Statistical Characterization of the Random Errors in the Radioimmunoassay Dose–Response Variable

David Rodbard, Robert H. Lenox, H. Linton Wray, and Douglas Ramseth

We have developed practical methods for evaluating the magnitude of the random errors in radioimmunoassay dose–response variables, and the relationship between this error and position on the dose–response curve. This is important to obtain appropriate weights for each point on the dose–response curve when utilizing least-squares curve-fitting methods; to evaluate whether the standards and the unknowns are subject to error of the same magnitude; for quality-control purposes; and to study the sources of errors in radioimmunoassay. Both standards and unknowns in radioimmunoassays for cAMP and cGMP were analyzed in triplicate. The sample mean (Ȳ), sample standard deviation, sY, and variance (sY2) of the response variable were calculated for each dose level. The relationship between sY2 and Ȳ was calculated utilizing several models. Results for standards and unknowns from several assays were pooled, and a curve smoothing procedure was used to minimize random sampling errors. This pooling increased the reliability of the analysis, and confirmed the presence of the theoretically predicted nonuniformity of variance. Thus, the calculation of results from these radioimmunoassays should utilize a weighted least-squares curve-fitting program. These analyses have been computerized, and can be used as a "preprocessor" for programs for routine analysis of results of radioimmunoassay.

Additional Keyphrases: quality control • computer analysis • precision • heteroscedasticity • weighted least-squares regression

Numerous methods have been proposed for the computer analysis of radioimmunoassay (RIA) results. These usually involve use of a least-squares regression, either linear in the case of the logit-log method (1, 2), or nonlinear as in the mass action law models (2) or the four-parameter logistic model (3, 4). In any least-squares method, we need to make assumptions regarding uniformity of variance of the response, or else derive an appropriate weighting function to compensate for nonuniformity of variance. It is clear that the logit transformation introduces severe nonuniformity of variance (1). However, controversy has persisted regarding the uniformity or nonuniformity of variance of the radioimmunoassay dose–response variables in the absence of a logit transformation, e.g., when using response variables such as "raw" counts bound, counts free, bound over total (B/T), bound over free (B/F), free over total (F/T), free over bound (F/B), percent bound relative to initial percent bound (B/B0) or its reciprocal 1/(B/B0). Rodbard et al. have consistently claimed nonuniformity of variance for RIAs for hLH, hFSH, hGH, angiotensin, ferritin, serum hepatitis B virus, and most other assays (1–4). This has been confirmed by Midgley et al. (5); however, many other workers have implicitly or explicitly made the assumption that there is uniformity of variance (6–11) for the RIA response variable. Some workers have

---

1 Nomenclature and nonstandard abbreviations: RIA, radioimmunoassay; B/T, bound-to-total ratio for labeled ligand corrected for nonspecific counts; B/F, bound-to-free ratio for labeled ligand corrected for nonspecific counts; B/B0, normalized fraction bound for labeled ligand, i.e., B/T relative to B/T for zero dose of labeled ligand.
implicitly assumed that there was uniformity of variance for the logarithm of counts bound (12). Bliss (13) and Duddleon et al. (14) have made the assumption that the response variables \( B/T \), \( B/B_0 \), or bound counts behave as "Poisson-like" variables, with variance directly proportional to the expectation of the response. Spona (15) implicitly assumed uniformity of variance for the logarithm of the logarithm of the counts bound. Others have implicitly assumed that 1/(\( B/B_0 \)) or "\( B_0/B' \)", or 1/(\( B/T \)), or 1/(counts bound) shows uniformity of variance, a clearly untenable assumption (16). Chang et al. (17) apparently advocate unweighted regression, but with close attention to experimental design in terms of choice of dose levels. However, their own data indicate severe nonuniformity of variance for \( B/B_0 \) (their Table I), and their results indicate better performance for a weighted rather than unweighted logit-log regression, in terms of the shape of confidence intervals.

