were removed. (The reported values were assumed to be single values so the Cochran test could not be applied.)

(*2*) The consensus technique as applied by the compilers (less than 1% of the data was removed as "clearly beyond the limits of acceptability"; then all values beyond the mean ±2 standard deviations were removed, and the mean and standard deviation were reported as the consensus mean and associated standard deviation).

(*3*) Values beyond the NIST assigned value ±3 times the associated uncertainty, which was taken to be the SD, were assumed to be outside the region of "acceptable" variability.

The resulting means (with concentration values from 5 ng/g to 4%) and associated standard deviations were recalculated as RSD to permit examination on a common basis as the HorRat. The results from this examination of the literature reports compared to the NIST values for these biologically related elements after the outlier removal treatments are summarized here.

Systematic Error

The consensus values from the literature as established by the compilers, which covered 7 decades of concentration,

showed substantial agreement with the NIST certified values. Only 5 of the 117 values differed by more than 10% from the certified value, and the reasons for the difference were easily explained: (*1*) The differences were from trace elements present at very low levels (m/m = E-6 to E-8), which present chemical problems (Be, Cr, and V); and (*2*) the differences were from elements that present environmental problems from contamination (Fe and Cr).

The values calculated by the original IUPAC treatment (removal by the Grubbs test at the 1% rejection level) left about 10% of the values that differed by more than 10% from the certified values. The deviations were very obvious because approximately 2/3 (which would correspond to the area encompassed in a normal curve of 1 SD above and below the center or average) of the values from both outlier treatment sets (consensus and IUPAC) were within 4% of the certified value.

Random Error

The distribution of 102 HorRat values from analyte/matrix combinations with 3 or more certified values were considered, avoiding those with only 1 or 2 values as meaningless for this purpose. The average HorRat values from the consensus data reported by the compilers were 0.8; the average HorRat values from the variability calculated by the IUPAC protocol (Grubbs at 1%) were 1.6, twice as large, although within the HorRat acceptable region of 0.5–2.0.

The Gladney treatise compiled thousands of results reported in the literature of the examination of SRMs of values from research, method development, and quality control papers. This compilation reflects the best performance that the analytical chemistry profession can exhibit. This review of the data from this compilation requires a conclusion that either the NIST assigned values are incorrect or a substantial fraction of the reported values are outliers. It is far more likely that the extreme individual values are outliers than the NIST of the United States incorrectly assigned the certified value. This examination led to the adoption of the use of the Grubbs tests for outlier removal at the 2.5% probability level (1.25% in each tail) for the revised IUPAC-AOAC protocol (29) in order to make the compilation values more in line with the assignments by NIST.

The estimates of the variability parameters, i.e., variances, standard deviations, and RSD, which are based on sampling variable data, are not fixed values but are also distributions of values that resemble the downward sweep of a parabola—a constantly decreasing curve asymptotically approaching the x-axis with % uncertainty of s_R as the y-axis and number of values as the x-axis. ISO 5725-1:1994 (30; Figure B.2) provides estimates of the inherent variability to be expected for the between-laboratory reproducibility estimates from interlaboratory studies as functions of the ratio of the repeatability to reproducibility standard deviations, the number of laboratories, and the number of replicates performed by each laboratory. The uncertainty for a set of 8 laboratories performing duplicates and showing a typical ratio of within- to among-laboratory precision of 0.5–0.7 is about 35%. This explains why different interlaboratory studies can provide substantially different estimates of variability by the same analytical method at the same and at different times. Similarly, the same laboratory can provide different estimates of its variability or uncertainty of the same method applied to the same analyte in the same matrix at different times. Therefore, a single laboratory must continuously sample its performance, preferably on certified reference materials but acceptably on "house standards," and monitor this performance with control charts. This suggests that critical examinations involving acceptance or rejection of valuable lots of goods should be conducted by laboratories and analysts with a verifiable record of acceptable performance. The true underlying variability in a laboratory may change from better equipment or better environment or fluctuations may result from the vagaries of samples taken from a fixed population.

