Table 2. Analytical Recovery of Phenylalanine and Galactose from Pooled Whole Blood

<table>
<thead>
<tr>
<th>Amount added, mg/L</th>
<th>Pool A</th>
<th>Pool B</th>
<th>Pool C</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA A GAL</td>
<td>0 0 0</td>
<td>30.0 30.0 150</td>
<td>80.0 80.0 400</td>
</tr>
<tr>
<td>Recovered, mg/L</td>
<td>24 18 112</td>
<td>37.0 30.6 140</td>
<td>95.0 74.0 380</td>
</tr>
<tr>
<td>% recovery</td>
<td>— — —</td>
<td>123 102 94</td>
<td>119 93 95</td>
</tr>
</tbody>
</table>

*Each value is the mean of 11 determinations. NA, not autoclaved; A, autoclaved. Pool B concentrations are near cutoff values; Pool C concentrations are clearly above normal.

greatest degree possible, a fail-safe operation. Several features are directed toward this goal. Addition of galactose to the extraction buffer produces an observable peak for each sample and facilitates identification by sequential numbering to ensure that each sample is accounted for. This results in a cutoff value corresponding to the 200 mg/L standard. The frequent use of controls (every 20th cup) ensures that the analyst reviewing the chart can easily locate groups of 20 specimens as well as identify specimens exceeding the cutoff values.

With this operating protocol we can easily recover from a mixup in specimens (transposition, misidentification of peak), the most likely error in screening programs. If the specimens are transposed in processing, and all results are negative, this is of no consequence. The presence of any presumptive positive result, however, necessitates confirmatory analysis. If upon re-analysis the result is not positive, a misidentification of specimens is assumed. Twenty specimens on either side of the initial positive specimen are reanalyzed. If no positive is found, a spurious (false) positive is assumed.

The final consideration is the elimination of false positives. In addition to the added expense, inconvenience, and anxiety caused for the parents, a high rate of false positives will result in a loss of confidence among referral physicians. The analysis/confirmation protocol as described optimizes information available from the initial specimen and keeps false-positive reports to an absolute minimum.

References

How to Improve Estimates of Imprecision

Jan S. Krouwer and Robin Rabinowitz

Authors sometimes report total imprecision as being less than within-run imprecision. This leads one to question how well the underlying statistical concepts are understood. This article explains statistical concepts relevant to imprecision studies and reviews proper use of the relevant statistical terms. Calculation of total imprecision and its components is illustrated by a numerical example. Two of the most common calculation mistakes are pointed out. Finally, we recommend improvements in experimental design that will result in more precise estimates.

Method evaluations published in Clinical Chemistry frequently contain estimates of total imprecision that are lower than the estimates of within-run imprecision for the same method. Although a previous Letter (1) pointed out that it is theoretically impossible for total imprecision to be lower than within-run imprecision, incorrect determinations of imprecision persist. One wonders how well the underlying statistical concepts are understood.

This article explains statistical concepts pertaining to imprecision studies, reviews proper use of statistical terms, shows by example how to calculate total imprecision and its components, and suggests improvements in the typical experimental designs used to estimate total and within-run imprecision.

Statistical Concepts and Terms

Figure 1 shows a model of how different components of imprecision act on an observation. In Figure 1b, the within-run imprecision of the observation y is the difference between the value of y and the mean of the run µy (i.e., the mean of the conceptual population of all results from this one run). Causes of within-run imprecision act as a random source of error on all observations in a run. The result is an...
Fig. 1. A model of how different components of imprecision affect observations
Adapted from Box et al. (3). The total error of the observation \( y = y - \mu \) is the sum of the within-run and between-run error components \( \sigma = \epsilon_{WR} + \epsilon_{BR} \). (a) Distribution of run means about grand mean. (b) Distribution of observations about a single run mean

approximately normal distribution (2) of observations, characterized by the run mean \( \mu_k \) and standard deviation \( \sigma_{WR} \), the within-run component of imprecision.

