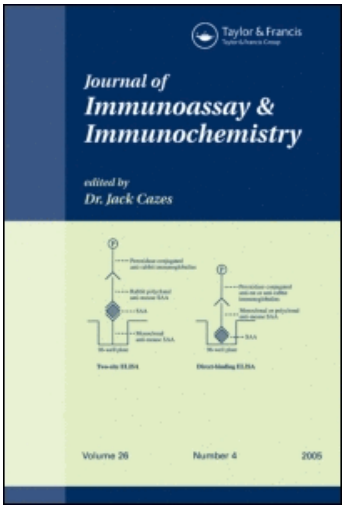

Journal of Immunoassay and Immunochemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597271

Estimation of Lot-to-Lot Relative Potency in the Preparation of Radioimmunoassay Calibrators

John H. Livesey[a]; Timothy G. Yandle[b]; Margaret J. Evans[a]; M. Jane Ellis[a]
[a] Department of Endocrinology, Christchurch Hospital, Christchurch School of Medicine and Health Sciences, Christchurch, New Zealand [b] Department of Medicine, Christchurch School of Medicine and Health Sciences, Christchurch, New Zealand

Online publication date: 28 July 2003

To cite this Article Livesey, John H. , Yandle, Timothy G. , Evans, Margaret J. and Ellis, M. Jane(2003) 'Estimation of Lot-to-Lot Relative Potency in the Preparation of Radioimmunoassay Calibrators', Journal of Immunoassay and Immunochemistry, 24: 3, 243 — 263
To link to this Article: DOI: 10.1081/IAS-120022935
URL: http://dx.doi.org/10.1081/IAS-120022935

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

JOURNAL OF IMMUNOASSAY & IMMUNOCHEMISTRY
Vol. 24, No. 3, pp. 243–263, 2003

Estimation of Lot-to-Lot Relative Potency in the Preparation of Radioimmunoassay Calibrators

John H. Livesey,[1,*] Timothy G. Yandle,[2] Margaret J. Evans,[1] and M. Jane Ellis[1]

[1]Department of Endocrinology, Christchurch Hospital, and
[2]Department of Medicine, Christchurch School of Medicine and Health Sciences, Christchurch, New Zealand

ABSTRACT

The aim of this study was to determine guidelines for estimating lot-to-lot differences in the potency of calibrator materials or batches of standards for radioimmunoassays. Thirty one lots of standards for thirteen different analytes were compared to the previous lot for that analyte with the relative potency computed by nine different methods. Assays were performed manually. The nine different calculation methods included non-simultaneous fitting of pairs of standard curves, full or partial simultaneous fitting, and least squares or

*Correspondence: John H. Livesey, Department of Endocrinology, Christchurch Hospital, Christchurch, New Zealand; E-mail: john.livesey@cdhb.govt.nz.

DOI: 10.1081/IAS-120022935 1532-1819 (Print); 1532-4230 (Online)

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

robust minimisation. The simultaneous methods were found superior to the non-simultaneous in minimising the variance of the relative potency estimates, while robust fitting procedures did not result in a lower variance than least-squares minimisation. The root mean square coefficient of variation for the simultaneous estimation of the relative potency by least squares was 6.1%. On this basis, it is recommended that relative potency estimations in radioimmunoassay be based on at least eight independent pair-wise standard curve comparisons. Additional guidelines for preparing and comparing batches of standards are also given.

Key Words: Calibration; Potency; Immunoassay; Standardisation; Radioimmunoassay.

INTRODUCTION

The long-term stability of the estimates of concentration obtained by radioimmunoassay and similar techniques is critically dependent on the ability to prepare successive lots of calibrators with low inter-lot variation in potency. On a practical level, for example, in a longitudinal study of a T_4 radioimmunoassay, the most disconcerting finding was the variability in successive lots of calibrator.[1] Further, it is readily demonstrated, on theoretical grounds, that the errors in preparing dilutions of a master lot of calibrator are unlikely to be negligible.[2,3] The problem is particularly acute for those analytes for which calibrators of essentially 100% purity are not available and, hence, determination of amount by simple physical methods, such as weighing, is not possible. This includes many peptides and proteins. Although radioimmunoassays and similar limited reagent methods have been largely supplanted in clinical laboratories by automated excess-reagent methods, radioimmunoassay methods are still very important in research laboratories such as ours because they are relatively easy to set up for new analytes.

The most common approach to minimising the inter-lot variability of calibrators appears to be to calibrate each lot against a previous lot. This can be the immediately previous lot[1] or a master lot.[4] However, despite its importance, there appears to be little published consideration of how to do this in a statistically adequate manner. Sadler and co-workers[1] compared the old and new lots of calibrators in at least three successive radioimmunoassays, but their study demonstrated that this was inadequate as the potency of successive lots varied by up to 5% for T_4 concentrations within the reference range and the total drift due

to calibration variation over 29 months ranged from 6% at high concentrations to 18% at low concentrations.

The aims of this work were, firstly, to compare procedures for estimating the relative potency of different lots of peptide radioimmunoassay standards using real radioimmunoassay data. Then, secondly, on this basis, to develop a statistically-based procedure for estimating the potency difference between lots of calibrators that has adequate precision to control long-term drift in calibration. In particular, guidelines for the appropriate number of lot-comparison assays were sought and the potential benefit of using robust estimation methods was examined.

