Drug Development and Industrial Pharmacy, 26(6), 661–669 (2000)

RESEARCH PAPER

Calibration and LOD/LOQ Estimation of a Chemiluminescent Hybridization Assay for Residual DNA in Recombinant Protein Drugs Expressed in *E. coli* Using a Four-Parameter Logistic Model

Kwan R. Lee,[1,*] Byron Dipaolo,[2] and Xiaoying Ji[2]

[1] *Statistical Sciences Department and* [2] *Bioanalytical Sciences Department, SmithKline Beecham Pharmaceuticals R&D, King of Prussia, PA 19406*

ABSTRACT

*Calibration is the process of fitting a model based on reference data points (*x, y*), then using the model to estimate an unknown* x *based on a new measured response,* y*. In DNA assay,* x *is the concentration, and* y *is the measured signal volume. A four-parameter logistic model was used frequently for calibration of immunoassay when the response is optical density for enzyme-linked immunosorbent assay (ELISA) or adjusted radioactivity count for radioimmunoassay (RIA). Here, it is shown that the same model or a linearized version of the curve are equally useful for the calibration of a chemiluminescent hybridization assay for residual DNA in recombinant protein drugs and calculation of performance measures of the assay.*
Key Words: *Calibration; Chemiluminescent hybridization assay; Limit of detection; Limit of quantitation.*

INTRODUCTION

Immunoassay methods such as radioimmunoassay (RIA) and enzyme-linked immunosorbent assay (ELISA) are used frequently to calibrate the concentration of a substance of interest in an analyte or the potency of a drug by comparing the analyte with a standard reference solution. Typically, samples of several unknown concentrations in replicates and replicates of the reference standard are applied to wells of a plate, and response readings for each well on the plate are obtained. From this raw data, the estimates of concentration are computed, and

* To whom correspondence should be addressed.

precision measures of the concentrations may be obtained. Confidence intervals are increasingly favored over the estimated plus/minus standard error alternative.

The DNA hybridization method is used for determination of residual *Escherichia coli* DNA in protein drug samples. Samples for analysis are extracted to remove protein, and DNA is subsequently denatured and immobilized on a nylon membrane using a slot blot manifold. Purified, sheared *E. coli* DNA is labeled with digoxigenin (DIG) using a random primed DNA labeling procedure. DIG-labeled DNA probes selectively bind to *E. coli* DNA on the membrane. After overnight incubation, the membrane is washed with high stringency to remove unbound probe. The hybridized DNA probe is detected using a horseradish peroxidase–labeled anti-DIG-Ab and an enhanced chemiluminescent substrate. X-ray film records the results, and the autoradiographic signal is quantified using image analysis. The amount of residual DNA is quantified by interpolating from an *E. coli* DNA standard curve that is derived from plotting the *E. coli* DNA standard concentration versus the measured signal volume.

DNA ASSAY CALIBRATION

Calibration Experiment

The data in Table 1 were obtained from a typical calibration experiment and show the measured signal volumes from X-ray film quantified by a three-laser densitometer. The signal volumes are shown in three replicates (v1–v3) together with some simple statistics. Calibration with a straight line was tried, but the fit was not satisfactory (Fig. 1). The figure suggests the presence of some nonlinearities not accounted for by the linear fit. A higher degree polynomial improves the fit (Fig. 2) and can be used for calibration purposes, but polynomials are unstable at the extreme boundaries of the fit. The difficulty in interpretation of the coefficient of the polynomial was regarded as a drawback. As an alternative, we decided to try a four-parameter logistic model for the calibration of the DNA assay.

The Four Parameter Logistic Model

The four-parameter logistic curve is a very flexible calibration model for data. It is sometimes referred to as a standard curve in the analytical community and is commonly given by

$$y = d + \frac{a - d}{1 + (x/c)^b}$$

Here, x is the concentration, and y is the instrument response, such as a gamma count. If x approaches 0, then y will be closer and closer to a; hence, a is the lower asymptote. If x approaches infinity, then y approaches d; hence, d is the upper asymptote. If $x = c$, $y = (a + d)/2$; hence, c is the predicted concentration at midlevel response. The parameter b is sometimes called the slope since it is the slope of the linearized curve (see logit-log model below). A typical shape of the curve when b is positive is given in Fig. 3.

By solving the equation in terms of x, the predicted concentration (‘‘potency’’) can be found by the following formula:

$$x = c\left(\frac{y - a}{d - y}\right)^{1/b}$$

which is sometimes divided by c and expressed as percentages.

