Effect of Counting Errors on Immunoassay Precision

George G. Kiese1 and Gregory Post2

Using mathematical analysis and computer simulation, we studied the effect of gamma scintillation counting error on two radioimmunoassays (RIAs) and an immunoradiometric assay (IRMA). To analyze the propagation of the counting errors into the estimation of analyte concentration, we empirically derived parameters for logit-log data-reduction models for assays of digoxin and triiodothyronine (RIAs) and ferritin (IRMA). The component of the analytical error attributable to counting variability, when expressed as a CV of the analyte concentration, decreased approximately linearly with the inverse of the square root of the maximum counts bound. Larger counting-error CVs were found at lower concentrations for both RIAs and the IRMA. Substantially smaller CVs for overall assay were found when the maximum counts bound progressively increased from 500 to 10 000 counts, but further increases in maximum bound counts resulted in little decrease in overall assay CV except when very low concentrations of analyte were being measured. Therefore, RIA and IRMA systems based in duplicate determinations having at least 10 000 maximum counts bound should have adequate precision, except possibly at very low concentrations.

Additional Keyphrases: analytical error · radioimmunoassay · immunoradiometric assay · statistics of quality control

Precision of RIAs and immunoradiometric assays (IRMAS) is related to many factors, one of which is the variation of the scintillation counts. It is well known that higher counts yield better precision, but the exact relationship between counting precision and the precision of the analytical measurement precision is complex, especially when nonlinear data-reduction systems are used. The College of American Pathologists' Commission on Laboratory Accreditation includes the following question in their inspection checklist: "Are counting times sufficiently long for statistical accuracy?" (1). The interpretation of "sufficiently long" depends on the opinion of the inspector, but a frequently cited guideline is at least 10 000 scintillations counted per tube. If the counts approximate the Poisson distribution, this count rate should have a theoretical CV of <1%.

The actual variation in immunoassay test results depends not only on the theoretical counting error, but also on other factors, including the performance characteristics of the scintillation counter, the mathematical calculations used to relate the scintillation counts to analyte concentration, and the inherent precision of the other components of the assay procedure such as pipetting, antibody-binding characteristics, and separation methods. Also the introduction of multi-well scintillation counters has made it impractical to count for a fixed number of counts; thus most laboratories are counting all specimens for a fixed length of time. Consequently, counting error varies with the concentration of the analyte. The relationship between the variation of the scintillation counts and the variation of the calculated analyte concentration depends on the shape of the dose–response curve and the data-reduction algorithm used.

The practice of counting for fixed lengths of time rather than for a fixed number of counts has important implications on counting error. The counting time generally is chosen to yield "appropriate" counts for the tubes having the maximum binding of label, the counts for the other tubes being lower. Larger relative counting errors occur with specimens having lower counts: i.e., high concentrations for RIAs, low concentrations for IRMAs. However, the precision of the final analyte measurement also depends on the slope of the dose–response curve. The steeper portions of the dose–response curve serve to reduce the imprecision of the analyte concentration estimates, whereas shallower portions exaggerate the relative imprecision.

In this study we investigated two aspects of immunoassay precision: (a) reproducibility of scintillation counts in a multi-well gamma counter, and (b) the mathematical propagation of counting error into errors of estimated analyte concentrations. We mathematically analyzed the counting errors associated with five levels of counts to determine their effects on the overall assay precision. We determined the propagation of error by using both a Monte Carlo simulation model and mathematical calculations based on the slope of the dose–response curve.

Materials and Methods

The within- and across-well reproducibility of a 12-well gamma scintillation counter (Model 1260; LKB-Wallac Instruments, Inc., Gaithersburg, MD) was measured experimentally for aliquots of 125I diluted in water. Three sets of 12 glass vials (12 × 75 mm) were filled with 100-μL aliquots of the radiolabel, diluted to give approximately 100, 1000, and 10 000 counts/min. Five replications of 1-min counts were made for each of the 36 vials in each of the 12 wells, for a total of 2160 determinations. The vial-to-vial differences in counts at each quantity of label were mathematically adjusted to correct for pipetting variations. We used analysis-of-variance techniques to estimate the within-well and across-well CVs at each of the three count activities. These CVs were compared with the theoretical precisions of the Poisson decay process, which

---

1 Department of Laboratory Medicine and Pathology, Mayo Foundation, 200 First St., SW, Rochester, MN 55905.
2 Department of Laboratory Medicine, St. Elizabeth Hospital, Lincoln, NE.


Received June 29, 1988; accepted April 15, 1989.
we predicted according to the formula CV, % = 100 × (average counts)\(^{-1/2}\).

