Biochem. J. (1972) **130**, 207–210
Printed in Great Britain

Statistical Analysis of Radioimmunoassay Data

By MICHAEL J. R. HEALY
Medical Research Council Clinical Research Centre, Northwick Park Hospital, Harrow, HA1 3UJ, Middlesex, U.K.

(*Received* 5 *June* 1972)

The statistical processing of radioimmunoassay data is discussed, with special emphasis on fitting the standard curve, screening the data for aberrant readings and combining separate estimations from a single sample.

The immunochemical kinetics of radioimmunoassay are susceptible to theoretical analysis and this theory has been developed by a number of authors [see, for example, Meinert & McHugh (1968), Rodbard *et al.* (1969), Ekins & Newman (1970), Ekins (1970), Yalow & Berson (1970)]. However, not all aspects of the method when used as a routine can be covered by theoretical treatment and it is of interest to introduce some statistical concepts from the more familiar bioassay field (see, for example, Finney, 1964), which has developed largely without this kind of theoretical guidance. Some of these concepts are not adequately implemented in many of the packaged computer programs now generally available.

Form of the Data

In a clinical context particularly, a single assay usually involves a substantial number of unknowns. These are accompanied by the set of standards aimed to cover the concentration range of interest, by a zero-concentration sample giving the maximum bound concentration and by a reagent blank, which, in view of the way the method works, corresponds to an infinite concentration. Each of the standard and unknown samples is normally run in duplicate and this will be assumed in what follows. For convenience, it is also assumed that only the bound fraction is counted.

Shape of the Standard Curve

There has been some discussion about the mathematical formula best suited to fit the standard curve. The curve forms a descending sigmoid with the zero concentration and blank values as the upper and lower asymptotes. On a log-concentration scale, the sigmoid is reasonably symmetric and a log-logistic curve is commonly found to give a good fit (Rodbard & Lewald, 1970).

The log-logistic curve has the following rather complicated formula, relating y, the expected radioactivity count rate, to x, the concentration:

$$\hat{y} = a + b[\exp(c - d\cdot\ln x)/\{1 + \exp(c - d\cdot\ln x)\}] \quad (1)$$

The structure of this formula is rather clearer if it is written as:

$$\hat{y} = a + bq \quad (2)$$

where

$$q = z/(1+z) \quad (3)$$

and

$$z = \exp(c - d\cdot\ln x) \quad (4)$$

From a statistical viewpoint, the important aspect of this curve is that it has four unknown parameters, a, b, c and d, that require estimation from the data. The blank and zero-concentration samples provide information directly about the parameters a and b, but this is subject to error just like the information obtained from the other samples; moreover, extra information about these parameters is available from the other standard samples, particularly those at high or low concentrations.

Currently, the method of analysis that is usually recommended for fitting this type of curve involves taking the zero-concentration and blank radioactivity count rates as providing error-free estimates of a and b. The other standard radioactivity count rates are then corrected for the blank and expressed as a proportion of the zero-concentration rate. This in turn is transformed into a quantity p, given by formula (5):

$$p = \ln[(y-a)/\{b-(y-a)\}] \quad (5)$$

A little algebra then shows that formula (1) can be transformed to give, for the expected value of p:

$$\hat{p} = c - d\cdot\ln x \quad (6)$$

so that c and d can be estimated by a linear regression calculation. This approach has the disadvantage of

ignoring the errors in the asymptote counts and also some of the available information on the asymptotic values; it is possible that some claims that formula (1) fails to fit certain types of data may have been due to methods of fitting which were inadequate in this way. Its apparent simplicity is also somewhat deceptive. The variance of p in formula (5) varies considerably over the range 0–1 so that weighted regression is required, and because the weights depend upon the fitted values the procedure has to be iterative.

Fitting the four parameters of formula (1) is a nonlinear regression problem, but this presents no great practical difficulty if good computing facilities are available. The formula is linear in the parameters a and b and it is most efficient to take advantage of this fact. The procedure is then to start with trial values of c and d, so that q in formula (3) can be evaluated for each value of x, and to estimate a and b from formula (2) by simple regression, calculating at the same time the residual sum of squares. The values of c and d can then be successively adjusted so as to minimize this sum of squares by a numerical technique such as that of Nelder & Mead (1965). Convergence is usually quite rapid in practice, even from crude initial estimates.

In fitting the regression equation (2) it would be feasible to allow for unequal variances in the y values by some form of weighting. The actual variances contain many sources of variation over and above the Poisson-distributed errors associated with the radioactivity counts. Rodbard & Cooper (1970) (see also Rodbard & Lewald, 1970) give a theoretical analysis of the errors in y and show that the higher radioactivity count rates tend to have greater variability. In practice, however, the extent of this increase in variance is not so great that weighting would be expected to give worthwhile improvement in the precision of the parameter estimates, especially when the rate of increase is not known in advance but has itself to be estimated from limited data. Note that the position is quite different when the transformed curve is fitted by using formula (6). The values of p derived from y values close to the asymptotes have very high variability and it is essential to give such values lower weights in the fitting process.