Table 1

Typical Data From DNA Calibration Assay

Row	Concentration (pg)	v1	v2	v3	Average	STD	RSD (%)
1	0	0.36	−5.76	−1.29	−2.23	3.17	—
2	2.5	39.01	31.67	37.91	36.2	3.96	10.94
3	5	85.39	88.71	76.68	83.59	6.21	7.43
4	10	331.19	346.12	337.4	338.24	7.5	2.22
5	25	1147.16	1549.28	1701.69	1466.04	286.48	19.54
6	50	3168.63	4239.02	4082.25	3829.97	578.07	15.09
7	75	5434.11	5891.64	5665.32	5663.69	228.77	4.04
8	100	6543.38	7144.19	6724.54	6804.04	308.19	4.53

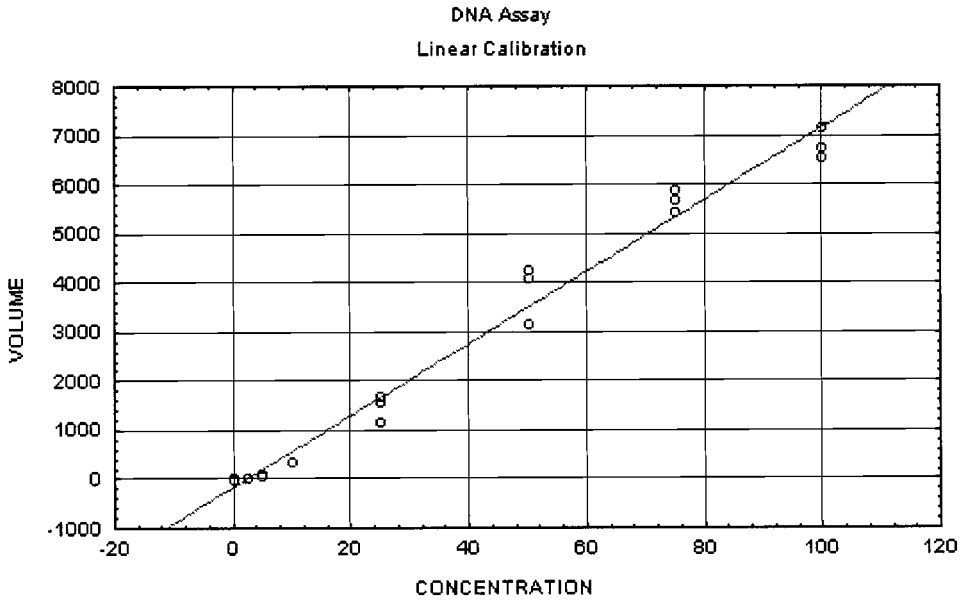

Figure 1. Calibration with a straight line.

Calibration with Logit-Log Model

An interesting simplification to the curve can be made by taking so-called logit transformation of the model; this is sometimes called the logit-log model. By taking a logit transformation of y, the above equation can be linearized as

$$\log\left(\frac{y-a}{d-y}\right) = b\,\log(x) - b\,\log(c)$$

Hence, the curve is now linear in $\log(x)$, and b is the slope in this transformed model. The logit-log model can be useful when a and d are known or they can be estimated precisely, which is usually the case. The a and d

