Ann. Inst. Statist. Math.
29 (1977), Part A, 467-477

ESTIMATION OF THE RESPONSE CURVE IN RADIOLIGAND ASSAYS

D. J. FINNEY

(Received Oct. 25, 1976)

1. Introduction and summary

For a wide range of substances, techniques of radioimmuno- and immunoradio-metric assays (RIA and IRMA respectively) have come into use because of specificity, speed, low costs, and high precision. The literature is extensive, especially in its accounts of different techniques but also in proposals for statistical treatment of results. I shall not attempt to summarize the immunological and radiochemical principles. The statistician or biometrician who wants further information at this level should start from a few key publications (such as Arrigucci *et al.* [1]; Ekins and Newman [4]; Midgley [10]; Midgley *et al.* [11]; Rodbard [14]; Rodbard and Frazier [16]).

Instead I content myself with a brief description of the data obtained in the course of such a radioligand assay, in a form understandable by those who are not familiar with the method. These data consist of counts of radioactivity (in an agreed fixed time), relating to the "bound" or the "free" portion of an antigen (in RIA) or antibody (in IRMA) that was labelled with a suitable radio-isotope before the assay began. I formulate the relation between expected count and dose of the standard preparation, and discuss the variance of individual counts about this expectation. I compare three principles that might be applied to estimate parameters, and apply them to estimating potencies for test samples. Finally I comment briefly on computer programs for the comprehensive analysis of data from radioligand assays.

2. The data

A radioligand assay involves radioactivity counts at several doses of a standard preparation, with independent replicates (usually 3 to 8) at each dose. The doses can include zero and also what is logically an infinite dose giving a non-specific or background count. For RIA, the expectations of "bound" counts decline from a maximum at zero dose

to a limiting minimum as the dose becomes very large; for IRMA the trend is from a minimum at zero dose to a limiting maximum. "Free" counts behave in a complementary manner. Of course, replicate counts at a dose will vary about their expectation. Individual counts will usually be fairly large, say at least 50 and possibly ranging as high as 20000; in the light of experience of an assay technique, counting time will be chosen to avoid the very small counts at which discreteness becomes important and also the very large that may imply wastage of time and resources.

An assay includes counts on test samples, the object being to use the counts on each sample for estimating its potency in terms of the standard. The nature of the experimental techniques makes inclusion of several test samples convenient—perhaps 6 or 8, perhaps several hundred. *Multidose assays* include two or more doses of each test sample, these having known ratios (e.g. 1 : 1/2 : 1/4) even though the true potency is unknown. *Single dose assays* have only one dose of each test sample; this may be because no suitable diluent that would not distort the potency is available, because in an assay for clinical diagnoses the number of different samples must be maximized, or because of inadequate understanding of the advantages of having more than one dose.

3. Nature of response curve

The relation between dose, z, and expectation of count is the *response curve* for the standard, usually a smooth monotonic function ranging from, say, D at $z=0$ to C at $z\to\infty$. Hereafter discussion is in terms of RIA and bound counts, for which $D>C$; for IRMA, or for free counts, $D<C$ and various changes of sign and inversions of terms are required without any essential differences.

For reasons that become apparent in connexion with multidose assays, a logarithmic dose metameter,

$$x=\log_{10} z \qquad (\text{or } x=\ln z)\,, \tag{3.1}$$

has advantages. The response curve still approaches C as a lower asymptote when $x\to\infty$, but it is now also asymptotic to D as $x\to-\infty$. Write u for a single count, and define

$$U=\mathrm{E}(u\,|\,x)\,. \tag{3.2}$$

Then

$$U=C+(D-C)F(x)\,, \tag{3.3}$$

where $F(x)\to 1$ as $x\to-\infty$, $F(x)\to 0$ as $x\to\infty$. Usually $F(x)$ will have a smooth sigmoidal shape, possibly with symmetry about $F(x)=0.5$.

The common practice in RIA is to write B (in place of u) for a bound count, with B_0, N for counts at zero and infinite dose respectively (corresponding with my D, C). This notation, however, does not distinguish between observed counts and their expectations, a distinction vital to sound estimation practice, and its users have formed values of $(B_0-B)/(B-N)$ to play the role of $F(x)$ without regard to the statistical errors of all counts. The new notation adopted here is intended to emphasize a different outlook. Estimation should be examined in relation to individual counts, the truly independent observations, rather than ratios or other derived quantities that forfeit this independence.