If there were no experimental errors in the pipetting of any of the reagents, nor in the separation of the bound and free fractions, such that the response variable was subject only to the counting error caused by the random radioactive decay process for the isotope, then one should have a true Poisson variance for the observed (raw) counts, governed by the equation \( \sigma_Y^2 = E(Y) \), where \( E(Y) \) represents the true mean or expectation of \( Y \). Thus, in general, one expects that under ideal assay conditions (where pipetting and classification errors are negligible) one would observe nonuniformity of variance. Further, it can be shown theoretically that there is usually severe nonuniformity of variance of the response variable, even in the presence of experimental errors (1, 18, 19, and unpublished studies). If we make a series of plausible assumptions (e.g., a random pipetting error for each of the reagents, and random errors in total reaction volume, reaction time, temperature, and rate constants), one can show that \( \sigma_Y^2 \) is nearly linearly related to \( E(Y) \) for assay conditions close to an "ideal" (19, and unpublished data). These theoretical studies also show that the "classification" errors (random errors in the separation of bound and free ligand) may either accentuate, diminish, or eliminate this nonuniformity of variance (1, 19). Thus it is not surprising that some assays in some laboratories may show uniformity of variance, while other assays in other laboratories may show severe nonuniformity of variance. We postulate that most claims of uniformity of variance are the result of a failure to use appropriate methods to detect nonuniformity of variance. In many bioassay programs, Bartlett's test is included to test for heterogeneity of variance. However, this test is extremely inefficient in the presence of low degrees of freedom. Further, it is quite insensitive to systematic (as opposed to "random") changes in \( \sigma_Y^2 \) along the dose-response curve.

Previously distributed versions of the logit-log program (2) contained subroutines that provided a plot of \( s_Y^2 \) vs. \( Y \) for the standard curve, and provided estimates of the variance at both extreme ends of the curve (\( B/B_0 = 0, 100 \)). More recent versions, used in our laboratories for the past two years, display a similar analysis for the unknowns whenever the necessary replicates are present. However, further analyses and decisions were left to the discretion of the user; as a result these data were rarely used appropriately. The present study was undertaken so that appropriate analyses of nonuniformity of variance would be made automatically, and incorporated into a "preprocessor" program before "fitting" the dose-response curve. RIAs for cyclic AMP and cyclic GMP were analyzed in detail, and found to display nonuniformity of variance of the type predicted by theory.

**Materials and Methods**

**Assay Methods**

Adenosine 3':5'-cyclic monophosphate (cAMP) and guanosine 3':5'-cyclic monophosphate (cGMP) were assayed by a modification of the radioimmunoassay procedure described by Steiner et al. (20). These assays are based on competition of the cyclic nucleotide with isotopically labeled cyclic nucleotide derivatives for binding sites on specific antibody. Derivatives of the cyclic nucleotides with high specific activity were prepared by iodinating the tyrosine methyl ether derivative of the succinylated cyclic nucleotide with \( ^{125} \)I. Free and antibody bound \( ^{125} \)I-labeled cyclic nucleotides were then separated by use of Carbowax (polyethylene glycol). Binding equilibrium is reached in 24 h. All assays were done by the same technician under identical conditions, and both standard and unknown samples were analyzed in triplicate. Similar specific activities, total counts, and antibody concentrations were used in successive assays.

**Statistical Methods**

Data were analyzed by using the weighted nonlinear four-parameter logistic model as described by Rodbard and Hutt (3). A new "preprocessor" program was written, to perform the series of calculations outlined in Table 1. For each set of replicates (triplicates in the present study), the sample mean (\( \bar{Y} \)), sample standard deviation (\( s_Y \)), and the variance (\( s_Y^2 \)) of the responses were calculated and stored. \( s_Y^2 \) was plotted vs. \( Y \), with separate symbols for standards and unknowns. Then, these data were analyzed by least-squares regression, by using the following equations:

\[
\sigma_Y^2 = a_0 + a_1 \{ \mu_Y \}
\]

or

\[
s_Y^2 = a_0 + a_1 \{ \mu_Y \} + e_1
\]

or

\[
s_Y^2 = s_0 + a_1 Y
\]
\[ \sigma^2_Y = a_0 + a_1 Y \]  
(1)

\[ \sigma^2_Y = a_0 + a_1 Y + a_2 Y^2 \]  
(2)

\[ \log(\sigma^2_Y) = \log(a_0) + j \log(Y) \]  
(3)