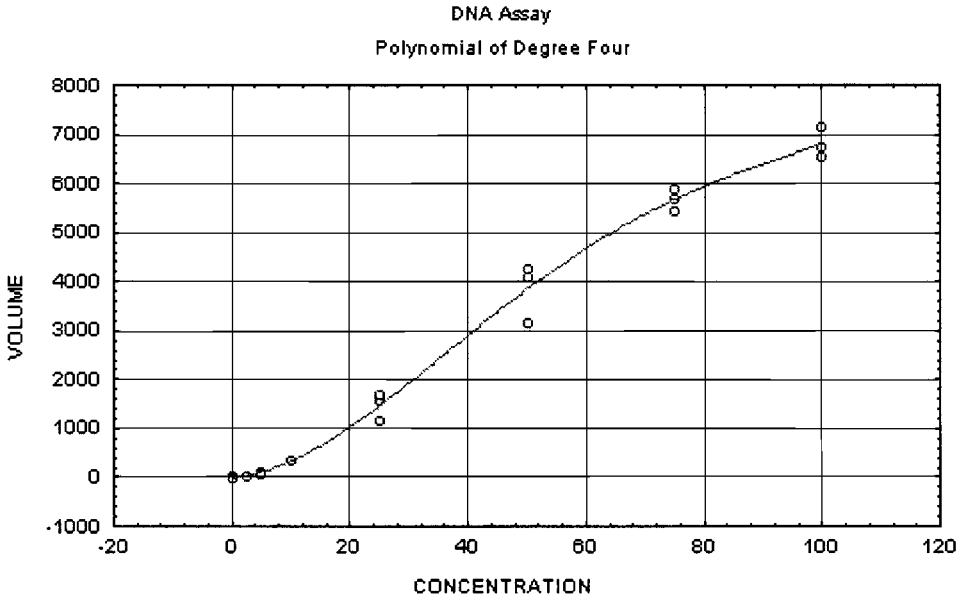

Figure 2. Calibration with a fourth-degree polynomial.

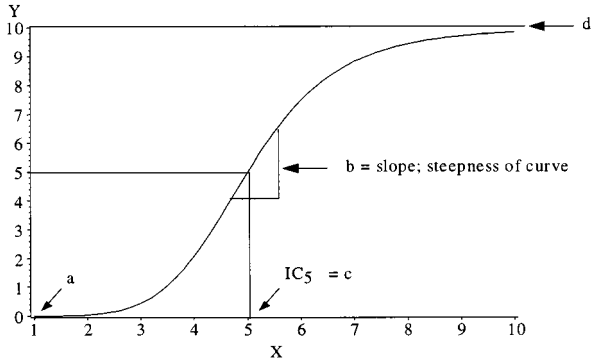

Figure 3. The four-parameter logistic model.

can be estimated reasonably well by taking the minimum and maximum of the data *y*'s. Then, *a* and *d* are regarded as constants in the remainder of the curve-fitting procedure. The two parameters *b* and *c* then can easily be estimated from the slope and intercept of the fitted straight line. The logit-log model was fit for the average signal volumes of the calibration assay (Table 1) and is shown in Fig. 4. The estimated parameters *b* and *c* are given in Table 2 together with an R^2 of .998.

Now, the curve can be used to predict the concentration using the potency formula for *x* above. Table 3 shows the spiking experiment for chosen concentrations (zero concentrations would be beyond the limit of detection)

Table 2

Logistic Parameter Estimation

a (min)	0.000
b (slope)	2.100
c (IC_{50})	49.396
d (max)	8000.000
Intercept	−8.191
R^2	0.998

and signal volumes measured (v1–v3). The percentage recovery is an accuracy measure, and it is the percentage of predicted concentration per spiked concentration. To gain some information on the precision of the assay, the concentration of the assay (c1–c3) was estimated for each of the signal volumes (v1–v3). As summary measures of precision, computations for standard deviation (STD) and relative standard deviation (RSD, sometimes called the coefficient of variation, CV) are shown in Table 4. All these tables can be computed on a simple spreadsheet, and they do not require any special statistical tools or nonlinear estimation procedures. Actually, you can go one step further and compute more measures of assay performance, such as limit of detection (LOD) and limit of quantitation (LOQ) for a straight-line regression model (see, e.g., Ref. 1). We used an Excel spreadsheet for implementation; a copy is available on request.