Debate about the response curve is primarily debate on $F(x)$, with C, D playing the same role in any equation proposed. Experience in bioassay suggests that conclusions will commonly not be very sensitive to the particular form of $F(x)$. At least two parameters are necessary in $F(x)$, one a location parameter indicating the region of the x-scale over which $F(x)$ is well-removed from its asymptotes and one a scale parameter telling something of how quickly $F(x)$ changes over this region. It is of course conceivable that more than two parameters are needed, but the simplest adequate curves will be those in which $F(x)$ is a function of

$$Y=\beta(x-\mu) \tag{3.4}$$

that involves no unknown parameters other than β, μ.

A candidate curve that satisfies all requirements so far stated is

$$F(x)=1/(1+e^{-2Y}) , \tag{3.5}$$

or

$$Y=\frac{1}{2}\ln[(U-C)/(D-U)] , \tag{3.6}$$

with $\beta<0$ for RIA. As is well known, the normal integral

$$F(x)=\int_{-\infty}^{Y}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)dt \tag{3.7}$$

is qualitatively very similar to (3.5), the factor "2" in (3.5) having the merit of making the correspondence good over a wide range of Y. The logistic equation, (3.5), has been widely used for RIA in various alternative formulations, notably by Rodbard and his colleagues (Rodbard [13], and other references; see also Burger *et al.* [2] and Healy [9]); it possibly has theoretical merits as well as mathematical simplicity, but only a very large body of data could distinguish empirically between (3.5), (3.7), and other reasonable sigmoids (cf. Finney [6]).

These curves are symmetric about $x=\mu$, $F(x)=1/2$, and may on

that account be inappropriate for some assays. Asymmetry seems almost certain to need at least a fifth parameter. One proposal (Harding *et al.* [8]) can be written

$$U=a+\frac{b}{1+c\ln[1+\exp(z-d)]}, \tag{3.8}$$

where the parameters bear no simple relation to those previously discussed. Despite qualitatively satisfactory properties, (3.8) is inadequate because its functional form is not invariant under changes of scale for dose measurement. To satisfy this requirement, an extra parameter must be introduced. One possibility is

$$U=a+\frac{b}{1+c\ln[1+\exp\{(z-d)/f\}]}, \tag{3.9}$$

but others could be tried. Though not discussed further here, these do not introduce new statistical problems beyond the general difficulty of estimating as many as five parameters adequately.

For any test sample, an arbitrary unit of dose can be adopted initially, for example simply a volume or weight of material. In a multidose assay, the ratios between doses of any one test sample are known, so that all are measurable in the same arbitrary units. In the usual way of bioassay, the unknown relative potency, ρ, is defined as the amount of standard equivalent to unit dose of the sample: the expected count for a dose z units of the sample is the same as that for ρz standard units. Consequently, in a valid assay the response curve for a test sample must be obtainable by substitution of ρz for z or $(x+\log\rho)$ for x. A multidose assay permits test samples also to contribute to estimates of parameters and enables a test of "parallelism" (as in parallel line bioassay) to be used as a check on validity.

4. Variance function

Individual counts at a dose will vary about the corresponding U. One would naturally hope that the distribution of counts was Poissonian. Rodbard and Cooper [15] and Rodbard [13] presented theoretical arguments for the augmentation of Poisson variance by many other components, and empirical evidence often shows that the variance exceeds U. Rodbard and his colleagues suggested the equivalent of

$$\mathrm{Var}(u\,|\,U)=V_1U+V_2U^2 \tag{4.1}$$

as a function that makes no attempt at being exact but is in practice an adequate representation of variance. I prefer to use

$$\text{Var}\,(u\,|\,U) = VU^J\,, \tag{4.2}$$

which can be scarcely distinguishable from (4.1) over a wide range of U. Other functions, involving two or more additional parameters (like V, J), could be devised. Essentially the same problems would arise with any variance function

$$\text{Var}\,(u\,|\,U) = \phi(U)\,; \tag{4.3}$$

it seems fairly sure that $\phi'(U)>0$ and $\phi''(U)>0$ over the useful range of U, and (4.1), (4.2) seem adequate approximations.

Had the distribution been Poissonian, for data with $U>50$ one would have had little hesitation in approximating to it by a normal. The greater variance is likely to mean that discreteness of counts is more important, but since most counts are large this is unlikely to interfere seriously with normality.

5. Estimation

The parameters fall into three categories:

(i) those that define the response curve, such as C, D, β, μ in (3.3)–(3,7);

(ii) those that occur in $\phi(U)$;

(iii) ρ, or the set of values of ρ for many test samples in an assay.

All could be estimated simultaneously from the data, using an accepted principle such as least squares or maximum likelihood. In practice, reasonably precise estimates of the variance parameters call for more data than can be expected from one assay, yet potency estimates are remarkably insensitive to change in these parameters.