Analysts do not often realize that estimates of standard deviations are just as variable as individual observations and means because they are also calculated from individual sets of values. Also, it is rare that the prediction interval (a confidence interval as applied to individual values) and the confidence interval (as applied to means) are used for the interpretation of results. These parameters are the regions within which it is expected that future individual values or means are expected to be found with an assigned probability. Table 4, calculated by Richard Albert, contains the factors by which to multiply the found standard deviation as a function of the number of values, *N*, used to calculate the interval within which the standard deviation will be found with the specified probability. This table is used as follows: Consider an estimate of a standard deviation of 10 ppm based on 5 independent concentration estimates. Note that the units for the standard deviation are the same as those for the concentration. In the long run, if the true concentration was 5.99 ppm, one would obtain this result of 10 or even higher in one case out of 40 (= 2.5% of the time) whenever one takes 5 independent concentration estimates and calculates a standard deviation from this quintet. Traditionally, one would say that true values as low as 5.99 (but not lower!) are consistent with the found value of 10. Similarly, in the long run, if the true concentration were 28.75, then one would obtain this result of 10, or even lower, in one case out of 40 whenever one takes 5 independent concentration estimates and calculates an SD from this quintet. Traditionally, one would say that true values as high as 28.75 (but not higher!) are consistent with the found value of 10. The major point is that SDs are not fixed parameters describing the variability of sets of data but rather have a very high variability themselves.

Applications

Use of HorRat by Methods-Endorsing Organizations

A number of professional and technical organizations and governmental regulatory bodies have implemented the use of the HorRat value as an appropriate criterion for interlaboratory as well as intralaboratory variability.

Table 4. Multiplication factors required to obtain the interval containing the standard deviation as a function of the number of values and the required probability

N	Lower 2.5%	Upper 2.5%	Lower 5%	Upper 5%
2	0.446	31.911	0.510	15.952
3	0.521	6.287	0.578	4.407
4	0.566	3.727	0.620	2.919
5	0.599	2.875	0.649	2.372
6	0.624	2.453	0.672	2.090
7	0.644	2.202	0.690	1.916
8	0.661	2.035	0.705	1.797
9	0.675	1.916	0.718	1.711
10	0.687	1.826	0.729	1.645
11	0.699	1.755	0.739	1.593

IUPAC was the first organization to endorse the use of the HorRat value (2) as a result of its organization of conferences leading to the adoption of the IUPAC/ISO/AOAC protocol for interlaboratory studies (29). For concentrations below μg/g (E-6), the acceptable calculated RSD_R is taken as 22%, but the basis for this generality does not appear to be well documented. It is apparently based on an examination of the EU mycotoxin studies and, specifically an FAPAS (Food Analysis Performance Assessment Scheme) proficiency study of aflatoxin M_1 in milk at a mass fraction of 0.6E-9 (ppb; 31).

Many methods-endorsing organizations, primarily in the food field, utilize the HorRat value as the basis for their acceptance. All chemical studies examined by AOAC INTERNATIONAL are now evaluated in terms of complying with the HorRat limits of 0.5–2.0. The lowest concentration before the HorRat is found to exceed 2.0 is often taken as the LOQ.

The Nordic Analytical Committee (NMKL) specifically advises its participants to use the HorRat values in its guidelines contained in the NMKL Procedure No. 4 (32).

The International Commission for Uniform Methods of Sugar Analysis (ICUMSA; 33) expressed skepticism about the use of HorRat when applied to products of high sugar concentrations. As was pointed out previously, the curve appears to deviate from the equation at the high concentrations (greater than 10%), which are often encountered in sugar product analysis.

The European Committee for Standardization (CEN), in a document circulated for acceptance (CEN/TC 275/WG7 N 0027; 34), states: "In general the values taken from this [Horwitz] curve are indicative of the precision that is achievable and acceptable of an analytical method by different laboratories. Its use provides a satisfactory and simple means of assessing method precision acceptability. This procedure is increasingly being used by organisations to assess the acceptability of an analytical method by different laboratories . . . . The use of the HorRat value is increasingly becoming prescribed by legislation as a result of the adoption of the 'criteria-approach' by such organisations. Each working group of CEN/TC 275 has been asked to consider including the calculated HorRat values in its standards to aid the analyst in the light of the legislative developments." At the 8th meeting of CEN/TC 275/WG7 in Berlin (January 16, 2004), the committee adopted the Resolution No. 32: " . . . agrees to include HorRat values in the evaluation of all future collaborative data submitted to evaluate methodology."

HorRat in EU Regulations

Regulatory authorities in Europe were the first to grasp the significance of the HorRat value to assess the acceptability of methods of analysis for regulatory purposes. A review of analytical quality control, calling attention to its potential utility in method assessment, from the Laboratory of the Government Chemist (now designated by its initials as LGC; 35) and endorsement of the concept by IUPAC (2) apparently stimulated incorporation of the HorRat value into food control legislation of the EU as an acceptance criterion at both the single and multiple laboratory levels.