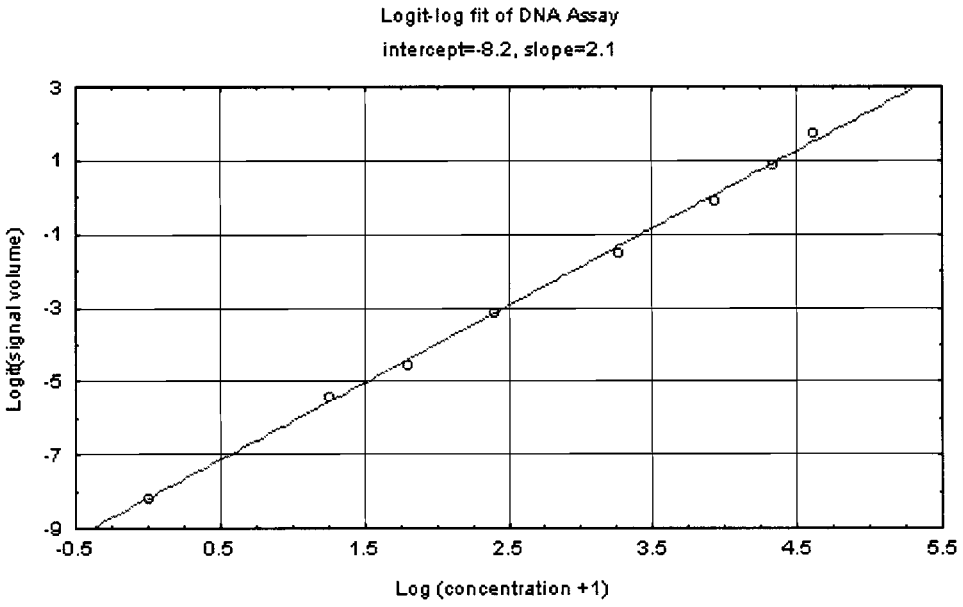

Figure 4. Linearized curve and straight-line fit.

Table 3

Spiked Samples and Measured Signal Volumes

Sample	Spike (pg)	v1	v2	v3
1	0	−5.67	4.60	6.51
2	0	−23.42	6.89	−4.93
3	10	339.65	335.76	346.75
4	50	4115.11	4150.16	3321.26
5	0	−19.97	40.75	−14.62
6	10	367.52	445.86	376.76
7	50	3885.05	3573.25	3513.14

If we are willing to use a full four-parameter logistic model with nonlinear estimation, the accuracy of the predictions can be improved, however. Moreover, some major improvements can be made on the precision estimate of the performance measures of the assay. For example, the RSD given in Table 4 does not take into account the uncertainty in the model fit and probably underestimates the true variation.

CALIBRATION WITH THE FOUR-PARAMETER LOGISTIC MODEL

A four-parameter logistic model has been used extensively for calibration of an immunoassay in which the response is optical density for ELISA or adjusted radioactivity count for RIA. Hence, there are large amounts of statistical literature on the estimation and prediction based on the curve. One of the very useful statistical developments of the calibration and assay development using the four-parameter logistic model was recently given by O'Connel et al. (2). The paper is tutorial in nature and shows the calculation of various performance measures of the assay, including precision profiles of fit and model goodness of fit test. Furthermore, they implemented the calculations using S-Plus functions (3); these are available free of charge on the Statlib. Statlib is a library of free statistical software routines and may be accessed by Internet web page (http://lib.stat.cmu.edu).

We have downloaded the S-Plus library to the PC Windows library directory and stored it in the usual library directory of S-Plus 4.5. To run the given S-PLUS functions in the named calibrate directory and store the functions in the attached data directory, some modification to the S-Plus source statement were needed, and attaching the directory as a working directory was necessary before sourcing the code as follows:

```
> attach(''C:/program Files/splus45/library/
  calibrate/_data'', pos=1)
> source(''C:/program Files/splus45/library/
  calibrate/source file'')
```

Once the library functions are in the data directory, they can be called by library statement in a data analysis session. One way of running the *fpl* library function for DNA assay is to prepare a data file (say, dna.dat) using a text editor and to copy the second to fifth columns of Table 1 to the file. Now, following S-Plus codes may be used to read the data file and define proper concentration and response columns (dna.conc and dna.resp, respectively). It is assumed that the data file is in a specific directory in the c drive (dnasc). It should also be noted that S-Plus has a built-in function under the same name (*fpl*). This built-in function computes the response for given parameter values and concentration of the logistic function, and it is not the function we need. Hence, it is necessary to attach the library on top of the search list (first=T) to

Table 4

Predicted Concentrations and Their Accuracy and Precision

Sample	Spike	c1	c2	c3	Predicted Concentration	Percentage Recovery	STD	RSD (%)
1	0	0.563	0.416	0.670	0.55	—	0.13	—
2	0	2.067	0.716	0.462	1.08	—	0.86	—
3	10	10.204	10.140	10.320	10.22	102.2	0.09	0.89
4	50	49.769	50.195	40.960	46.97	93.9	5.21	11.10
5	0	1.844	3.008	1.452	2.10	—	0.81	—
6	10	10.653	11.839	10.798	11.10	111.0	0.65	5.83
7	50	47.062	43.607	42.965	44.54	89.1	2.20	4.95

make the library function have the priority over the built-in function.

```
> library(calibrate, first=T)
> detach(1)
> attach(''c:/dnasc/_data'', pos=1)
> dna ← read.table(''c:/dnasc/dna.dat'', header=T)
> dna.conc ← rep(dna[,''conc''], 3)
> dna.resp ← c(dna[,''v1''], dna[,''v2''],
  dna[,''v3''])
> fpl.dna ← fpl(dna.conc, dna.resp, dos=T,
  b3start=50, b4start=2))
> plot(fpl.dna)
```