These three models are designated, for convenience, the “linear”, “quadratic”, and “log-log” models, respectively. Use of the “log-log” model was suggested for this purpose by Prof. David J. Finney (personal communication). Both weighted and unweighted regression was used to fit the parameters of these equations (see below). For equation 2, the method of orthogonal polynomials was used (21). To reduce the sampling error in the calculated sample variances with low degrees of freedom, we “pooled” the individual estimates of \( \sigma^2_Y \) by clustering the data. This was done in either of two ways: (a) by dividing the observed range of \( Y \) into 10 (or 30) equal segments, or (b) by dividing the observed range of \( Y \) into unequally spaced regions containing an equal (or nearly equal) number of “points” (\( \bar{Y} \), \( \sigma^2_Y \) pairs). The former method resulted in clusters (“bins”) with drastically different numbers of points and hence degrees of freedom. This necessitated the use of weighted regression when finding estimates of parameters for equations 1–3. Accordingly, we adopted the second approach to “clustering.” The median \( \sigma^2_Y \) was found for each cluster or “bin”. We used the median, rather than the mean, in order to minimize the effects of “outliers” in the raw data, and (or) in the calculated \( \sigma^2_Y \). Then the regression analyses for equations 1–3 were repeated on these “smoothed” or “binned” data. In some preliminary studies, we also used the models:

\[ \sigma^2_Y = a_1 Y \]  
(4)

and

\[ \sigma^2_Y = a_0 Y^j \]  
(5)

Equation 4 is the “Poisson-like” model, a degenerate case of equations 1–3 or 5. Equation 5 is the explicit form of the “exponential” model proposed by D. J. Finney (personal communication). It is a generalization of equation 4, and it may be viewed as an alternative model to equations 1 or 2. Equation 5 was “fit,” with and without weighting, by using the “MLAB” system (22) at the NIH Division of Computer Research and Technology. We used weighting, to compensate for the nonuniformity of variance of \( \sigma^2_Y \). The weighting functions included

\[ W = df \]  
(6)

\[ W = \frac{df}{\sigma^2_Y} \]  
(7)

\[ W = \frac{df}{(\sigma^2_Y)^2} \]  
(8)

where df is the degrees of freedom for the local estimate of \( \sigma^2_Y \). The \( \sigma^2_Y \) in the denominator could be the median \( \sigma^2_Y \) in a “bin”, or the \( \sigma^2_Y \) predicted from a previous attempt (iteration) at curve fitting. Usually, similar results were obtained for the parameters of models 1–5 irrespective of the weighting used. Non-linear curve fitting by use of equation 5 with weights given by equations 6–8 often failed to converge and was unsatisfactory for purposes of routine analysis. The use of equation 3 with unweighted regression is nearly equivalent to the use of equation 5 with weights given by equation 8, because the logarithm of the sample variance is asymptotically normally distributed and shows uniformity of variance. Using the log transform does not permit direct comparison of the residual variance for equation 3 with that of the other equations; accordingly, the residual variance for the fit to equation 3 was recalculated in terms of \( \sigma^2_Y \).

After calculating the several regressions (equations 1–5), we selected the “best” model as the one with the smallest residual variance. In comparing equation 2 (which involves three fitted parameters) with the equations 1 or 3 (which involve two fitted parameters), we used an approximate F test:

\[ F = \frac{\Delta \text{SS} / \Delta \text{df}}{\text{MS}} \]  
(9)

where \( \Delta \text{SS} \) is the difference in the sum of squares for the two models, \( \Delta \text{df} \) is the change in number of df, and MS is the Mean Square (residual variance) for the model with the larger number of fitted parameters (equation 2 in this case). Similar considerations apply when comparing model 4 with the others. Also, to be suitable for use as a weighting function in subsequent curve-fitting procedures, the predicted variance (\( \sigma^2_Y \)) must be greater than zero over the allowable or observable range for the assay. If any model resulted in a predicted \( \sigma^2_Y \) that was zero or negative with \( Y \) in the observed range for an assay, then this

### Table 1.

1. Calculate and store \( \sigma^2_Y \), \( \bar{Y} \) for each set of replicates.
2. Plot \( \sigma^2_Y \) vs. \( \bar{Y} \) for standards.
3. Plot \( \sigma^2_Y \) vs. \( \bar{Y} \) for unknowns. (Plots 2 and 3 may be superimposed on the same graph.)
4. “Pool” the results for standards and unknowns after testing for homogeneity, if sufficient data are available.
5. “Fit” the relationship between \( \sigma^2_Y \) and \( \bar{Y} \) by using three models (linear, quadratic, log-log).
6. Plot the “fitted” relationships.
7. Select the “best” model by objective, quantitative criteria (e.g., residual variance, F test).
8. “Bin” the results in local regions of the dose response curve, and repeat steps 2–7.
9. Pool results over assays, and repeat steps 2–8.
model was rejected and the next best model was selected. Appropriate calculations for transformed response variables (e.g., B/T, B/F, B/B0) are given in the Appendix.