In Figure 1a, the between-run component of imprecision is the difference between the run mean \( \mu_k \) and the grand mean \( \mu \) (i.e., the mean from the conceptual population of all runs). Causes of the between-run component of imprecision act as a random source of error on all run means. The result is an approximately normal distribution of run means characterized by the grand mean \( \mu \) and standard deviation \( \sigma_{BR} \), the between-run component of imprecision.

The two sources of imprecision act independently on the observation \( y \) and combine to give the total imprecision, the difference between the value \( y \) and the grand mean \( \mu \) (Figure 1b). The distribution of all observations (not shown) is characterized by the grand mean \( \mu \) and standard deviation \( \sigma_T \), the total imprecision. Mathematically, \( \sigma_T = \sqrt{\sigma_{WR}^2 + \sigma_{BR}^2} \). Experimentally, by collecting \( n \) observations for each of \( k \) runs, one estimates the parameters \( \mu_k \), \( \sigma_{WR} \), \( \mu_k \), \( \sigma_{BR} \), and \( \sigma_T \) as the statistics \( y_k \), \( s_{WR} \), \( y \), \( s_{BR} \), and \( s_T \), respectively.

Note that the between-run component of imprecision is both conceptually and operationally distinct from the total imprecision. If the word "component" is omitted from "between-run component," the term "between-run" imprecision is frequently confused to mean either the between-run component of imprecision or total imprecision.

An example of this lack of clarity is contained in the guidelines for estimating imprecision established by this journal (4):

Studies must include estimates of "within-run" and "between-run" precision. Each should be determined at low, normal, and above-normal concentrations, with use of specimens that are in an appropriate biological fluid matrix. For the between-run precision evaluation, at least four observations should be accumulated for each of five separate groups of samples. The mean, standard deviation, and coefficient of variation should be reported.

Note that this advice is not very specific on what to do with the data. Moreover, clinical chemists usually assay a patient's specimen only once. Hence, they are concerned with the imprecision of a single observation taken from any run, which correctly should be labeled "total imprecision," not "between-run precision."

Calculations

The purpose of imprecision evaluations is to estimate the magnitude of total imprecision and of its components. Calculation of an example data set is shown in the Appendix and explained below.

To calculate the within-run imprecision, one simply takes the standard deviation (SD) of the observations of the run (equation 1, Appendix). When more than one run is conducted, the SDs for each run are combined as a weighted average, termed the "pooled within-run SD" (equation 2, Appendix).

To calculate the between-run component, one might be tempted to take the SD of the run averages about the grand average. This would only be valid if each run average were free from within-run imprecision; however, within-run imprecision does occur, so it is necessary to subtract it from the SD of the run averages to obtain a pure between-run component of imprecision (5) (equation 3, Appendix).

The total imprecision is the combination of the within-run and between-run components (equation 5, Appendix).

A common calculation error when there is more than one observation per run is to report the SD of the run averages as total imprecision. This practice results in underestimates of the total imprecision (equation 6, Appendix). Another calculation error is to report the SD of all observations as total imprecision. This practice of treating observations as if they are independent when they are not also results in underestimates of the total imprecision (1).

Experimental Design of Imprecision Studies

The calculated SD is a quantity that itself suffers from imprecision. The experimental design—that is, how many observations are taken in all, and how they are divided into observations per run and number of runs—is a controlling factor in the imprecision of the estimates of total imprecision and its components.

Figure 2 compares the approximate CVs of the estimates of total and within-run components of imprecision for two types of experimental design for the case when the within-run and between-run components of imprecision are equal. Design type A, representative of some designs that appear in Clinical Chemistry, is characterized by many observations per run, for a small number of runs. Design type B, similar to NCCLS designs (6), is characterized by relatively few observations per run, but for many runs. Each design estimates the within-run imprecision with about the same CV, but design B is much better than design A in estimating total imprecision because it adequately samples runs, which is the component of interest. For designs like A, four runs is too small a sample to give a precise estimate of either the between-run component or total imprecision.