METHODS

Notation

For convenience in the following description of methods, we denote the total number of different lots (or batches) of freshly prepared calibrator (or standard) material that were compared with a previous lot of calibrator in this study as M, with the individual new lots being denoted $1, 2, \ldots i, \ldots M$. The ith lot could be a new lot of standards for any one of a variety of analytes, for example, C-peptide or IGF-1. Each new lot (i) is to be compared to a previous lot, i', of standard material for the same analyte in order to estimate the relative potency of the new lot compared to the old.

Within the procedure for estimating the relative potency of the ith pair of lots, that is to say comparing the potency of calibrator lots i and i', a number of independent comparisons of pairs of standard curves (one constructed from the new lot and one from the old) are performed. Denote as n_i the total number of such independent pair-wise comparisons performed to estimate the relative potency of lots i and i'. Hence N, the total number of pair-wise standard curve comparisons performed, is given by

$$N = \sum_{i=1}^{M} n_i$$

Experimental Design

Preliminary simulations suggested that the number of independent pair-wise standard curve comparisons (n_i) required to determine relative

potency with reasonable precision for one lot comparison would be about eight. Consequently, the experimental approach adopted in most cases was to perform, for each lot comparison, four separate assays each containing two instances of the old and new lots of calibrators in the design:

1. S1, QCs, S2, S2, QCs, S1
2. S1, QCs, S2, S2, QCs, S1
3. S1, QCs, S2, S2, QCs, S1
4. S1, QCs, S2, S2, QCs, S1

where S1 is a standard curve constructed from the old lot of calibrator and S2 is a standard curve constructed from the new lot of calibrator. QCs are quality control specimens and the numbers 1 to 4 refer to the four independent assays. Taking the four assays together, this design furnishes eight independent comparisons between S1 and S2 and is balanced as regards the order of S1 and S2. For various practical reasons, in some cases, greater or fewer than four assays of balanced comparisons were made, or unbalanced comparisons (that is to say just S1, S2) were incorporated into routine clinical assays. All standard concentrations were run in duplicate and standard curves included from 6 to 11 different concentrations in addition to the zero standard, which was run in quadruplicate.

Preparation of Standards

Dilutions of calibrators for use in standard curves were prepared by diluting the concentrated calibrator by weight, rather than by volume, in order to increase precision.

Radioimmunoassays

Brief details of the radioimmunoassays for aldosterone, AVP, BNP, C-peptide, glucagon, IGF-1, N-terminal BNP, and VIP have been published,[5] as have the procedures for adrenomedullin[6] and ANP.[7] Renin activity was measured by an antibody trapping assay based on that of Nussberger and colleagues[8] and endothelin by a radioimmunoassay based on that of Kitamura and co-workers.[9] The radioimmunoassay for N-terminal ANP was similar to that for N-terminal BNP except that NT-ANP (1–30) was used as standards and the antiserum

(Peninsular Labs., San Carlos, CA) was raised against NT-ANP. All assays were performed manually.

Relative Potency Estimation

The relative potency (or potency ratio) of lots of calibrators was estimated by nine different procedures based on four different basic methods. In methods 1 and 2, each pair of standard curves (one prepared from the previous lot and one from the new lot) were fitted simultaneously by the five-parameter logistic equations:

$$B_1 = d + \frac{(a-d)}{(1+\beta x_1^{\lambda})^{\mu}} \tag{1}$$

$$B_2 = d + \frac{(a-d)}{(1+\beta (r x_2)^{\lambda})^{\mu}} \tag{2}$$

where B is the fraction of tracer bound to antibody, x is the concentration of calibrator, and a, d, β, λ, μ and r are adjustable parameters. Equation (1) was fitted to the standard curve given by the previous lot of calibrators and Eq. (2) to that given by the new lot. The fitted value of parameter r was then an estimate of the relative potency of the new lot of calibrator compared to the old. Parameters a and d approximate the maximum and minimum binding, respectively, for the standard curves. Finney[10] has shown that the five-parameter logistic is a good approximation to RIA standard curve shapes and the advantages of the simultaneous analysis of sigmoidal curves are outlined by DeLean and colleagues.[11]

Equations (1) and (2) were fitted to the standard curve data by the downhill simplex method[12] using a purpose-written computer program (available from the authors). Downhill simplex minimisation was used so that the fitting could be readily performed using a variety of minimisation criteria.[12] If z represents the weighted residual after curve fitting, the three criteria used were the minimisation of z^2 (least squares), $|z|^{1.25}$ and $\log(1+z^2/2)$.

The attempt was made initially to use the absolute deviation, $|z|$, as a minimisation criterion, since it may be more robust against outliers than least squares and, if the residuals are distributed as a double exponential, it is the optimal criterion.[12] However, since convergence proved to be poor for $|z|$ for the present fitting problem, an exponent of 1.25 was added. Convergence proceeded smoothly for $|z|^{1.25}$ and this criterion

will be referred to as the modified mean absolute deviation (MMAD) fitting.

The minimisation of $\log(1+z^2/2)$ was used an alternative robust estimator as it is optimal for a Cauchy distribution of the residuals and the Cauchy distribution has even more extensive tails than the double exponential.[12] This criterion will be referred to as Cauchy fitting.