Results

Figure 1 shows a representative dose–response curve. Gross inspection suggests the presence of nonuniformity of variance of the response variable (bound counts) in RIA for both cAMP and cGMP. Figure 2 indicates the relationship between \( s_\chi^2 \) and \( Y \) for the standards and unknowns from a single assay for cGMP. The results for the standards and unknowns appear compatible, so they have been “pooled” for subsequent analyses. This relationship has been fit by a linear model, a quadratic polynomial model, and a log-log model (equations 1–3), designated LIN, QUAD, and LOG, respectively. Each point on this graph represents a sample variance based on only two degrees of freedom (df = N - 1 = 2 for triplicates). Thus, there is enormous sampling error in the sample standard deviation and sample variance and indeed the 95% sampling limits for \( s_\chi^2 \) are at .025 and 3.69 times the expected results. To simplify the analysis and to make clear the relationship between \( s_\chi^2 \) and \( Y \), we divided the horizontal axis representing the response variable into 10 segments between the lowest and highest values, such that each segment contains the same number of “points” (and, hence, df in the present case). The median variance in each of these intervals was then calculated. Results from assays for cGMP are shown in Figure 2B. Now, each of the points represents an estimate of

\[ 3 \] A program, written for a CDC 3500 in Fortran IV, is available on request from Douglas Ramseth, Walter Reed Army Institute of Research, Washington, D. C., or from the Editorial Office of this journal.

the variance based on approximately 24 degrees of freedom. The linear, parabolic, and log-log models have been fitted to these data, combining results from standards and unknowns. Note the excellent agreement between the three models utilized over the observed range for the response variable. The results of the three models diverge as one extrapolates from the data. Results from 22 consecutive assays for cGMP are summarized in Table 2. Figure 3 shows what happens when we “pool” results (for \( s_\chi^2 \) vs. \( Y \)) for 22 cGMP assays. Although Figure 3A appears to be completely “random” at first blush, regression analysis shows a definite, significant relationship when equations 1, 2, or 3 are used. By pooling or “binning” and thus obtaining about 400 df for each local estimate of \( s_\chi^2 \), a very strong underlying relationship is seen (Figure 3B). Again, there is an excellent correlation, whether one utilizes a linear, quadratic, or log-log model. Quite clearly, there is nonuniformity of variance with about a fourfold increase in \( s_\chi^2 \) over the observable range of counts. A similar analysis can be applied separately to the points from the standard curve (Figure 4) or the unknowns (Fig-
Table 2. Summary of Relationship between $\sigma_Y^2$ and $Y$ for cGMP