References

3. Ibid., p 571.
5. Ref. 2, chap. 17.
6. NCCLS Proposed Standard PSEP-3 Protocol for Establishing
Fig. 2. CVs of \( s_{WR} \) and \( s_t \) for two ways of dividing the total number of observations into observations per run and number of runs

Design A, \( n \) observations per run, four runs. Design B, four observations per run, \( k \) runs. Total imprecision \( s_t \) shown by solid line. Within-run imprecision \( s_{WR} \) shown by dashed line. CVs were calculated with use of approximation (7) based on the number of degrees of freedom. \( CV = \sqrt{1/2} \times \sqrt{\text{d.t.}} \). The total degrees of freedom were calculated from Satterthwaite's formula (8) and the assumption \( s_{WR} = s_{BR} \).


Appendix

Calculation of total imprecision and its components, with an example, for equal numbers of observations per run. For unequal numbers of observations per run, see reference 9.

<table>
<thead>
<tr>
<th>Runs</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,4,3,3</td>
</tr>
<tr>
<td>2</td>
<td>3,7,7</td>
</tr>
<tr>
<td>3</td>
<td>1,2,3,2</td>
</tr>
<tr>
<td>4</td>
<td>6,4,5,4</td>
</tr>
<tr>
<td>5</td>
<td>4,4,4,5</td>
</tr>
</tbody>
</table>

Note:

\( Y_{kn} \) The value of the \( n \)th observation of the \( k \)th run where \( n = 1,2,3,\ldots,N_k \)

\( N_k \) The total number of observations in the \( k \)th run

\( \gamma_k \) The average of the \( k \)th run, where \( k = 1,2,3,\ldots,K \), estimates \( \mu_k \)

\( K \) The total number of runs

\( \bar{y} \) The grand average, estimates \( \mu \)

\( s_{WR} \) Within-run component of imprecision, estimates \( \sigma_{WR} \)

\( s_{BR} \) Between-run component of imprecision, estimates \( \sigma_{BR} \)

\( s_T \) Total imprecision, estimates \( \sigma_T \)

\( s_{RA} \) SD of run averages, \( \gamma_k \)

1. Within-run component of imprecision

For each of the \( k \) runs

\[
s_{RA} = \left[ \sum_{i=1}^{N_k} \frac{(Y_{ki} - \bar{y})^2}{K - 1} \right]^{1/2}
\]

(1)

\( s_{WR1} = 0.82 \)

\( s_{WR2} = 2.31 \)

\( s_{WR3} = 0.82 \)

\( s_{WR4} = 0.96 \)

\( s_{WR5} = 0.50 \)

For all runs, the pooled within-run component of imprecision, \( s_{PWR}, \) is:

\[
s_{PWR} = \left[ \sum_{i=1}^{K} \sum_{j=1}^{N_k} \frac{(Y_{ki} - \gamma_k)^2}{K(N_k - 1)} \right]^{1/2}
\]

(2)

\( s_{PWR} = 1.25 \)

2. Between-run component of imprecision

\[
s_{BR} = \left[ s_{RA}^2 - s_{PWR}^2 \right]^{1/2}
\]

(3)

where

\[
s_{RA} = \left[ \sum_{i=1}^{K} \frac{(Y_{ki} - \bar{y})^2}{K - 1} \right]^{1/2}
\]

(4)

\( s_{BR} = 1.27 \)

\( N_k = 4 \)

\( s_{BR} = 1.10 \)

In equation 3, \( s_{RA}^2 \) can be less than \( s_{WR}^2/N_k \), which means that the quantity under the square-root sign can be negative. In this case, one sets \( s_{BR} = 0 \).

3. Total imprecision

\[
s_T = \left( s_{BR}^2 + s_{WR}^2 \right)^{1/2}
\]

(5)

\( s_T = 166 \).

4. Run averages SD lower than total imprecision when there is more than one observation per run

Rearranging equation 3 gives

\[
s_{RA} = \left[ s_{BR}^2 + s_{WR}^2/N_k \right]^{1/2}
\]

(6)

which is less than the total imprecision when \( N_k \) is > 1.

That is:

\[
s_{RA} < \left[ s_{BR}^2 + s_{WR}^2 \right]^{1/2}
\]