For the fitting of the logistic functions, the residuals were weighted using the radioimmunoassay variance function (or response-error relation):

$$\hat{\sigma}^2 = h_0 + h_1 B^{1.4} \tag{3}$$

where B is the fractional binding of tracer to antibody corresponding to the residual and $\hat{\sigma}^2$ is the estimated variance of the residual. The parameters h_0 and h_1 of the variance function were determined by iterative weighted linear regression on all the squared differences between the duplicate fraction bound values within the relative potency determination assay. The exponential form of Eq. (3) follows the recommendation of Dudley and co-workers[13] and the exponent of 1.4 is that used in this laboratory for all radioimmunoassay calculations.

Outlying duplicates were rejected by calculating the kurtosis of the standardised differences between duplicate values and successively deleting the duplicates with the largest standardised differences until the kurtosis was less than 5. Standardisation of the differences between duplicate values was by division by $\hat{\sigma}$.

In the first method (Method 1) for estimating r, all the parameters in Eqs. (1) and (2) other than r were forced to take the same values in both equations during the curve fitting. Or, in other words, the two standard curves were assumed to be of identical shape, but to differ in their positions on the x axis.

The second method (Method 2) for estimating r was the same as Method 1 except that the value of a, as well as r, was allowed to differ between Eqs. (1) and (2). This allowed for any drift of the maximum binding within the assay.

The third method (Method 3) was to use the 'optimal' least squares estimate. That is to say, using the least squares estimate from Method 1 unless the least squares estimate by Method 2 was significantly better (by F test of the sums of squares of the residuals) in which case the latter estimate was used.

The fourth method (Method 4) for estimating r was to fit Eq. (1), independently, to all standard curves and calculate the ED_{50} for each standard curve. The ED_{50} is the concentration of standard (x) for

which the binding (B) equals $(a+d)/2$. The estimate of r for a pair of standard curves is then taken in this method to be the ratio of the ED_{50} values. Method 4A was to simply fit Eq. (1), method 4B was to fit Eq. (1) then, if there was still significant deviation of the residuals from the fitted line, to fit sequential straight-line segments.[14]

In all, nine different procedures for estimating relative potency were performed for each pair of standard curves, namely Method 1 with three different criteria for minimising the residuals (denoted 1LS, 1MMAD, and 1Cauchy), Method 2 with the three different minimisation criteria (denoted 2LS, 2MMAD and 2Cauchy), Method 3 (denoted 3LS), and Methods 4A and 4B.

For each lot comparison for an analyte (comparison of the ith lot to the i'th lot), the n_i independent comparisons of pairs of standard curves resulted in n_i estimates of the value of r for each of the nine different estimation methods. The mean ($\bar{r}$) and standard error of the mean (*sem*) for each method was calculated from the n_i values of r for that method. The hypothesis that $\bar{r}=1$ was tested using the single sample Student's t test and the coefficient of variation for the estimates of r (denoted CV_r) for each of the nine methods was also calculated.

Parallelism

Parallelism of the standard curves S1 and S2 was checked by performing a preliminary run of the simultaneous fitting of Eqs. (1) and (2) to S1 and S2, respectively, with $\mu=1.0$, while allowing the value of the parameter λ to differ between Eqs. (1) and (2). Only λ and r differed between Eqs. (1) and (2) (except that, when Method 2 was used, a differed also). The ratio of the two values of λ, here denoted by s, was taken to be a measure of the similarity of the slopes of the two standard curves.

In a similar manner to the estimation of relative potency, s was estimated by seven different procedures (the two ED_{50} methods were not applicable). The mean ($\bar{s}$) and its *sem* was calculated from the n_i values of s and the hypothesis that $\bar{s}=1$ was tested using the single sample Student's t test. The coefficient of variation for the estimates of s (denoted CV_s) for each of the seven methods was also calculated.

Weighted Coefficient of Variation Estimates

The weighted root mean square of the M independent estimates of CV_r for each of the nine estimation methods were calculated using

weights of $n_i - 1$ and the weighted 25th percentile, median, and 75th percentiles for CV_r calculated using weights of n_i. Weighed summary statistics were calculated similarly for the estimates of s.

Comparison of Variances

For each of the M lot comparisons, the variances of the estimates of relative potency obtained by the nine different procedures, each applied n_i times to independent pairs of standard curves, were compared using the median version of the Levene test for equality of variance.[15] In order to combine the data from the lot comparisons, the M different probabilities (p_i) that the null hypothesis (all variances equal) was false, were combined using the weighted Stouffer procedure.[16] In brief, each p_i is converted to a normal distribution Z score and the Z scores summed over the M lot comparisons using weights of n_i. The weighted sum is divided by the square root of M and the resulting Z score converted back to a probability value.

To determine which of the nine methods of computing the relative potency showed the greatest variance in the estimates, the variance of each method was compared in turn to that of the 'optimal' least-squares (LS) estimate, again using the Levene test. The hypothesis tested in each case was that the 'optimal' LS method had the lower variance of the two. The resulting one-sided probabilities were combined over all the M lot comparisons using the weighted Stouffer procedure.