A search of the Websites of the EU (36, 37) for Directives and Regulations containing the term "HorRat" (and its variations) located a number of publications in the *Official Journal of the European Union* utilizing the term. However, this search did not locate all the pertinent documents. Several additional documents were located by a search through Google and a specific search with the EU search engine for analytes of regulatory importance. Therefore, Table 5 of EU publication references may not be a complete compilation of publications of analytes of regulatory importance that use or reference the formula.

The references to documents with a "Y" in the last column (Table 5) contain statements similar to the following: "The precision values are calculated from the Horwitz equation: $RSD_R = 2^{(1 - 0.5 \log C)}$, where RSD_R is the relative standard deviation calculated from results generated under reproducibility conditions:

$$(s_R / \bar{x}) \times 100$$

where C is the concentration ratio (i.e., 1 = 100 g/100 g, 0.001 = 1 000 mg/kg).

That is a generalized precision equation, which has been found to be independent of analyte and matrix but solely dependent on concentration for most methods of analysis." The Fusarium toxins document (Directive 2005/38) contains an alternative "fitness for purpose" approach specifying "the maximum level of uncertainty regarded as fit for purpose. . . The laboratory may use a method which produces results within the maximum standard uncertainty [*Uf,* μg/kg]" as calculated from the following formula:

$$Uf = [(LOD/2)^2 + (\alpha C)^2]^{0.5}$$

where LOD is the limit of detection (μg/kg), C is the concentration of interest (μg/kg), and α is a fractional constant varying from 0.2 (at ≤50 μg/kg) to 0.1 (at >10 000 μg/kg),

Table 5. European Union (EU) publications referring to analytes of regulatory importance

Identification	*Official Journal* reference	Subject (Y = Contains reference to equation)
Directive 96/23/EC of 29 April 1993	*L 125*, 23.05.1996, pp 0010–0032	Tissue residues
Directive 98/53/EC of 16 July 1998	*L 201*, 17.07.1998, pp 93–101	Contaminants in food (aflatoxins); performance criteria for methods (Y)
Regulation 466/2001 of 8 March 2001; Regulation 221/2002 of 6 February 2002; Regulation 472/2002 of 12 March 2002; Regulation 563/2002 of 2 April 2002	*L 77*, 16.03.2001, pp 1–13; *L 37*, 7.2.2002, pp 4–6; *L 80*, 23.03.2002, p. 42; *L 86*, 3.4.2002, pp 5–6	Maximum levels for nitrates in spinach and lettuce; mycotoxins in nuts, dried fruit, cereals, milk; Pb and Cd in foods; Hg in fish; 3-MCPD in hydrolyzed vegetable protein and soy sauce
Directive 2001/22/EC of 08 March 2001	*L 77*, 16.03.2001, pp 0014–0021	Performance criteria for methods for Pb, Cd, Hg, 3-MCPD (Y)
Regulation 2375 of 29 November 2001	*L 321*, 6.12.2001, pp 0001–0005	Maximum levels for dioxins
Regulation 257/2002 12 February 2002	*L 41*, 13.02.2002, pp 0012–0015	Maximum levels for aflatoxins
Directive 2002/26/EC 13 March 2002	L 75 16.03.2002, pp 0038–0043	Ochratoxin A in cereals, fruits (Y)
Directive 2002/63/EC 11 July 2002	*L 187*, 18.07.2002, pp 0030–0043	Sampling for pesticide residues
Directive 2002/69/EC 26 July 2002	*L 209*, 6.8.2002, pp 0005–0014	Dioxins in foods
Directive 2002/70/EC 26 July 2002	*L 209*, 6.8.2002, pp 0015–0023	Dioxin in feeds
Decision of 12 August 2002	*L 221*, 17.08.2002, pp 0008–0036	Performance of methods; interpretation of results (Y)
Decision of 13 March 2003	*L71/17*, 15.3.2003	Minimum required performance limits for chloramphenicol, medroxyprogesterone, nitrofuran metabolites
Regulation (EC) 1425/2003 of 11 August 2003	*L 203*, 12.08.2003, pp 0001–0003	Patulin in apple products (Y)
Directive 2003/78/EC of 11 August 2003	*L 203*, 12.08.2003, pp 0040–0044	Methods for sampling and performance criteria for patulin (Y)
Regulation (EC) No. 128/2004 of 23 January 2004	*L 19*, 27.01.2004, pp 0003–0011	Analysis of wines (Y)
Directive 2004/16 of 12 February 2004	*L 42*, 13.02.2004, pp 0016–0022	Performance criteria for methods for tin in canned food (Y)
Regulation 242/2004 of 2 December 2004	*L 42*, 13.02.2004, pp 0003–0004	Tin in canned foods (Y)
Directive 2005/10/EC of 4 February 2005	*L 34/15*, 8.2.2005 pp 15–20	Benzo(a)pyrene in foodstuffs (Y)
Directive 2005/38/EC of 6 June 2005	*L 143*, 7.6.2005 pp 18–26	Fusarium toxins in food (deoxynivalenol, zearalenone, fumonisins B_1 and B_2, and T2, and HT2 toxin) (Y)