A better policy seems to be to use evidence from many assays for estimating parameters that define the "shape" of $\phi(U)$, V_2/V_1 in (4.1) or J in (4.2), then to assume these fixed for future assays of the same kind but still to estimate a factor of proportionality such as V_1 or V. This utilizes the experience that the exact form of $\phi(U)$ is not vital while still allowing for a general factor to distinguish assays with intrinsically high or low variability. Rodbard *et al.* [18] have discussed a similar suggestion. If (4.2) is used, almost certainly J will lie between 1.0 and 2.0; either of these extremes is also a special case of (4.1). A series of related assays examined recently (Finney, [7]) suggested a value of J between 1.5 and 2.0. It is sometimes useful to confirm the comparative unimportance of the value of J (or V_2/V_1) by repeating the potency calculations with alternative choices (see below).

As already mentioned, in a multidose assay the counts for a test sample also contribute information on the parameters C, D, β. Even

in a single dose assay a test sample gives some information on C, D, as is evident when asample of unexpectedly low or high potency happens to give the highest or lowest counts of the whole assay. Commonly, though, counts on the standard contribute almost all the information on C, D. If this were sufficient to determine C, D very precisely, the problem would reduce to estimating a series of $F(x)$ functions, as defined by (3.3), identical except for displacement of x by the $\log \rho$ appropriate to each test sample. For a single dose assay, this amounts to estimating $\log \rho$ by the horizontal distance of a point for the mean count from the estimated standard curve.

If the model expressed by (3.3)–(3.5) is appropriate, then Y as defined by (3.6) is a linear function of x. Instead of (3.4), by analogy with simple linear regression the form

$$Y=\alpha+\beta x \tag{5.1}$$

may lead to estimates of α, β that in general behave better than estimates of β, μ in respect of near-normality of distribution. (Further investigation may be worth while.) In the (x, Y) space, the standard preparation and all test samples are represented by parallel straight lines, horizontal distances between the estimated lines being taken as estimates of $\log \rho$. The analysis has much in common with the logit analysis for bioassays that use quantal responses, though there are important differences in distributional and variance assumptions. Replacement of (3.5) by (3.7) brings similar connexions with probit methods (Finney [5], [6]).

6. Alternative estimation principles

Weighted non-linear least squares is perhaps the most obvious basis for estimation. It is described here with particular reference to the logistic response curve and the variance function (4.2), but many alternatives require only minor changes. Define

$$S=\sum (u-U)^2/U^J\ , \tag{6.1}$$

where summation is over all relevant counts, J is taken as known,

$$U=C+(D-C)/(1+e^{-2Y}) \tag{6.2}$$

and

$$Y=\alpha+\beta x\ . \tag{6.3}$$

In a single dose assay, a test sample that has a very low or very high mean count may invite modified estimation of C or D, but such a sam-

ple cannot lead to any precise statement about precision and can be neglected without loss; less extreme mean counts scarcely influence estimation of C and D. Consequently (6.1) can then be restricted to the standard preparation, from which C, D, α, β can be estimated. A multidose assay, on the other hand, requires all test samples to be included in S and must provide a separate α parameter for each sample even though C, D, β are the same for all.

With a suitable optimization routine (Nelder and Mead [12]), minimization of S in respect of C, D, β, and the separate α in a multidose assay is not an excessive task unless the number of test samples is very large. Moreover, the minimal value of S can be regarded as a sum of squares (with degrees of freedom equal to the number of observations minus the number of fitted parameters) for estimating V. A matrix of asymptotic variances and covariances, obtained in the standard manner, can be used in constructing probability statements about relative potencies. The estimate of $\log \rho$ for a test sample is of course the horizontal distance

$$(\hat{\alpha}_T - \hat{\alpha}_S)/\hat{\beta}\,, \tag{6.4}$$

where α_S, α_T are the α parameters for the standard and the test respectively.

As an alternative to least squares, maximum likelihood can be used. The log likelihood can be written

$$L = -\frac{1}{2}\sum \ln(VU^J) - \frac{S}{2V}\,; \tag{6.5}$$

maximization of L is evidently closely related to minimization of S. Apart from the fact that V is also being estimated, the computations required are much the same as before. Iterative convergence is likely to be a little slower because of the additional parameter.

A third possibility is to seek a transformation of the counts that will approximately stabilize the variance. This cannot be done in any simple way for (4.1), but for (4.2) the usual first order approximation leads to the transform

$$u^* = u^{1-J/2}\,, \qquad U^* = U^{1-J/2}\,; \tag{6.6}$$

if $J=2$, $u^* = \ln u$ is appropriate. The assumption of a normal distribution for u^* about U^* conflicts with the previous assumption of normality for u, but can be tried. Minimization of

$$S^* = \sum (u^* - U^*)^2 \tag{6.7}$$

is then both the least squares and the maximum likelihood procedure.