To analyze the variations in estimated analyte concentrations produced by propagation of counting error, we developed a computer-simulation program based on the logit-log data-reduction model (2). In this model the logit of the ratio of the counts bound to the counts bound at zero concentration is a linear function of the logarithm of the analyte concentration. The simulation program was written in FORTRAN and was run in an IBM-XT computer. The variation of the scintillation counts was assumed to follow the Poisson distribution, but in the model this was approximated with the gaussian distribution, with the standard deviation of the sample counts equal to the square root of the average of the sample counts, because at counts of 500 or more the two distributions are essentially identical. A set of 10,000 random numbers in a gaussian distribution was obtained from a Statistical Analytical Systems (SAS Institute, Inc., Cary, NC) program run in an IBM 370 computer, and used to generate the simulated scintillation counts. The slope and intercept of the logit-log dose-response curves were chosen from representative assays performed in our clinical laboratory. We determined these curve parameters by a weighted least-squares fit, using a Pascal version of the Rodbard program (3).

The following equations were used in developing the simulation program.

\[
\ln \left[ \frac{B(x)}{B_0} \right] = m \cdot \ln(x) + I
\]

Solving for \(B(x)\):

\[
B(x) = B_0 \cdot e^{\exp(m \cdot \ln(x) + I)} \quad \frac{B(x)}{B_0} \quad \text{or} \quad \frac{1}{B(x)} \quad \text{or} \quad m \cdot \ln(x) + I
\]

Simulated counts:

\[
\hat{B}(x) = \text{gaussian-distributed random number with mean and variance} \quad B(x)
\]

Simulated dose:

\[
\hat{x} = e^{\exp \left( \ln(\hat{B}(x)/B_0) \right)} = \left[ 1 + e^{\exp(m \cdot \ln(x) + I)} \right]
\]

Slope of response curve:

\[
d(\hat{B}(x))/dx = m \cdot \hat{B}(x)/\left[ 1 + e^{\exp(m \cdot \ln(x) + I)} \right]
\]

Where:

- \(e\) = base of the natural logarithm raised to exponential power listed
- \(x\) = concentration of analyte being measured
- \(\hat{x}\) = simulated analyte concentration
- \(m\) = slope of logit-log dose-response line
- \(I\) = intercept of logit-log dose-response line
- \(B(x)\) = counts in bound phase at concentration \(x\)
- \(B_0\) = counts in bound phase at zero concentration
- \(\hat{B}(x)\) = simulated counts bound at concentration \(x\)
- \(\bar{B}(x)\) = average counts bound at concentration \(x\)
- \(d(\hat{B}(x))/dx\) = first derivative of \(\hat{B}(x)\) relative to \(x\)

Using the simulation model, we tested three representative assay systems—two RIAs (dioxin and triiodothyronine; in-house preparations) and one IRMA (ferritin; Ranco Laboratories, Inc., Houston, TX). Five values of maximum counts bound were examined for each of the RIAs: 500, 1000, 5000, 10,000, and 50,000. In the IRMA the maximum counts bound were calculated as 1.2 times the average counts obtained with the 2000 μg/L standard, as specified by the manufacturer.

To perform Monte Carlo simulations for each value of maximum counts bound, we sampled from gaussian distributions having means and variances corresponding to the average counts for the four selected analyte concentrations for each of the assays. The average counts for these gaussian distributions were calculated from the \(B(x)/B_0\) ratios corresponding to the selected concentrations, the logit-log curves, and the specified maximum bound counts. Five-hundred iterations were simulated at each selected dose. After each trial, we calculated the averages, standard deviations, and CVs of the simulated counts and the simulated analyte concentrations.

The counting errors were mathematically related to total assay variation by using an analysis of variance model. We empirically determined the total assay variation at each level, using data from the quality-control pools. These numbers were based on the average monthly CVs from four consecutive months, representing 201 digoxin assays, 79 triiodothyronine assays, and 80 ferritin assays. These calculations contain the implicit assumption that the counting error is independent of other error sources. The simulation program with 500 iterations was used to estimate the counting errors corresponding to the average quality-control pool concentrations and counts used during this period of time. We estimated the overall assay variation associated with various levels of counting error, using the following equations:

\[
\text{CV}^2(\text{other sources}) = \text{CV}^2(\text{total}) - \text{CV}^2(\text{counting error of duplicate counts})
\]

\[
\text{CV}^2(\text{counting error of duplicate counts}) = \frac{1}{2} \text{CV}^2(\text{counting error of singlet counts})
\]

\[
\text{CV}(\text{total}) = \sqrt{\text{CV}^2(\text{counting error of singlet})/2 + \text{CV}^2(\text{other sources})}
\]

The assay variation from other sources corresponds to zero counting error.