depending on C. When this formula is applied to various combinations of LOD and C, it is found that that the uncertainty (*Uf*) differs appreciably from (αC) only when C approaches LOD, a situation where the estimated C is very uncertain because of the high RSD! Because different laboratories are likely to provide larger differences in their estimates of LOD than of C, it would be more conducive to harmonization to apply a constant factor as the allowance for uncertainty at the action limits for the individual toxins. However, this point may be academic because the criteria only require that an "assessment" of the performance criteria be made, with no indication of its content.

Use of HorRat Values in Proficiency Exercises

Interlaboratory studies are classified in 3 ways, depending on the primary variable of interest: method (collaborative studies), matrix (reference material studies), and laboratories (proficiency studies). Although the original function was developed from method performance studies, it has been found useful as a benchmark in the long-running FAPAS proficiency studies of the United Kingdom (38), where the Horwitz equation has been modified at both the high and low ends to account for the improved precision noted in these regions.

Thompson et al. (20) describe a proficiency scheme for geochemical analysis characterized by a test material containing over 50 analytes with a concentration range from about 0.1 μg/g to about 60% where "the Horwitz function described the overall interlaboratory precision well." An even more impressive example of the generality of the function "comes from a recent interlaboratory study of the analysis of a volcanic glass by microprobe methods (LA-ICP spectrometry

and electron probe)." The mass of material analyzed by each analyst was only a few micrograms.

The Clinical Laboratory Improvement Act of the U.S. requires laboratories examining clinical specimens to participate in proficiency testing programs. The College of American Pathologists has been providing such a program for many years. An examination of the reported results from the 2005 surveys with participants numbering in the thousands indicates that with most "standard" analytes the HorRat values are considerably better than 1.0 (39).

Summary

The simple equation showing that the relative standard deviation between laboratories (RSD_R, %) equals twice the concentration (C, as a mass fraction) raised to the –0.15 power, $RSD_R = 2C^{-0.15}$, represents interlaboratory precision very well. It is an empirical summary of more than 100 000 blind interlaboratory examinations from numerous fields of analytical chemistry, independent of analyte, matrix, method, and state of the art. When used with full knowledge of possible limitations, the equation provides a useful performance-indicating parameter for method-approval organizations and regulatory agencies.

References

(1) Horwitz, W. (1997) "The Variability of AOAC Methods of Analysis as Used in Analytical Pharmaceutical Chemistry," *J. Assoc. Off. Anal. Chem.* **60**, 1355–1363

(2) Pocklington, W.D. *Guidelines for the Development of Standard Methods by Collaborative Study*, 5th Ed., Laboratory of the Government Chemist, Teddington, UK (originally published in *Pure & Appl. Chem.* (1990) **62**, 149–162

(3) U.S. Code of Federal Regulations, 21 CFR 2.19, address corrected in Federal Register July 15, 2005, p. 40880, Access through Google: 21CFR2.19 "Methods of Analysis" (accessed January 16, 2006)

(4) AMC Technical Brief No. 17 (July 2004) "The Amazing Horwitz Function," http://www.rsc.org/pdf/amc/brief17.pdf (accessed January 16, 2006)

(5) Horwitz, W., Kamps, L.R., & Boyer, K.W. (1980) "Quality Assurance in the Analysis of Foods for Trace Constituents," *J. Assoc. Off. Anal. Chem*. **63**, 1344–1354

(6) Thompson, M. (1999) "A Natural History of Analytical Methods," *Analyst* **124**, 991

(7) Hall, P., & Selinger, B. (1989) "A Statistical Justification Relating Interlaboratory Coefficients of Variation with Concentration Levels," *Anal. Chem*. **61**, 65–66

(8) Taylor, J.K. (1987) "Quality Assurance of Chemical Measurements," Lewis Publishers Inc., Chelsea, MI