One might at first imagine that the computations here would be quicker because variance functions do not have to be calculated; this is counterbalanced by the need for recalculation of U^* at each iteration, so that there is practically no time advantage for one over the other.

I have applied alternative calculations to counts from an oestradiol assay. This single dose RIA had quadruplicate counts at 6 doses of the standard, with 4 counts under non-specific conditions (infinite dose) and 8 counts at zero dose (Table 1). As a selection from many computations, I show (Table 2) estimates of the parameters for the standard curve using weighted least squares and three values of J, with corresponding estimates by the other two procedures for $J=1.5$. Agreement is excellent, and there are no obvious reasons for preferring one set of estimates to another. Table 3 illustrates, for four typical test samples, how this agreement is carried over to potency estimates: appreciable differences between estimation procedures appear only for samples of very low or very high potency, and precision is then necessarily so poor that no trust will be placed in any estimate.

Table 1 Counts recorded for the standard curve in a radioimmunoassay of oestradiol

Dose (units of 10 pg)	Counts				Means
0	1627	1567	1720	1660	1681
	1704	1689	1759	1722	
0.625	1182	1291	1294	1312	1270
1.25	1029	1112	986	1074	1050
2.5	702	784	733	777	749
5.0	486	485	501	460	483
10.0	307	277	285	275	286
20.0	196	164	193	182	184
non-specific	25	39	51	38	38

Table 2 Estimated parameters for the standard curve

Variance parameter	$\hat{a}$	$\hat{b}$	$\hat{C}$	$\hat{D}$
Weighted least squares				
$J=1.0$	0.330±0.025	−1.175±0.030	39.8±4.3	1681.5±19.2
$J=1.5$	0.325±0.033	−1.169±0.032	40.1±2.5	1683.8±28.2
$J=2.0$	0.317±0.055	−1.160±0.046	40.6±2.0	1688.3±54.7
Maximum likelihood				
$J=1.5$	0.325±0.031	−1.169±0.030	39.7±2.3	1681.1±26.6
Transformed count				
$J=1.5$	0.325±0.033	−1.165±0.033	37.6±2.4	1681.7±29.1

Table 3 Estimated potencies (pg/ml plasma) for four oestradiol test samples, with limits at probability 0.95, by method of weighted least squares

Variance parameter	Sample 2	Sample 8	Sample 10	Sample 9
Potency				
$J=1.0$	58	217	1700	13000
$J=1.5$	58	217	1720	13000
$J=2.0$	57	217	1750	14000
Lower limit				
$J=1.0$	49	190	1270	3000
$J=1.5$	46	187	1360	5000
$J=2.0$	38	173	1360	6000
Upper limit				
$J=1.0$	69	249	2290	52000
$J=1.5$	73	253	2200	33000
$J=2.0$	85	271	2290	32000

7. Computing

An excellent range of computer programs has been developed by Rodbard ([14]; Rodbard and Hutt [17]), who appears to be continually extending the facilities and introducing greater generality. Although many special experimental complications have been taken into account, I do not think these or any other programs (Cook [3]) yet provide for the full estimation processes that I have outlined. For research purposes, I have been able to use general least squares and maximum likelihood programs, augmented by special calculations for potency estimates.

The great number of radioligand assays now being performed, and their importance as a routine diagnostic aid, justify considerable programming effort for a comprehensive and flexible system. In addition to such obvious requirements as simplicity of input, clarity and comprehensiveness of output, and ease of use by many who are not professional statisticians, other features of a good program call for special care. Versatility in adaptation to different response curves and different variance functions is important, not least in order to permit robustness of conclusions to be checked by modifying uncertain assumptions. The program should handle estimation of C, D, and should draw attention to test samples of such extreme potency that closeness to an asymptote prohibits any precision of estimation. Single dose assays present no major computational problems. Multidose are more demanding if proper use is to be made of their superior information. With p test samples, a $(p+4)\times(p+4)$ matrix must be inverted even for potency estimation

and more than this for full exploitation of validity tests, but the pattern of the matrix is such as to permit a reasonable scheme of computation. A common practice in routine assays is to analyze each test sample separately with the standard. Often this is adequate; however, it does not make full use of the information available in a multidose assay, and is unlikely to represent any economy of modern computational facilities. Any good program will also include various provisions for detecting anomalous behaviour, and allowing action to be taken in order to ensure that potency estimation for some samples is not prevented or distorted because others obviously fail to conform to the model on which the analysis is based.