Preliminary Screening

The procedures involved in radioimmunoassay are sufficiently complex that a certain proportion of aberrant observations has to be expected in practice and it is one objective of running all samples in duplicate to detect and eliminate these. It is common practice to calculate an estimated standard deviation from the differences between duplicates and to screen out any sample for which the difference exceeds some multiple of the standard deviation. The difficulty, of course, is that the excessive differences are liable to enter into the estimated standard deviation and to inflate it disproportionately, particularly if the usual root-mean-square estimator is used. One simple alternative possibility is to take the median of the absolute values of the differences between duplicates; computationally this is not too time-consuming if the partial-sorting method of Hoare (1961) is used. The result can be taken as a direct estimate $\hat{\sigma}$ of the standard deviation of a single reading: the exact conversion factor, assuming a normal distribution, is 0.984.

Further screening is advisable during the fitting of the standard curve. The residual (observed minus fitted value) can be calculated at each of the observed points (including blank and background). If any residual exceeds (say) three times its standard error (which is $\hat{\sigma}/\sqrt{2}$ for the mean of two duplicates), the corresponding point can be omitted and the fitting process repeated. This procedure should not be used lightly, however; the occurrence (in duplicate) of aberrant standard values must cast considerable doubt on the validity of the assay as a whole.

The goodness of fit of the standard curve can be roughly assessed by comparing the mean square of the residuals (with $n-4$ degrees of freedom, where n is the number of standard pairs) with the error variance $\hat{\sigma}^2$. This corresponds to the usual bioassay test of linearity. The F distribution, which would provide the usual means of assessing this comparison, is not theoretically appropriate here for several reasons, but F tables may be used with caution to give some indication of statistical significance. $\hat{\sigma}^2$ may be regarded as having $0.6N$ degrees of freedom, where N is the total number of pair differences used in its estimation.

Combination of Estimates

There is no difficulty in reading off estimated concentrations from observed responses, by using the formula $x = \exp\{(c-p)/d\}$, with p given by formula (5). Responses outside the range of blank to background obviously have to be excluded, but otherwise there seems no reason to omit extreme responses, unless the choice of standards has been unfortunate so that a substantial part of the response scale is not covered. However, it is clear that concentrations estimated from the steeply sloping central portion of the curve will be considerably more precise than those corresponding to the near-horizontal stretches near the asymptotes. If errors due to estimation of the standard curve parameters are ignored, the variance of an estimated log concentration (using natural logarithms to the base e) can be shown to be given approximately by:

$$V(\ln x) \doteqdot \left(\frac{b}{d(y-a)\cdot[b-(y-a)]}\right)^2 \cdot \tfrac{1}{2}\sigma^2 \qquad (7)$$

where $\frac{1}{2}\sigma^2$ is the variance of the mean of a pair of duplicates. If a set of samples are in fact different dilutions from a single original specimen, the corresponding results can be combined by taking a weighted mean of the quantities

$$g = \ln x + \ln(\text{dilution factor}) \qquad (8)$$

with the reciprocals of the variances used as weights. The consistency of the several samples can be assessed by calculating the weighted sum of squares (formula 9) and referring this to the ordinary tables of the χ^2 distribution.

$$\Sigma wg^2 - (\Sigma wg)^2/\Sigma w \qquad (9)$$

This test corresponds to the tests of parallelism and curvature in straight-line bioassay. The weights corresponding to formula (7) tend to be slightly too large (and quoted standard errors based on them slightly too small) because the errors in estimating the standard curve have been ignored. In that no extrapolation is involved, this effect is not likely to be important; this is certainly true for the weights, for which only relative values matter. However, formula (7) deals only with within-assay variation. It is common experience that between-assay variation is considerably in excess, so that the total weights of the mean estimates from single assays may require adjustment when combining estimates made on separate occasions. A possible rule-of-thumb procedure that makes approximate allowance for between-assay differences is: (1) find the median of all the weights in the assays to be combined; (2) reject all estimates whose weights are less than $1/k$ times the median weight as being too imprecise to be usable; (3) reduce all weights greater than k times the median weight so that they equal that value (k may be chosen to be between, say, 3 and 10, the smaller values being appropriate for high between-assay variability). Estimates formed in this way will not be unduly dominated by individual values that are very precise on a within-assay basis.

It will in any case usually be necessary to estimate the precision of these combined values. A weighted sum of squares can be calculated for each combined estimate from formula (9) and the total of these sums of squares divided by the corresponding degrees of freedom. This produces an estimated variance per unit weight, s^2, analogous to the heterogeneity factor of probit analysis (Finney, 1952). The variance of a weighted mean is now estimated as $s^2/\Sigma w$ in the usual way.

The presence of between-assay variability is too often accepted without comment. The use of standard preparations (Cotes, 1970) is aimed at eliminating this source of variability and if this aim is not achieved an opportunity for improving assay technique is being overlooked. Often the potential improvement here would greatly outweigh that available from reductions in within-assay variability, the topic usually treated by theoretical approaches. In particular, the energetic use of quality-control techniques should be an indispensible part of routine assay practice (Rodbard *et al.*, 1969; Whitby *et al.*, 1967).