<table>
<thead>
<tr>
<th>Assay</th>
<th>Linear N</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>Quadratic $a_0$</th>
<th>$a_1$</th>
<th>$a_1 \times 10^3$</th>
<th>Log-Log $a_0$</th>
<th>j</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>122</td>
<td>3376</td>
<td>4.634</td>
<td>-4478</td>
<td>8.41</td>
<td>-0.391</td>
<td>0.461</td>
<td>1.206</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>-6904</td>
<td>7.834</td>
<td>-10310</td>
<td>10.20</td>
<td>-0.285</td>
<td>0.017</td>
<td>1.653</td>
</tr>
<tr>
<td>3</td>
<td>95</td>
<td>15086</td>
<td>3.184</td>
<td>-25481</td>
<td>26.97</td>
<td>-2.777</td>
<td>0.277</td>
<td>1.286</td>
</tr>
<tr>
<td>4</td>
<td>76</td>
<td>19191</td>
<td>4.034</td>
<td>-72820</td>
<td>53.93</td>
<td>-5.936</td>
<td>5.391</td>
<td>0.986</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>15505</td>
<td>2.459</td>
<td>-17282</td>
<td>18.67</td>
<td>-1.684</td>
<td>0.548</td>
<td>1.182</td>
</tr>
<tr>
<td>6</td>
<td>112</td>
<td>16348</td>
<td>3.543</td>
<td>-29530</td>
<td>27.36</td>
<td>-2.701</td>
<td>57.910</td>
<td>0.686</td>
</tr>
<tr>
<td>7</td>
<td>148</td>
<td>12599</td>
<td>2.716</td>
<td>-8938</td>
<td>11.18</td>
<td>-0.747</td>
<td>232.300</td>
<td>0.480</td>
</tr>
<tr>
<td>8</td>
<td>133</td>
<td>6244</td>
<td>2.856</td>
<td>-38499</td>
<td>24.73</td>
<td>-2.331</td>
<td>0.194</td>
<td>1.274</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>9039</td>
<td>0.929</td>
<td>-17785</td>
<td>18.52</td>
<td>-2.338</td>
<td>113.900</td>
<td>0.501</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>24941</td>
<td>3.963</td>
<td>-43608</td>
<td>40.60</td>
<td>-4.020</td>
<td>36.800</td>
<td>0.795</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>-5512</td>
<td>7.884</td>
<td>13256</td>
<td>-6.28</td>
<td>1.621</td>
<td>0.137</td>
<td>1.374</td>
</tr>
<tr>
<td>12</td>
<td>116</td>
<td>11920</td>
<td>7.207</td>
<td>-2929</td>
<td>21.83</td>
<td>-2.023</td>
<td>0.859</td>
<td>1.199</td>
</tr>
<tr>
<td>13</td>
<td>105</td>
<td>8741</td>
<td>5.199</td>
<td>5069</td>
<td>7.38</td>
<td>-0.267</td>
<td>27.800</td>
<td>0.669</td>
</tr>
<tr>
<td>14</td>
<td>125</td>
<td>4385</td>
<td>2.488</td>
<td>-365</td>
<td>7.65</td>
<td>-0.906</td>
<td>33.380</td>
<td>0.787</td>
</tr>
<tr>
<td>15</td>
<td>95</td>
<td>28116</td>
<td>2.891</td>
<td>-21314</td>
<td>29.70</td>
<td>-2.633</td>
<td>60.610</td>
<td>0.688</td>
</tr>
<tr>
<td>16</td>
<td>88</td>
<td>19453</td>
<td>3.555</td>
<td>-13625</td>
<td>17.87</td>
<td>-1.231</td>
<td>12.070</td>
<td>0.913</td>
</tr>
<tr>
<td>17</td>
<td>70</td>
<td>23085</td>
<td>3.620</td>
<td>-27863</td>
<td>29.86</td>
<td>-2.566</td>
<td>31.630</td>
<td>0.769</td>
</tr>
<tr>
<td>18</td>
<td>148</td>
<td>14961</td>
<td>3.788</td>
<td>-5432</td>
<td>15.15</td>
<td>-1.239</td>
<td>49.220</td>
<td>0.579</td>
</tr>
<tr>
<td>19</td>
<td>98</td>
<td>6933</td>
<td>0.587</td>
<td>-1863</td>
<td>7.00</td>
<td>-0.859</td>
<td>1518.000</td>
<td>0.256</td>
</tr>
<tr>
<td>20</td>
<td>99</td>
<td>14870</td>
<td>2.566</td>
<td>-284</td>
<td>10.42</td>
<td>-0.864</td>
<td>284.500</td>
<td>0.484</td>
</tr>
<tr>
<td>21</td>
<td>90</td>
<td>17954</td>
<td>2.492</td>
<td>23887</td>
<td>0.21</td>
<td>0.187</td>
<td>19.510</td>
<td>0.751</td>
</tr>
<tr>
<td>22</td>
<td>88</td>
<td>4166</td>
<td>3.141</td>
<td>-11316</td>
<td>14.64</td>
<td>-1.528</td>
<td>5.858</td>
<td>0.917</td>
</tr>
</tbody>
</table>

Mean

Pooled

Pooled and binned

$^4$Results are shown for each of the three regression models for several assays. The means of the coefficients are presented, together with the coefficients obtained by regression analysis for the data pooled over assays ("pooled"), or for data that have been both pooled over assays and "binned" into ten intervals containing the same number of degrees of freedom. It is apparent that the coefficients for the quadratic model are extremely unstable (between assays). Ideally, one should examine these coefficients together with their joint 95% confidence limits.

Figure 5. This provides a more refined test for similar behavior of standards and unknowns than simply analyzing the data from one assay (e.g., Figure 2).