Comparison of Mean Estimates

For each of the N pairs of standard curve comparisons (old lot versus new lot), estimates of the potency ratio by methods 1LS, 1MMAD, 1Cauchy, 2LS, 2MMAD, and 2Cauchy were divided by the value of the potency ratio estimated by method 3LS. The significances of the differences between the means of each of these six ratios and unity were determined using the one-sample Student's t test.

Assay Characteristics and Correlations

For the purpose of summarising assay characteristics and performing correlation studies, the variability of the duplicates within an assay was taken to be the variance calculated from Eq. (3) for a

fractional binding of tracer to antibody of 0.25. Lot standard deviations (denoted $SD_{0.25}$) were the root mean square of the n_i within-lot variances estimated for each pair-wise comparison for the lot. The binding of tracer to antibody at zero standard concentration (B_0) was expressed a fraction of the total tracer activity added to each assay tube. Correlations were performed by the Spearman (non-parametric) procedure.

Sample

All comparisons of lots of calibrators made in our laboratory over a period of three years were included in this study.

RESULTS

Sample

New lots of standards had been prepared on 31 occasions, in total, five times for BNP, four times for C-peptide, three times each for IGF-1, N-terminal BNP and VIP, twice each for ANP, adrenomedullin, aldosterone, AVP, and glucagon, and on one occasion each for endothelin, N-terminal ANP, and renin activity. From the potency estimation assays for these 31 lots, a total of 233 independent pairs of standard curves (old lot versus new lot) were available for study. Thus, in terms of the notation in the Methods section, M was 31 and N was 233. The mean value of n_i was 7.52 (minimum 4, maximum 11).

Assay Variability and Maximal Binding

The weighted RMS of the $SD_{0.25}$ values for all 31 lot comparisons was 0.00653 (median 0.00535, quartiles 0.00463, 0.00670) and the weighted mean of the B_0 values was 0.440 (median 0.413, quartiles 0.351, 0.509).

Mean Relative Potency Estimates

The means and medians of the ratios of the method 1 and method 2 estimates to the method 3LS estimate are given in Table 1 for relative

Table 1. Mean and median ratios of potency estimates by methods 1 and 2 to the potency estimate by method 3LS. Student's t and P relate to the differences between the means and unity.

	Method					
	1LS	1MMAD	1Cauchy	2LS	2MMAD	2Cauchy
Mean	0.9914	0.9900	0.9908	0.9971	0.9957	0.9973
SEM	0.0023	0.0024	0.0026	0.0015	0.0020	0.0023
N	233	233	233	233	233	233
t	3.78	4.08	3.46	1.94	2.13	1.19
P	< 0.0001	< 0.0001	0.0003	0.026	0.017	0.12
Median	1.0000	0.9943	0.9960	1.0000	0.9968	0.9974

Table 2. Summary of weighted coefficients of variation (CV_r) for estimates of relative potency (r) by different methods for 31 lot comparisons.

	Method								
	1LS	1MMAD	1Cauchy	2LS	2MMAD	2Cauchy	3LS	4A	4B
RMS	6.10	6.06	6.15	6.22	6.39	6.51	6.07	16.98	16.65
Median	5.10	4.89	4.70	5.35	5.38	5.16	5.47	8.97	9.14
25th centile	3.08	3.46	3.25	3.69	4.33	4.09	3.50	5.28	6.09
75th centile	6.84	7.20	7.02	6.69	6.89	6.92	6.67	19.50	19.59
IQR	3.76	3.74	3.77	3.00	2.56	2.83	3.17	14.22	13.50

RMS = root mean square, IQR = interquartile range.

potency estimates. Only the mean of the ratio of the 2Cauchy estimator to the 3LS estimator (0.997) did not differ significantly from unity. Although statistically significantly different ($p = 0.026$ or less, $n = 233$), the ratios of the estimates by the other methods to the 3LS estimates were, nevertheless, also close to unity (range 0.990 to 0.997).

Coefficients of Variation of Relative Potency Estimates

The weighted estimates of the root mean square (RMS) values of CV_r for the nine different methods for estimating r, along with the weighted median and the 25th and 75th percentiles are given in Table 2. The

RMS values range from 6.06 to 6.51 except that for the two method 4 procedures the RMS CV_r was 17.0 and 16.6. The median CV_r values were lower than the RMS CV_r values for all methods.

Comparison of the Variances of Relative Potency Estimates

The combined Levene test probabilities for the variance comparisons from the 31 lots of standards showed that the variances of the nine estimators for the relative potency (r) differed significantly ($Z = 8.13$, $p < 0.0001$) between the different estimation methods.

Comparison of the variances of the other eight estimators of r to the variance of the 'optimal' LS method (method 3LS) in pairwise Levene tests found that the variance of the ED_{50}-based methods (4A and 4B) was significantly greater than that of the 'optimal' LS method ($Z = 5.36$ and 5.79 respectively, both $p < 0.0001$). In contrast, the variance of none of the other six methods taken individually was significantly greater than that of the 3LS method. The most negative Z value of the six was -0.64 for the 1LS estimator ($p = 0.26$) and the most positive 1.32 for the 2C estimator ($p = 0.094$).

Applying the Levene test for homogeneity of variance only to the potency ratio estimates by the seven non-ED_{50}-based methods (3LS and the LS, MMAD and Cauchy variants of 1 and 2), and combining the thirty one p values as above, found that the variances of the estimates did not differ significantly between these seven methods ($Z = 0.90$, $p = 0.19$).