(9) Lea, P. (2004) MATFORSK, Norwegian Food Research Institute, Osloveien, Norway, personal communication, August 9

(10) Helrich, K. (1984) "The Great Collaboration: The First 100 Years of the Association of Official Analytical Chemists," AOAC INTERNATIONAL, Gaithersburg, MD

(11) Analytical Methods Committee (2001) "Measurements of Near Zero Concentration: Recording and Reporting Results that Fall Close to or Below the Detection Limit," *Analyst* **126**, 256–259; Helsel, D.R. (2005) "Nondetects and Data Analysis: Statistics for Censored Environmental Data," Wiley-Interscience, Hoboken, NJ

(12) Albert, R., & Horwitz, W. (1997) "A Heuristic Derivation of the Horwitz Curve," *Anal. Chem.* **69**, 789–790

(13) Youden, W.J., & Steiner, E.H. (1975) "Statistical Manual of the AOAC," AOAC INTERNATIONAL, Gaithersburg MD; Wernimont, G.T., & Spendley, W. (1985) "Use of Statistics to Develop and Evaluate Analytical Methods," AOAC INTERNATIONAL, Gaithersburg, MD

(14) Lynch, J., JL72@cornell.edu

(15) MAF Food Assurance Authority (2002) "A Guide to the Validation and Approval of New Marine Biotoxin Test Methods," Shellfish Quality Assurance Program, February, Wellington, New Zealand, http://www.nzfsa.govt.nz/animalproducts/seafood/guidelines/validation-guidelines-biotoxin.pdf (accessed January 16, 2006)

(16) Canadian Council of Ministers of the Environment Mercury Laboratory, (2003) Round Robin Project 257-2003, Phase 1, CRM/RM Sample Report, http://www.ceamercuryprogram.ca/EN/Pdf/CCME%20Mercury%20Phase%20II%20Fourth%20Quarter%20Report.pdf (accessed January 16, 2006)

(17) Buchgraber, M., & Anklam, E. (2003) "Method Description for the Quantification of Cacao Butter Equivalents in Cacao Butter and Plain Chocolate," Institute for Reference Materials and Measurements, Geel, Belgium, http://www.irmm.jrc.be/html/publications/technical_reports/publications/EUR20831EN.pdf (accessed January 16, 2006)

(18) Charlton, S., Giroux, R., Hondred, D., Lipton, C., & Worden, K., "PCR Validation and Performance Characteristics—AEIC Biotech Consensus Paper," http://www.aeicbiotech.org/guidelines/pcr_valid_final.pdf, (accessed October 26, 2005; not available January 16, 2006)

(19) Rosell-Melé, A. (July 6, 2001) "Precision of the Current Methods to Measure the Alkenone Proxy $U^{K'}_{37}$ and Absolute Alkenone Abundance in Sediments: Results of an Interlaboratory Comparison Study," Geochemistry Geophysics Geosystems 2, Paper No. 2000GC000141, http://www.icrea.es/ficheros/PaginaPersonal/Secciones/fsec_7910.pdf (accessed January 16, 2006)

(20) Thompson, M., Potts, P.J., Webb, P.C., & Kane, J.S. (1997) "GeoPT–A Proficiency Test for Geoanalysis," *Analyst* **122**, 1249–1254

(21) Potts, P.J., Thompson, M., & Wilson, S. (2002) *Geostandards Newsletter* **26**, 197–235; from www.rsc.org/pdf/amc/brief17.pdf (accessed January 16, 2006)

(22) Thompson, M., & Lowthian, P.J. (1997) "The Horwitz Function Revisited," *J. AOAC Int.* **80**, 676–679

(23) Horwitz, W., Albert, R., & Nesheim, S. (1993) "Reliability of Mycotoxin Assays–An Update," *J. AOAC Int.* **76**, 461–491

(24) EU Studies on Mycotoxins: Dragacci, S., Grosso, F., Pfauwathel-Marchond, N., Fremy, J.M., Venant, A., & Lombard, B. (2001) *Food Addit. Contam.* **18**, 405–415, aflatoxin M_1; Dragacci, S., Grosso, F., & Gilbert, J. (2001) "Immunoaffinity Column Cleanup with Liquid Chromatography for Determination of Aflatoxin M_1 in Liquid Milk: Collaborative Study (**2000.08**)," *J. AOAC Int.* **84**, 437–443; Visconti, A., Pascale, M., & Centonze, G.