The linear regressions for all the cases given above are highly statistically significant, i.e., the slope is significantly different from zero ($P < .01$). In this example it appears that $s_Y^2$ for the high Y values is approximately four times larger than the variance at the other end of the curve. The intercept ($a_0$) does not appear to be significantly different from zero. Hence, we are dealing with a Poisson-like variable, and in lieu of the two-parameter linear model, we could utilize a simple linear model forced through the origin, i.e., equation 4. Also, when utilizing the log-log model of Finney, the exponent j does not differ significantly from unity (Table 2). In this particular example, all three of the equations (1-3)—or, for that matter, all five of the models—would have provided a satisfactory model for the nonuniformity of variance. Indeed, in view of the large sampling error, it appears impossible to discriminate between the performance of these five models for the present data. Similar results, summarized in Figure 6A, were obtained with the cyclic AMP assay. Figure 6B again shows the median variance in each of 10 bins, pooling data from both the standard and the unknowns from 23 assays.

Discussion

Use of the Predicted Variance for Weighted Curve-fitting

The major practical application of these analyses is the use of one of the models to calculate weights for the curve-fitting procedure used to describe the standard curve, and used as a basis for dose interpolation for unknowns. Jacquez and Norusis (23) have pointed out that use of empirically determined weights may actually result in a decreased performance of a linear curve-fitting procedure. This is true if the weights are based on a sample variance with a very small number of degrees of freedom (df < 9). Accordingly, in some commercial adaptations of the "logit-log" curve-fitting method, which uses weights based on a sample variance (for logit(B/B0)) with only one or two df, the use of weights may actually seriously degrade the performance of the curve-fitting procedure. In contrast to Jacquez and Norusis, the approach advocated here and elsewhere (1-5, 14) results in an estimate of $\sigma_Y^2$ with extremely high degrees
Fig. 3 A. Relationship between $s_2^2$ and $\bar{Y}$ for a cyclic GMP assay, on pooling data over assays containing 2000 points altogether. Although at first inspection, this appears to be a random scattergram, utilization of linear, quadratic, or log-log regression indicates a clear-cut relationship, such that $s_2^2$ varies systematically with $\bar{Y}$.

B. The data of Figure 3A has been clustered into 10 bins containing 200 observations per bin. The median variance within each “bin” is determined.

Fig. 4 A. Relationship between $s_2^2$ and $\bar{Y}$ for the cyclic GMP RIA, on pooling over 22 assays but utilizing data for “standards” alone and excluding data from “unknowns.”

B. Data of Figure 4A, after clustering into bins and finding the median variance within each bin. There is marked nonuniformity of variance, and any of the three models used is adequate to describe the data.

Fig. 5 A. Relationship between the $s_2^2$ and $\bar{Y}$ for data from “unknowns” only, excluding data from standards.

B. Same data, after binning and finding the local median.

Fig. 6 A. $s_2^2$ vs. $\bar{Y}$ for RIA for cyclic AMP, on combining data from both standards and unknowns over 23 assays, including 2000 points.

B. Some data as Figure 6A, after pooling data into 10 bins with 200 observations per bin. Again, there is marked nonuniformity of variance, which might have escaped detection if one utilized only casual perusal of Figure 6A.
of freedom and reliability. For instance, we do not "weight" each point according to the reciprocal of any one "dot" in Figure 2A; instead, we weight according to the value of \( \sigma_Y^2 \) predicted by the smooth curve in Figures 2 to 6. By "pooling" information from all dose levels, from standards and unknowns, and from many assays, we have drastically reduced the effect of sampling error and extracted information about the underlying nature of the error in the response variable. Other methods of pooling or smoothing could be used. For purposes of estimating parameters for a given model, the "binning" of points (Figures 2B-6B) is unnecessary. We obtained very similar parameter estimates when using curve fitting based on all of the points (Table 2). However, the use of "binning" makes it possible to see the underlying relationship more clearly. This is even more important when individual estimates of \( s_Y^2 \) are based on duplicates, and hence have only one degree of freedom.