The variance of the 1LS estimator did not significantly exceed that of the 2LS estimator when just these two methods were compared by Levene test ($Z = -0.25$, $p = 0.40$).

Mean Slope Ratio Estimates

The means and medians of the ratios of the method 1 and method 2 estimates to the method 3LS estimate are given in Table 3 for the slope ratio (s). For each of the Method 1 estimators (1LS, 1MMAD, 1Cauchy) the means of the ratio to the 3LS estimator differed significantly from unity ($\bar{s} = 0.995$, 0.994, 0.995, $p = 0.0007$, 0.0005, 0.008 respectively, all $N = 233$), whereas, the ratios of the Method 2 estimators to the 3LS estimator did not differ significantly from unity. The differences from unity for the Method 1 estimators were, however, small, being 0.6% or less.

Table 3. Median and mean ratios of slope ratio estimates by methods 1 and 2 to the slope ratio estimate by method 3LS. Student's t and P relate to the differences between the means and unity.

	Method					
	1LS	1MMAD	1Cauchy	2LS	2MMAD	2Cauchy
Mean	0.9954	0.9943	0.9952	0.9998	0.9979	0.9994
SEM	0.0015	0.0017	0.0020	0.0010	0.0015	0.0020
N	233	233	233	233	233	233
t	3.18	3.32	2.40	0.18	1.40	0.29
P	0.0007	0.0005	0.008	0.43	0.081	0.39
Median	1.0000	0.9980	0.9967	1.0000	0.9990	0.9990

Table 4. Summary of coefficients of variation (CV_s) for estimates of slope ratio (s) by different methods for 31 lots weighted according to the number of pair-wise comparisons within each lot comparison.

	Method						
	1LS	1MMAD	1Cauchy	2LS	2MMAD	2Cauchy	3LS
RMS	4.39	4.55	4.70	5.18	5.29	5.48	4.52
Median	3.97	4.08	4.23	3.99	4.30	4.93	3.59
25th centile	2.26	2.49	2.10	2.46	2.85	2.78	2.80
75th centile	5.41	5.29	5.58	6.51	6.19	6.04	5.59
IQR	1.44	1.21	1.35	2.52	1.89	1.11	2.00

RMS = root mean square, IQR = interquartile range.

Coefficients of Variation of Slope Ratio Estimates

The weighted estimates of the RMS value of CV_s for the different methods for estimating s, along with the weighted median and the 25th and 75th percentiles are given in Table 4 and range from 4.4 to 5.5%. The median CV_s values for the slope ratio were all lower than the corresponding RMS values.

Comparison of the Variances of Slope Ratio Estimates

Applying the Levene test for homogeneity of variance to the slope ratio estimates for each lot by the seven methods (3LS and

the LS, MMAD and Cauchy variants of 1 and 2), and combining the thirty one p values as above indicated that the variances of the estimates did not differ significantly between the seven estimators ($Z = 1.13$, $p = 0.13$).

Comparison of the variances of the other six estimators of r to the variance of the 'optimal' LS method (method 3LS) in pairwise Levene tests found that the variance of none of the other six methods taken individually was significantly greater than that of the 3LS method. The most negative Z value of the six was -0.92 for the 1LS estimator ($p = 0.18$) and the most positive 1.07 for the 2C estimator ($p = 0.14$).

The variance of the 1LS estimator was significantly lower than that of the 2LS estimator when just these two methods were compared by the Levene test ($Z = -1.98$, $p = 0.024$).

Correlations

There was no significant correlation between the mean within-lot variability of duplicates within assays ($SD_{0.25}$) and the coefficients of variation of either the 3LS estimates of potency (CV_r) or slope (CV_s) ($r = 0.05$, 0.13; $p = 0.77$, 0.48, $n = 31$, 31, respectively). However, despite a tendency for $SD_{0.25}$ to increase with B_0 ($r = 0.32$, $p = 0.083$), both CV_r and CV_s were significantly negatively correlated with B_0, the maximum standard curve binding, ($r = -0.49$, -0.47, $p = 0.0056$, 0.0072, $n = 31$, 31 respectively), as shown in Fig. 1.

DISCUSSION

This study has been performed using real, rather than simulated, data so that conclusions can be based on the actual rather than the idealised variability of endocrine radioimmunoassays as performed under routine conditions in the research and clinical laboratory. The standard curve comparisons were performed by eight different technicians with varying experience in radioimmunoassay and the assay procedures used included both polyethylene glycol precipitation and solid-phase second antibody for separating 'bound' from 'free' tracer. Hence, the results are likely to be relevant to assays performed by a variety of staff and methods.