When experimental data must frequently be analyzed for very practical purposes, there is a temptation to say that in routine use a simplified and approximate statistical analysis is adequate, any more sophisticated analysis being reserved for research investigations in which a professional statistician is consulted. Often exactly the opposite is true. An experienced statistician may be able to judge from his scrutiny of data whether an approximate analysis is safe, for example whether C, D are so well-estimated that they can be assumed error-free without fear of producing a false appearance of high precision in potency estimates. Routine radioligand assays, on the other hand, will be interpreted and used medically without statistical expertise; such expertise, therefore, needs to be built into the program, so that the occasional assay in which some test samples cannot safely be estimated, or $\hat{\beta}$ is so low in precision as to jeopardize all estimation, or the assumed form of response curve fails to fit the data, declares itself unmistakably by messages output by the computer after internal computations and tests that often need not be reported. Only a highly developed program can be adequate.

I am greatly indebted to Dr. S. Z. Cekan for the data mentioned in the text and used for Tables 1–3, and also to Dr. D. Rodbard for much helpful correspondence.

UNIVERSITY OF EDINBURGH

REFERENCES

[1] Arrigucci, A., Forti, G., Fiorelli, G., Pazzagli, M. and Serio, M. (1973). Mathematical analysis of the results of competitive binding methods, *The Endocrine Function of the Human Testis*, vol. 1, 74–90.

[2] Burger, H. G., Lee, V. W. K. and Rennie, G. C. (1973). A generalised computer program for the treatment of data from competitive protein-binding assays including radioimmunoassays, *Journal of Laboratory and Clinical Medicine*, **80**, 302–312.

[3] Cook, B. (1975). Automation and data processing for radioimmunoassays, *Steroid Im-*

munoassay—Proceedings of the Fifth Tenovus Workshop, Alpha Omega Publishing Ltd., Cardiff, 293-310.

[4] Ekins, R. and Newman, B. (1970). Theoretical aspects of saturation analysis, *Acta Endocrinologica*, suppl. **147**, 11-36.

[5] Finney, D. J. (1964). *Statistical Methods in Biological Assay* (2nd edition), Charles Griffin & Co. Ltd., London.

[6] Finney, D. J. (1971). *Probit Analysis* (3rd edition), The University Press, Cambridge.

[7] Finney, D. J. (1976). Radioligand assay, *Biometrics*, **32**, 721-740.

[8] Harding, B. R., Thomson, R. and Curtis, A. R. (1973). A new mathematical model for fitting an HPL radioimmunoassay curve, *Journal of Clinical Pathology*, **26**, 973-976.

[9] Healy, M. J. R. (1972). Statistical analysis of radioimmunoassay data, *Biochemical Journal*, **130**, 207-210.

[10] Midgley, A. R. (1966). Radioimmunoassay: A method for human chorionic gonadotrophin and human luteinizing hormone, *Endocrinology*, **79**, 10-18.

[11] Midgley, A. R., Niswender, G. D. and Rebar, R. W. (1969). Principles for the assessment of the reliability of radioimmunoassay methods (precision, accuracy, sensitivity, specificity), *Acta Endocrinologica*, supplementum no. **142**, 163-184.

[12] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization, *Computer Journal*, **7**, 308-313.

[13] Rodbard, D. (1971). Statistical aspects of radioimmunoassays, *Principles of Competitive Protein Binding Assays*, Lippincott,Philadelphia, 204-259.

[14] Rodbard, D. (1974). Statistical quality control and routine data processing for radioimmunoassays and immunoradiometric assays, *Clinical Chemistry*, **20**, 1255-1270.

[15] Rodbard, D. and Cooper, J. A. (1970). A model for prediction of confidence limits in radioimmunoassays and competitive protein binding assays, *Proceedings, Symposium on Radioisotopes in Medicine*, 659-674.

[16] Rodbard, D. and Frazier, G. R. (1975). Statistical analysis of radioligand assay data, *Methods in Enzymology*, **37**, 3-22.

[17] Rodbard, D. and Hutt, D. M. (1974). Statistical analysis of radioimmunoassays and immunoradiometric (labelled antibody) assays: a generalized weighted, iterative, least-squares method for logistic curve fitting, *Symposium on RIA and Related Procedures in Medicine*, **1**, 165-192.

[18] Rodbard, D., Lenox, R. H., Wray, H. L. and Ramseth, D. (1976). Statistical characterization of the random errors in the radioimmunoassay dose-response variable, *Clinical Chemistry*, **22**, 350-358.