Having established the relationship between \( \sigma_Y^2 \) and \( Y \), one can now weight each of the points \((X, Y)\) in a least-squares regression according to the inverse of the predicted (rather than the observed) variance for that point. Because the variances are calculated on the basis of the response variable rather than on the basis of the dose or \( X \) variable, we should re-adjust the weights after each iteration, recalculating the weights on the basis of the response level predicted for that dose level on the basis of the previous iteration. In this manner, the weights become an implicit function of the dose level, and all replicates for a given dose are given the same weight. This technique has been used in the methods given in references 1-4 and is analogous to what is commonly done in probit or logit analysis with use of quantal response variables (24). This approach to weighting is also applicable to all other methods of least-squares curve fitting, including multiple polynomials, spline-fitting, mass-action law models, etc. With computerized calculations no additional effort is required on the part of the user.

The present results confirm previous findings of significant nonuniformity of variance for RIAs. Hence, in the assays for cGMP and cAMP and in general, unless proven otherwise, it is desirable to utilize weighted regression. Failure to do so will not necessarily result in serious error in the estimate of the curve (9, 17, 19, 23), but it invalidates attempts to calculate the 95% confidence limits or the coefficient of variation for the potency estimate based on the scatter of the points around the standard curve. Use of a nonweighted logit-log regression and the assumption of uniformity of variance will result in serious underestimation of the error at certain positions (the extremes) of the dose–response curve and overestimation of the error at other positions (the center) of the curve [see Table 2 in (17)]. Use of a nonweighted four-parameter logistic (4) will (usually) result in underestimation of the error in the low-dose region and overestimation of the error in the high-dose region of the curve. Use of weighted regression makes it possible to improve estimation of potency when the same sample is analyzed at multiple dose levels, by use of the weighted, rather than the simple average of the individual potency estimates (26). Moreover, the present findings, on a large data base, confirm the theoretical predictions (1, 18, 19) of nonuniformity of variance for radioimmunoassay dose–response variables.

Choice between Models

In the cAMP and cGMP assays analyzed here, any of the models 1–5 would have been satisfactory. We expect that for other assay systems one or more of these models will fail [e.g., see (3)]. Indeed, for some assays, a "best" model may clearly emerge. Model 1 has the simplicity of a linear model and involves two parameters. In contrast, model 2 involves three parameters and is slightly more general; for instance, it can handle the case where \( \sigma_Y \) is proportional to \( Y \) (i.e., a constant coefficient of variation) (4). This is quite important in some applications (3), particularly immunoradiometric assays. Model 2 can handle the case where \( Y \) is a binomial variate, and thus can be applied to quantal assays. However, in model 2, the parameters are very unstable from assay to assay, and it becomes much more difficult, if not impossible, to pool results (coefficients) from several successive assays (Table 2). Equation 2 also has the undesirable property of frequently predicting a zero or negative variance. To circumvent this, Equation 2 could be "fit" subject to the nonlinear constraint that \( \sigma_Y^2 \) must be positive over the range of \( Y \) observed in the assay.

Model 3 is "linear" after log transforms of both axes. This model is not directly applicable to linearly transformed response variables such as \( B/T \) or \( B/B_0 \) when the mean nonspecific counts (\( N \)) is not zero. This model has the advantages that it confers a degree of uniformity of variance to the dependent variable, \( \log(s^2) \), which is quite important for analysis of results from a single assay (before "binning"). Also, model 3 (or 5) avoids the problem of having the predicted \( \sigma_Y^2 \) becoming negative. Model 3 has a potential problem when the observed sample variance (\( s^2 \)) is zero. This is readily remedied by use of \( \log(s^2 + c) \), where \( c \) is a small constant. Model 5 has the serious problem that for limited data of the type usually obtained for a single assay it is very difficult to obtain convergence of a non-linear curve-fitting routine, and (or) estimates of \( a_0 \) and \( j \) are very unstable.4

In our laboratory, we have adopted the strategy of always using the model with the smallest residual

4 A reviewer suggested another model that may be appropriate for the present data:

\[
\sigma_Y^2 = \frac{a_0 Y}{a_1 + a_2 Y}
\]

Indeed, any simple model may be used. We prefer equations 1–3 to this hyperbolic model, especially because upward rather than downward curvature is expected in most RIA systems.
variance provided it predicts a positive $\sigma^2$ over the observable range. There are advantages and disadvantages to each of the models. It is essential that the final model that is selected fits the data characteristic for each assay system in each laboratory. The methods described above allow an investigator to evaluate his assay for an appropriate model. Our current programs for analysis of RIA results allows the user to select the model he prefers (26). These programs should be useful for a wide variety of analyses in the clinical chemistry laboratory, in addition to RIAs.3