The ED_{50}-based methods were included in the study because they might have constituted a quick way of estimating relative potency

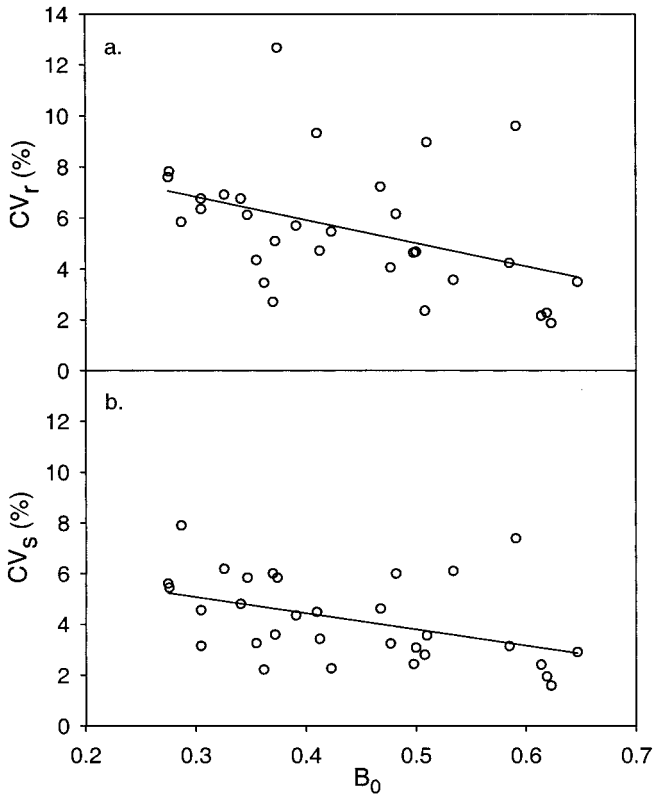

Figure 1. Relationship between the maximal standard curve binding (B_0) and the coefficient of variation of the mean 3LS estimate for (a) the relative potency (CV_r) and (b) the slope ratio (CV_s). Correlations coefficients are (a) -0.49 ($p = 0.0056$, $n = 31$) and (b) -0.47 ($p = 0.0072$, $n = 31$). The lines show the linear regressions of CV_r and CV_s on B_0.

from routine RIA calculations without having to run data through a separate formal potency comparison procedure. However, we found the ED_{50}-based methods to have a significantly greater variability in the estimation of relative potency than did the 'optimal' least-squares method using the Levene test. This is presumably because two entirely independent fits of Eq. (1) have, between them, 10 adjustable parameters, whereas, in the simultaneous fits (methods 1, 2, and 3), there are only five or six adjustable parameters. The fewer degrees of freedom in the latter estimates probably result in the greater

precision. Consequently, the present work indicates that less variable potency estimates can be obtained by constraining all or most parameters (apart from relative potency) to be identical for the two standard curves being compared.

The mean relative potency estimates by method 3 and the variants of methods 1 and 2 were all very similar, with the mean method 1 and 2 estimates differing by 1 percent or less from the mean 3LS estimates (Table 2). Although all these differences, except for method 2Cauchy, reached statistical significance (Table 2), they are probably too small to be of practical importance.

The finding that the variability of the potency estimates by method 3 and the variants of methods 1 and 2 do not differ significantly between these seven methods suggests, firstly, that there is no advantage in using robust fitting procedures (the MMAD and Cauchy procedures) compared to least-squares and, secondly, that for the least-squares procedures, the variability of the potency estimates is similar whether strict simultaneous fitting is used (1LS), whether B_0 is assumed to vary (2LS) or whether an optimal combination of these is used (3LS). It, thus, appears for our radioimmunoassays that the distribution of the residuals (z) after fitting Eqs. (1) and (2), and the variability of B_0, are such that neither the type of residual minimisation technique used nor the variant of least-squares used is critical for the estimation of relative potency.

The mean slope ratio estimates by the seven different procedures were all very similar (Table 3). The variants of methods 1 and 2 differed from those by method 3LS by less than 0.6% and, although statistically significant differences were found (Table 3), they are probably too small to be of practical importance.

The Levene test finding that there was no significant difference between the variances of the seven estimators of relative slope suggests again that the distribution of the residuals (z) in the curve fitting for real radioimmunoassays is such that robust procedures offer no advantage over least-squares procedures. Among the least-squares procedures for relative slope estimation, while neither the variance of the 1LS or the 2LS procedure differed significantly from that for the 3LS procedure, a direct comparison of the variances of the 1LS and the 2LS methods suggested that the former was preferable to the latter. It, thus, appears that the slope ratio is best estimated by either the 1LS or 3LS procedure.

Our observation that the variance of estimates of both relative potency and the slope ratio are negatively correlated with the standard curve binding for the zero standard ($r = -0.49$ and -0.47, respectively,

Fig. 1) suggests that care should be taken to adjust assay conditions so that B/B_0 is at least 0.5 in order to obtain the most precise estimates. The virtual lack of a relationship between the variance of the assay duplicates (as indicated by $SD_{0.25}$) and the variance of the potency and slope estimates ($r = 0.05$ and 0.13, respectively), is somewhat surprising and suggests that factors other than the goodness of assay duplicates largely determine the variability of the estimate of relative potency.