Experimental Design

Questions of experimental design have been extensively treated by R. Ekins and his colleagues (Ekins & Newman, 1970). These authors show that the parameters of the standard curve are to some extent under the experimenter's control and indicate methods of producing a curve, which (put fairly crudely) is steep in the region of interest.

A matter which does not seem to have received attention is the choice of the number of standard concentrations. Normally 10–15 concentrations are used with twofold or smaller dilution ratios between them. The theory of experimental design for non-linear curve-fitting situations is incomplete, but work by Box (1968) makes it likely that the information obtained from the standards would be increased if the standard concentrations were reduced in number with correspondingly greater replication on each. At least four concentrations (including zero and blank) are, of course, essential, and more are needed if departures from the assumed form of the curve are to be detectable.

The stability of the standard curve should not go unquestioned: Welborn *et al.* (1970) have shown that changes in both position and shape of the curve can occur during a lengthy run. This suggests that drift standards (of a kind familiar in automated biochemical assays) should be inserted at intervals in the run and that these should be at more than one concentration. The need for quality-control measures mentioned above has also been stressed in many published articles, though perhaps without the influence on practice that could be desired.

Conclusion

Although radioimmunoassay is a complex technique, its results do not differ statistically in any fundamental way from those in other more familiar fields. It is suggested that the use of a number of more or less standard statistical techniques will lead to the better utilization of assay values.

A computer program, in the form of a Fortran IV subroutine, embodying some of the ideas outlined above, is available from the author on request.

I am grateful to Dr. N. Veall and Mrs. S. Chalkley for introducing me to radioimmunoassay and for discussing its problems.

References

Box, M. J. (1968) *J. Roy. Statist. Soc. Ser. B* **30**, 290–302

Cotes, P. M. (1970) *In Vitro Proced. Radioisotop. Med., Proc. Symp.* 675–699

Ekins, R. (1970) *In Vitro Proced. Radioisotop. Med., Proc. Symp.* 325–352

Ekins, R. & Newman, B. (1970) *Acta Endocrinol. (Copenhagen) Suppl.* **147**, 11–36

Finney, D. J. (1952) *Probit Analysis*, Cambridge University Press

Finney, D. J. (1964) *Statistical Method in Biological Assay*, Griffin, London

Hoare, C. A. R. (1961) *Commun. Ass. Comput. Machinery* **4**, 322

Meinert, C. L. & McHugh, R. B. (1968) *Math. Biosci.* **2**, 319–338

Nelder, J. A. & Mead, R. (1965) *Comput. J.* **7**, 308–313

Rodbard, D. & Cooper, J. A. (1970) *In Vitro Proced. Radioisotop. Med., Proc. Symp.* 659–673

Rodbard, D. & Lewald, J. E. (1970) *Acta Endocrinol. (Copenhagen) Suppl.* **147**, 79–103

Rodbard, D., Bridson, W. & Rayford, P. L. (1969) *J. Lab. Clin. Med.* **74**, 770–781

Welborn, T. A., Stenhouse, N. S., Curnow, D. H. & Johnson, C. J. (1970) *In Vitro Proced. Radioisotop. Med., Proc. Symp.* 525–538

Whitby, L. G., Mitchell, F. S. & Moss, D. W. (1967) *Advan. Clin. Chem.* **10**, 65–156

Yalow, R. S. & Berson, S. A. (1970) *In Vitro Proced. Radioisotop. Med., Proc. Symp.* 455–482

APPENDIX

A Specimen Analysis: Assay of Thyrotrophin

1. *Estimated standard deviation* (from median absolute difference) = 316

   Rejection limit (3 × S.D.) for differences = 1340
   for residuals = 670

2. *Standard curve*

Concn. (μunits/ml)	Obs. (c.p.m.) 1	Obs. (c.p.m.) 2	Mean	Abs. diff.	Fitted values	Residuals
0	37	5076	2557	5039*	5110	(−2553)
0.1	4789	4928	4859	139	5050	−191
0.5	4979	4961	4970	18	4896	+74
1.0	4769	5131	4950	362	4748	+202
2.0	4270	4571	4420	301	4514	−94
5.0	4462	3939	4201	523	4026	+175
10.0	3374	3500	3437	126	3512	−75
16.0	2793	2966	2879	173	3114	−234
20.0	2482	3423	2952	941	2918	+35
50.0	2111	2278	2195	167	2149	+46
100.0	1676	1904	1790	228	1677	+113
Blank	686	700	693	14	744	−51

Parameters of fitted curve: $a = 744$; $b = 4366$; $c = 2.403$; $d = 0.805$. Residual S.D. = 240.

3. *Unknowns*

Obs. (c.p.m.) 1	Obs. (c.p.m.) 2	Mean	Abs. diff.	Concn. (μunits/ml)	Wt. of ln concn.
4779	3913	4346	866	2.89	5.2
5599	4944	5272†	655	—	0
3799	3802	3800	3	6.91	10.9
70	8375	4222	8305*	—	0
6305	5013	5659†	1292	—	0
4657	4602	4630	55	1.52	2.5
..	..	..	..	..	..

* Point rejected—difference too great.

† No estimate—mean count rate above upper asymptote.