Two specific counts, corresponding to a low concentration and a high concentration, were mathematically analyzed for triiodothyronine and ferritin, to further validate the simulation program. For each of these counts, the counting-error SD and CV were calculated as well as the corresponding concentration and the SD and CV of the concentration.

The ratios of the CVs of the concentration estimates to the CVs of the counting error from the simulation program were compared with the propagation-of-error ratio calculated mathematically. The slope of the dose–response curve was calculated from its derivation (see equations listed previously). The slope of the dose–response curve at the concentration being evaluated was used to linearly approximate the propagation of error from the counts to the concentration. The propagation of relative error was estimated as the ratio of \((\text{change in concentration}/\text{average concentration}) + (\text{change in counts}/\text{average counts})\), where the derivative of \(B(x)\) was used to estimate \((\text{change in counts}/\text{change in concentration})\). These ratios were compared with the ratio of the corresponding CVs obtained from the simulation program to help verify the simulation results.

**Results**

The precision of the 12-well counter approximated the theoretical precision of the Poisson process at each of the levels tested. Table 1 shows that the well-to-well variations...
Table 1. Precision Characteristics of the 12-Well Counter

<table>
<thead>
<tr>
<th>Total counts accumulated</th>
<th>100</th>
<th>1000</th>
<th>10 000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-well CV*, %</td>
<td>10.8</td>
<td>3.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Among-well CV, %</td>
<td>10.8</td>
<td>3.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Predicted Poisson CV, %</td>
<td>10.0</td>
<td>3.2</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* CV, coefficient of variation, calculated as (100 × SD/mean).

were only slightly larger than the within-well variations. However, it should be noted that the counter was calibrated and background counts were determined just before this experiment.

The logit-log dose–response parameters used in the simulation model were selected from representative assays for these analytes (digoxin, m = -1.174, I = 0.743; triiodothyronine, m = -1.040, I = 5.551; ferritin, m = 1.110, I = -7.004). The slope of the curve for the ferritin assay is positive, because increasing counts correspond to increasing concentrations in IRMAs.

The effects of counting error on analyte precision are graphically represented as precision profiles in Figure 1. For each of the assays, including the RIAs, CVs for relative counting error were larger at lower concentrations. Although the lower concentrations in the RIAs are associated with a higher percentage of bound counts, the propagation of error across the dose–response curve causes the imprecision to increase at low concentrations. For the IRMA ferritin assays, the imprecision of the low concentration measurements is predominantly caused by the large counting-error imprecision rather than the propagation of error.

When assessing only the assay imprecision caused by counting error, larger total counts bound produced progressively lower imprecision profiles for each of the assays. This decrease in assay CVs was approximately proportional to the reciprocal square root of the maximum counts bound. That is, a fivefold increase in counts more than halved the CV, while a doubling of counts only decreased the CV by about 30%. Higher counts were uniformly associated with lower CVs; however, the only component of variation considered in Figure 1 is that due to counting error.

The practical question concerning counting error is its net effect on assay precision. Figure 2 illustrates the effects of counting errors on total assay variation for the three assays investigated. In all cases, increased counts decreased the overall assay variation, but the effect of using higher counts was more pronounced when the counting error was large relative to the other sources of variation. For each of the assays, substantial decreases in total assay CV were noted as the total count bound increased from 500 to 1000 and from 1000 to 5000. A smaller, but still important, decrease in assay CV was noted when the bound counts increased from 5000 to 10 000. However, the precision profiles were almost flat for the change from 10 000 to 50 000 counts bound, except for those corresponding to the lowest concentrations. Note that these precision profiles are based on using the average of duplicate tubes for each

Fig. 1. Precision profiles for the component of the relative analytical variability due to counting errors as a function of analyte concentration and the maximum counts bound
(Top) digoxin (RIA), (middle) triiodothyronine (RIA), and (bottom) ferritin (IRMA)

Fig. 2. Precision profiles for the total relative assay variability as a function of the maximum counts bound and analyte concentration
(Top) digoxin (RIA), (middle) triiodothyronine (RIA), and (bottom) ferritin (IRMA)
Table 2. Examples Comparing Variation of Counts and Variation of Analyte Concentrations for an RIA and an IRMA

<table>
<thead>
<tr>
<th></th>
<th>RIA (T₄)</th>
<th>IRMA (ferritin)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Count</td>
<td>8000</td>
<td>2000</td>
</tr>
<tr>
<td>SD (count)</td>
<td>89</td>
<td>45</td>
</tr>
<tr>
<td>CV (count), %</td>
<td>1.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Concentration, μg/L</td>
<td>54.8</td>
<td>789</td>
</tr>
<tr>
<td>SD (concn), μg/L</td>
<td>3.0</td>
<td>22.0</td>
</tr>
<tr>
<td>CV (concn), %</td>
<td>5.4</td>
<td>2.7</td>
</tr>
<tr>
<td>Ratio: CV(concn)/CV (count)</td>
<td>4.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Calculated slope of B(x), counts per μg/L</td>
<td>-30.4</td>
<td>-2.1</td>
</tr>
<tr>
<td>Calculated ratio: (Δconcn/Δmean concn)/(Δcount/Δmean count)</td>
<td>4.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Examples are based on 10 000 maximum counts bound.

concentration. Errors would be larger if singlet determinations are used.