To determine guidelines for the practical estimation of the relative potency of lots of standard materials, it is reasonable to assume that the estimate of the relative potency of a new lot, compared to the previous lot, should be within 5% of the true value 95% of the time. For peptide radioimmunoassays in endocrinology, errors smaller than this are likely to be of little research or clinical consequence. For the relatively small number of pair-wise standard curve comparisons likely to be made, say about eight, Student's t is approximately 2.3. Hence, the coefficient of variation for the mean of the eight individual estimates must be approximately 5/2.3 ($= 2.2\%$) if the above criterion is to be met. The RMS of the CV_r values found in this study is 6.1% for the 1LS and 3LS methods (Table 1) so that n independent potency estimates must be averaged where

$$2.2 = \frac{6.1}{n^{0.5}} \tag{4}$$

From Eq. (4), n is about 7.7, indicating that as a rule-of-thumb eight independent comparisons should usually be performed when estimating relative potencies. Since Table 4 shows that the variance of slope ratio estimates is generally less than that of relative potency estimates (Table 2), eight independent standard curve comparisons should also suffice for estimating the slope ratio with adequate precision.

Recently, it has been suggested that at least 20 comparison assays should be run when calibrating secondary standards against primary standards[17] but, on the basis of our work, this number of comparisons is probably excessive if the preparative and computational methods described here are used. The same author[17] suggested the hand-drawing of calibration curves, which may account for the high number of comparisons recommended.

For analytical systems, possibly automated, where a CV_r of better than 6.1% is routinely obtainable, fewer than eight independent relative potency estimates will be required to obtain a 95% confidence of the mean estimate being within 5% of the true value. In this case, the analyst can choose either to perform less than four comparison assays, if this

level of accuracy is acceptable, or to perform all four (or more) comparison assays to increase the accuracy of the estimate.

On the basis of this study and our experience, we make the following recommendations for preparing peptide standards and estimating the relative potency of different lots of calibrator material.

1. If weighing out peptide in powder form, it should be dried before use if it has not been stored with desiccant. Freeze-drying will rapidly reduce the water content.
2. Where the standard material is sufficiently stable, prepare a sufficient number of aliquots of concentrated stock standard solution to last many years. Store these at $-40^{\circ}C$ or $-80^{\circ}C$ in suitable air-tight, largely filled vials or freeze-dried.
3. Make dilutions of working standards as required from the stock aliquots using as few serial dilutions as practicable; a fewer number of higher dilutions are more accurate and precise than a larger number of smaller ones.
4. The number of different concentrations comprising the standard curve should be between 8 and 12 (including zero).
5. The diluent must contain sufficient inert protein or surfactant to prevent significant losses by adsorption.[18]
6. Use weighing rather than volumetric procedures for preparation and dilution of standards; the precise concentrations of the working standards can then be calculated from the record of the weighings.
7. The lots of working standards should be as large as practicable, generally lasting at least a year at the usual rate of performing assays. Depending on stability, store the working standards at $-20^{o}C$ or below. Again use air-tight, largely filled vials.
8. To estimate the potency of the newly diluted lot of standards compared to the previous lot, set up four radioimmunoassays of the balanced design described above under Experimental Design.
9. For each of the eight independent pairs of standard curves, compute the slope ratio and the relative potency using the 1LS or the 3LS procedure.
10. Using the one-sample Student's t test compare the mean of the eight slope ratio estimates to 1.000. If the mean slope ratio does not differ significantly from unity, proceed to the next step.
11. Using the one-sample Student's t test, compare the mean of the eight relative potency estimates to 1.000. If it does not differ

significantly from unity, the new lot of standards can be considered to be equipotent with the previous lot.

12. If the mean relative potency estimate differs significantly from unity, the nominal concentration values of the new lot of standards must be adjusted by multiplying them by the estimated mean relative potency.
13. The coefficient of variation of the mean relative potency should ideally be about 2.2% or less in order to obtain adequate precision. Alternatively, a CV of 2.7% or less will give 90% confidence that the mean is within 5% of the true value. If the CV of the mean of the eight relative potency estimates initially obtained is greater than the desired CV of 2.2 or 2.7%, additional balanced comparisons must be performed until the CV of the mean relative potency is sufficiently low.

Adoption of these recommendations when preparing standards for peptide radioimmunoassays should help to reduce lot-to-lot variation in potency and, thus, reduce the long-term variability and drift in results obtained by radioimmunoassay techniques. This is important for large scale or long-term research studies, for adherence to previously established reference ranges and for reducing that portion of total measurement uncertainty that is due to uncertainty in the calibration traceability chain.[19]

CONCLUSIONS

In summary, this work suggests that when calculating the relative potency of standard materials for peptide radioimmunoassay:

1. Simultaneous fitting of two standard curves gives less variable estimates of potency than do calculations based on independently fitted standard curves.
2. Robust fitting techniques offer no advantage over least-squares for potency estimation.
3. Strict simultaneous least-squares fitting, or allowing B_0 to vary, or a combination of these, are equally efficient for estimating relative potency.
4. The slope ratio is best estimated by either the 1LS or 3LS procedures with robust techniques offering no benefits.
5. The binding of tracer to antibody (B/B_o) should be at least 50% to obtain the most precise relative potency estimates.

6. Factors other than the variance of duplicates within an assay largely determine the variance of potency and slope estimates.
7. At least eight independent comparisons are probably desirable when estimating relative potency.