Conclusions

The radioimmunoassays for cAMP and cGMP analyzed here show nonuniformity of variance for the response variable. Accordingly, weighting should be used for optimal data processing. Several models are available to describe the relationship between the variance of the response and the response level. Parameters of these models should be obtained by objective, least-squares regression methods. It is desirable to compare the results from the standards and the unknowns and to pool the data from these two sources, if appropriate. Further, data can be pooled over several assays in a stable assay system. Then, parameters for the model (e.g., $a_0$, $a_1$, $a_2$) can be used for curve-fitting in subsequent assays if the assay remains in satisfactory quality control. These parameters can then be updated periodically, for example, after every 10 assays. With an automated, computerized approach, these parameters can be estimated and updated after each assay. Direct inspection of the error in the response variable facilitates quality-control programs because this error, together with the slope of the dose-response curve, completely determines the precision of an estimate at any point on the dose-response curve.

D. M. Hutt of the Standard Information Systems, McLean, Va., assisted with computer programming for a prototype of this program. V. B. Faden provided equations 12 and 13. R. Sheets provided expert assistance with the computations involving the MLAB system of the DEC-10 computer. T. Sellner prepared the manuscript on the WYLBUR text-editing system of the NIH Division of Computer Research and Technology. A. Guirguis expertly performed the assays. D. Tang critically reviewed the manuscript. D. J. Finney provided many helpful suggestions and access to unpublished material.

Appendix

The present analyses have been performed on the original response variable or raw data, i.e., counts bound. This appears to be the most efficient approach when subsequent data processing is by a method such as the four-parameter logistic (3, 4, 6) or others (7, 10, 12, 15) that also make use of the raw counts. The present program may also be used when subsequent processing makes use of a linear transform of the response (e.g., $B/T$, $B/B_0$) or a nonlinear transform, e.g., $\logit (B/B_0)$. For instance, if $Y$ and $a_0$, $a_1$, $a_2$ apply to “raw counts”, and

$$Y' = \frac{B/B_0}{B_0 - \bar{N}}$$  \hspace{1cm} (10)

where $B_0$, $\bar{N}$ are constants, then

$$\text{Var}(Y') = c_0 + c_1 Y' + c_2 (Y')^2$$  \hspace{1cm} (11)

where

$$c_0 = (a_0 + a_1 \bar{N} + a_2 \bar{N}^2)/(B_0 - \bar{N})$$
$$c_1 = (a_1 + 2a_2 \bar{N})/(B_0 - \bar{N})$$
$$c_2 = a_2$$  \hspace{1cm} (12)

Alternatively:

$$a_0 = (B_0 - \bar{N})^2 c_0 - \bar{N}(B_0 - \bar{N}) c_1 + \bar{N}^2 c_2$$
$$a_1 = (B_0 - \bar{N}) c_1 - 2\bar{N} c_2$$
$$a_2 = c_2$$  \hspace{1cm} (13)

Thus, determination of the coefficients for models 1–5 allows one to calculate the variance for any linearly transformed variable.

Note that models 1, 2, and 4 can be applied to linearly transformed variables (e.g., $B/T$, $B/B_0$) directly. Equations 3 or 5 can be applied to variables transformed by a constant multiplicative factor, but cannot be applied directly to variables involving a translation of the axes.

It is possible to combine results over assays only when the curves are stable with regard to their position, in terms of both the X and Y axis, and with regard to the magnitude of the scatter and the dependence of the scatter on the position on the dose response curve. In the present studies, we have “pooled” results over assays, on the basis of the original response variable (Y). Alternatively, one could pool the results over assays based on the dose levels (25, 19). Under certain assumptions the method used here is superior, although in practice either of these methods should be satisfactory. It is a moot point as to whether it is better to pool results over assays (e.g., Figures 3–6) and use analyses in terms of raw counts or to use $B/B_0$. Certainly if total counts, $(B/T)_{m}$, or $B_0$ changes appreciably between assays, then it is highly doubtful that one can pool results over assays (either in terms of $s^2$, Y, or in terms of $a_0$, $a_1$, $a_2$). Likewise, if one is to pool results over assays when using $B/B_0$ as the response variable, the assay system should be “stable”. However, analyses in terms of $B/B_0$ should be less affected by changes in total counts, counting time, or specific activity of the tracer.
References