(2001) "Determination of Ochratoxin A in Wine and Beer by Immunoaffinity Column Cleanup and Liquid Chromatographic Analysis with Fluorometric Detection: Collaborative Study (**2001.01**)," *J. AOAC Int.* **84**, 1818–1827; Visconti, A., Solfrizzo, M., & De Girolamo, A. (2001) "Determination of Fumonisins B1 and B2 in Corn and Corn Flakes by Liquid Chromatography with Immunoaffinity Column Cleanup: Collaborative Study (**2001.04**)," *J. AOAC Int.* **84**, 1828–1837, Project SMT4 CT97 2193; Josephs, R.D., Krska, R., MacDonald, S., Wilson, P., & Pettersson, H. (2003) "Preparation of a Calibrant as Certified Reference Material for Determination of the *Fusarium* Mycotoxin Zearalenone," *J. AOAC Int.* **86**, 50–60, Project SMT4-CT98-2228

(25) Lynch, J.M., Barbano, D.M., Fleming, J.R., & Nicholson, D. (2004) *Inside Laboratory Management* **8**, 24

(26) Thompson, M. (2000) "Recent Trends in Interlaboratory Precision at ppb and Sub-ppb Concentrations in Relation to Fitness for Purpose Criteria in Proficiency Testing," *Analyst* **125**, 385–396

(27) Gladney, E.S., O'Malley, B.T., Roelandts, I., & Gills, T.E. (1987) "Standard Reference Materials: Compilation of Elemental Concentration Data for NBS Clinical, Biological, Geological, and Environmental Standard Reference Materials," National Bureau of Standards Special Publication 260-111, Superintendent of Documents, Washington, DC

(28) Horwitz, W., & Albert, R. (1991) Biologically Related National Institute of Standards and Technology Standard Reference Materials: Variability in Concentration Estimates, ACS Symposium Series 445, Biological Trace Element Research Multidisciplinary Perspectives, K.S. Subramanian, G.V. Iyengar, & K. Okamoto (Eds), pp 50–73

(29) AOAC INTERNATIONAL Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis (1995) *J. AOAC Int*. **78**, 143A–160A; also reprinted in *Official Methods of Analysis of AOAC INTERNATIONAL* (2000) 17th Ed. and (2005) 18th Ed. Original publication: Horwitz, W. (1995) "Protocol for the Design, Conduct, and Interpretation of Method Performance Studies," *Pure Appl. Chem.* **67**, 331–343

(30) ISO 5725-1:1994 "Accuracy (trueness and precision) of measurement methods and results–Part 1: General principles and definition" *in Statistical Methods for Quality Control,* Vol. 2, 4th Ed., ISO Central Secretariat, Geneva, Switzerland, and National Standards Organizations, p. 28

(31) Thompson, M. (2000) "Recent Trends in Interlaboratory Precision at ppb and sub-ppb Concentrations in Relation to Fitness for Purpose Criteria in Proficiency Testing," *Analyst* **125**, 385–386

(32) NMKL Procedure No. 4 (1996) "Validation of Chemical Analytical Methods;" Under revision (2004) (in Scandinavian languages only); NMKL Report No. 11 (2000) 2nd Ed., "Guide for Referees Within Chemistry, Elaboration of Analytical Methods Within NMKL" (in Danish only); NMKL Report No. 20 (2003) "Guide for Referees Within Microbiology, Elaboration of Analytical Methods Within NMKL" (in Danish and English only)

(33) International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (2004) Berlin-Proceedings of ICUMSA, 22nd Conference, http://www.pollach.at/pages/icumsa/horrat.html (accessed January 16, 2006)

(34) European Committee for Standardization (CEN) CEN/TC 275/WG7 N 0027

(35) Mesley, R.J., Pocklington, W.D., & Walker, R.F. (1991) "Analytical Quality Assurance–A Review," *Analyst* **116**, 975–990

(36) http://europa.eu.int/eur-lex/en/search/search_lif.html (accessed January 16, 2006)

(37) http://europa.eu.int/comm/food/food/chemicalsafety/contaminants/report-sampling_analysis_2004_en.pdf (accessed October 24, 2005) (not accessible January 16, 2006)

(38) Food Analysis Performance Assessment Scheme (FAPAS) (2002) Protocol for the Organization and Analysis of Data, 6th Ed., Central Science Laboratory, Sand Hutton, York, UK, http://www.fapas.com/?CFID=1304099&CFTOKEN=13797300 (accessed October 24, 2005) (not accessible January 16, 2006)

(39) Application of the HorRat Value to Clinical Data (in preparation)