SYMBOLS

β	Scale parameter in logistic function (Eqs. (1) and (2))
λ	Slope parameter in (Eqs. (1) and (2))
μ	Asymmetry parameter in logistic function (Eqs. (1) and (2))
$\hat{\sigma}^2$	Variance of duplicates within an assay
a	Parameter in logistic function (Eqs. (1) and (2)). Approximates B_0
B	Fraction of tracer bound to antibody
B_0	Value of B when $x = 0$
CV_r	Coefficient of variation of r
CV_s	Coefficient of variation of s
d	Parameter in logistic function (Eqs. (1) and (2))
ED_{50}	Value of x for which $B = (a + d)/2$
h_1, h_2	Parameters in variance function (Eqs. (3))
M	Total number of new lots of calibrators Σ_i
i	The ith new lot of calibrator
i'	The previous (or master) lot of calibrator with which the ith lot is to be compared
N	Total number of pair-wise standard curve comparisons (Σn)
n_i	Number of pair-wise standard curve comparisons for lots i and i'
p_i	Probability that the variance of two or more estimators of s or r are the same for the ith lot
r	Relative potency of calibrator lots i and i'
$\bar{r}$	Mean of n_i values of r
s	Slope ratio (the ratio of the λ values in equations 1 and 2)
$\bar{s}$	Mean of n_i values of s
$SD_{0.25}$	Within-lot standard deviation of assay duplicates at $B = 0.25$
sem	Standard error of the mean
x	Concentration of calibrator

z	Residual in curve fitting
Z	Normal distribution Z score

ACKNOWLEDGMENT

This work was supported by the Canterbury District Health Board.

REFERENCES

1. Sadler, W.A.; Smith, M.H.; Murray, L.M.; Turner, J.G. A pragmatic approach to estimating total analytical error of immunoassays. Clin. Chem. **1997**, *43*, 608–614.
2. Gaines Das, R.E. Dilution as a source of error: implications for preparation and calibration of laboratory standards and for quality control of radioimmunoassay. Clin. Chem. **1980**, *26*, 1726–1729.
3. Higgins, K.M.; Davidian, M.; Chew, G.; Burge, H. The effect of serial dilution error on calibration inference in immunoassay. Biometrics **1998**, *54*, 19–32.
4. Hölzel, W. Analytical variation in immunoassays and its importance for medical decision making. Scand. J. Clin. Lab. Invest. **1991**, *51* (suppl. 205), 113–119.
5. Evans, M.J.; Livesey, J.H.; Ellis, M.J.; Yandle, T.G. Effects of anticoagulants and storage temperatures on stability of plasma and serum hormones. Clin. Biochem. **2001**, *34*, 107–112.
6. Lewis, L.K.; Smith, M.W.; Yandle, T.G.; Richards, A.M.; Nicholls, M.G. Adrenomedullin(1–52) measured in human plasma by radioimmunoassay—plasma concentration, adsorption, and storage. Clin. Chem. **1998**, *44*, 571–577.
7. Yandle, T.G.; Espiner, E.A.; Nicholls, M.G.; Duff, H. Radioimmunoassay and characterization of atrial natriuretic peptide in human plasma. J. Clin. Endocr. Metab. **1986**, *63*, 72–79.
8. Nussberger, J.; Fasanella d'Amore, T.; Porchet, M.; Waeber, B.; Brunner, D.B.; Brunner, H.R.; Kler, L.; Brown, A.N.; Francis, R.J. Repeated administration of the converting enzyme inhibitor cilazapril to normal volunteers. J. Cardiovasc. Pharm. **1987**, *9*, 39–44.
9. Kitamura, K.; Tanaka, T.; Kato, J.; Eto, T.; Tanaka, K. Regional distribution of immunoreactive endothelin in porcine tissue: abundance in inner medulla of kidney. Biochem. Biophys. Res. Comm. **1989**, *161*, 348–52.

10. Finney, D.J. Response curves for radioimmunoassay. Clin. Chem. **1983**, *29*, 1762–1766.
11. Delean, A; Munson, P.J.; Rodbard, D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. Am. J. Physiol. **1978**, *235*, E97–102.
12. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes in C*, 2nd Ed.; Cambridge University Press: Cambridge, UK, 1992.
13. Dudley, R.A.; Edwards, P.; Ekins R.P.; Finney, D.J.; McKenzie, I.G.; Raab, G.M.; Rodbard, D. Guidelines for immunoassay data processing. Clin. Chem. **1985**, *31*, 1264–1271.
14. Livesey, J.H. Computation of radioimmunoassay data using segmentally linearized standard curves. Comput. Biomed. Res. **1974**, *7*, 7–20.
15. *NIST/SEMATECH e-Handbook of Statistical Methods*; http://www.itl.nist.gov/div898/handbook/ (accessed Oct 2002).
16. Rosenthal, R. *Meta-Analytic Procedures for Social Research*; Sage: Beverly Hills, CA, 1984; 93.
17. Wild, D. Standardization and calibration. In *The Immunoassay Handbook*, 2nd Edn.; Wild, David, Ed.; Nature Publ. Grp.: New York, 2001; 140–148.
18. Livesey, J.H.; Donald, R.A. Prevention of absorption losses during radioimmunoassay of polypeptide hormones: effectiveness of albumins, gelatin, caseins, Tween 20 and plasma. Clin. Chim. Acta **1982**, *123*, 193–198.
19. Kristiansen, J. Description of a generally applicable model for the evaluation of uncertainty of measurement in clinical chemistry. Clin. Chem. Lab. Med. **2001**, *39*, 920–931.

Received December 1, 2002
Accepted January 8, 2003
Manuscript 3084

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016