Table 2 shows the propagation-of-error calculations for the two specific counts for triiodothyronine and ferritin assays. For the RIA, the CV of the counts associated with the high concentration is more than double the CV of the counts associated with the low concentration, whereas the CV of high analyte concentration is half that of the low concentration. On the other hand, for the IRMA, the CV of the counts associated with the high concentration is less than one-third the CV of the counts associated with the low concentration, but the CV of the high analyte concentration is approximately half that of the low concentration. These differences are caused by the nonlinear propagation of counting variation to the variation of the analyte concentration.

Table 2 also compares the ratios of the relative errors in concentration estimates to the relative counting errors as determined both by the simulation program results and mathematical calculations by using the derivative of the dose–response curve. In each of the cases these ratios agree within 0.1 unit, confirming that the two techniques give comparable results. The calculation of the slope of the dose–response curve using the derivative of B(x) shows that for both assays the dose–response curve is steeper at the lower concentrations; however, when the slope is converted to relative units (CVs), the high end of the RIA and the low end of the IRMA had the smaller magnifications of the counting-error effect on concentration estimation.

Discussion

The College of American Pathologists' inspection criteria attempt to relate counting times to statistical accuracy. This is confusing, because counting time is related to precision rather than to accuracy. Even low counts should be unbiased estimates of the true counts and, therefore, should be accurate on the average if a valid data-reduction system is used. However, other counter-performance characteristics, such as linearity and variations in background counts, may influence assay accuracy.

Rodbard (4) showed that the precision of RIAs and IRMAs is dependent on the position on the dose–response curves. Generally, concentrations near the center of the range have better relative precision, but this depends on the specific assay characteristics and the methods of data reduction (5, 6). These authors also have emphasized the importance of accurate data-reduction methods. In our experience the weighted logit vs log-concentration data-reduction model, with provisions for eliminating outlying points, has been satisfactory for most RIAs and some IRMAs. Routines such as the cubic spline, four-parameter logistic, and five-parameter logistic have been useful for other IRMA and immunoenzymometric assays (7).

Ekins (8) used the concept of the "precision profile" to describe the precision of measurements as a function of analyte concentration. He showed that the absolute error, expressed as a standard deviation, increases with increasing analyte concentration, while the relative error, expressed as CV, decreases within increasing analyte concentration for RIAs. His relative curves for total assay CV are similar to those depicted in Figure 2; however, he did not specifically analyze the component of the CV contributed by counting error.

The counting error profiles for the IRMA are similar in shape to those for the RIAs—surprising, considering the inverse shape of their dose–response curves. RIAs have higher counts and therefore lower counting errors at low concentrations, while IRMAs have higher counts at high concentrations. However, for RIAs the data-reduction calculations propagate the count variation into larger relative variations in concentration estimates at low concentrations, while for IRMAs the data-reduction calculations propagate the count variation into larger relative variations in concentration estimates at high concentrations. The net effect is that both systems have similar effects of counting error on the relative precision of concentration estimates.

The question of adequate counts for adequate statistical precision still is a judgment call. The current study relates counts to analytical assay precision, but no performance criteria are available for analytical precision. Some authors have suggested that the analytical SD should be less than one-quarter of the assay normal range (9). Others have related analytical precision to the biological variation of the analytes (10, 11). Several groups have advocated the establishment of medically significant performance goals (12–16). Once appropriate analytical precision limits are defined, the question of adequate counting times can be readily answered by the approach outlined. The question of duplicate vs singlet determinations also can be analyzed by this method, once precision performance limits are defined.

Until analytical performance standards are developed, the relative size of the components of variation can help to define reasonable limits for counting error. Ironically, the better the inherent assay precision from non-counting sources of error, the greater the requirement for decreased counting error. With assay methodologies examined, the counting error achieved with 10 000 counts bound generally resulted in small increases of the overall assay variation, but if non-counting sources of error decrease with improved methodology, higher counts will be necessary to maintain this low relative contribution to overall measurement error. Also if singlet determinations are used, larger count rates will be required to achieve the same precision as found with average of duplicate determinations.

References