Now, the S-Plus object fpl.dna contains all the information on the four-parameter logistic fit and performance measures of the assay. We are not going to repeat the theoretical results given in the paper by O'Connel et al., but the theory is based on the general nonlinear model theory. The parameter estimates and their standard errors along with other performance measures are shown in Fig. 5. Here, b1 = *a*, b2 = *d*, b3 = *c*, b4 = *b* in our parameterization of the same model. We had some trouble in convergence of the nonlinear estimation procedure, but specifying the starting values of b3 and b4 from a logit-log model of Table 2 solved the problem. In that sense, the logit-log model has another use: to provide good initial values to the nonlinear estimation of the four-parameter logistic curve.

Other performance measures in Fig. 5 need explanation. The minimum detectable concentration (MDC) is defined as ''the lowest concentration which results in an expected response significantly more than the expected response at zero concentration'' (2, p. 103). The MDC is usually called the LOD in the analytical community. The reliable detection limit (RDL) is also shown and is defined as ''the lowest concentration that has a high probability of producing a response that is significantly greater than the response at zero'' (2, p. 103). The RDL is sometimes called the LOQ. As indicated by Rodbard (4), more than one definition of LOD and LOQ are available. The LOQ in the fpl output has a different meaning than usual. It is tied to the user-defined coefficient of variation (CV) and is defined as ''the lowest concentration at which the CV of the calibrated concentration is less than the required CV'' (2, p. 103). It is the lowest concentration at which the usable range of the assay starts at which the usable range is defined by a given CV.

The goodness of fit of the model can be assessed by the so-called lack-of-fit test. The goodness of model fit is assessed by comparing the model lack of fit to the pure error with a variance-ratio F test. A small p value (say, less than .05) indicates lack of fit, hence indicating improper model fit. For this DNA data, model fit looked adequate ($p = .99$).

The minimum required CV can be specified in the fpl function, and the precision profiles of the assay can be assessed using the S-Plus function *precprof*, which makes

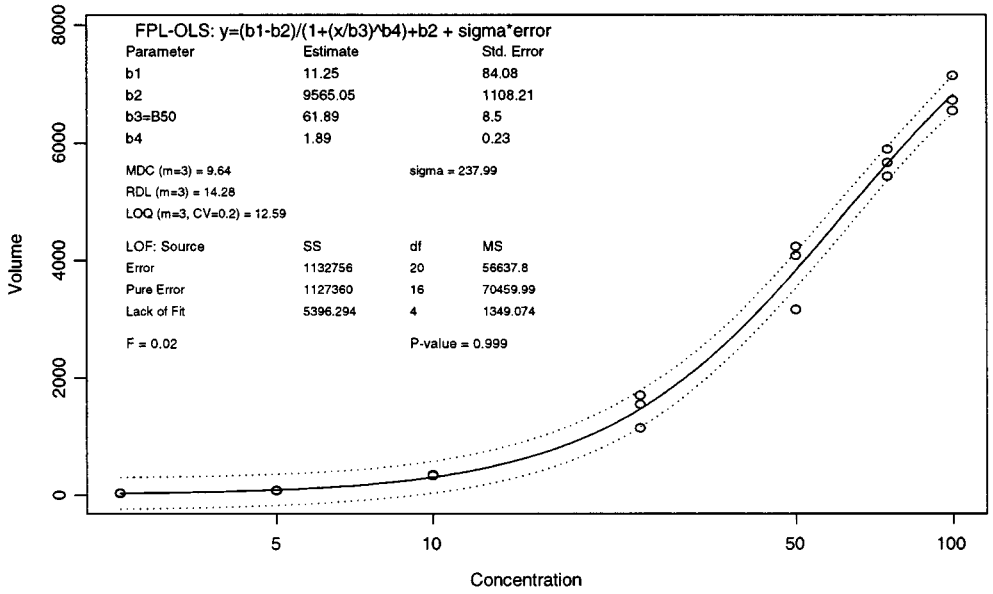

Figure 5. Four-parameter logistic model estimation and performance measures.

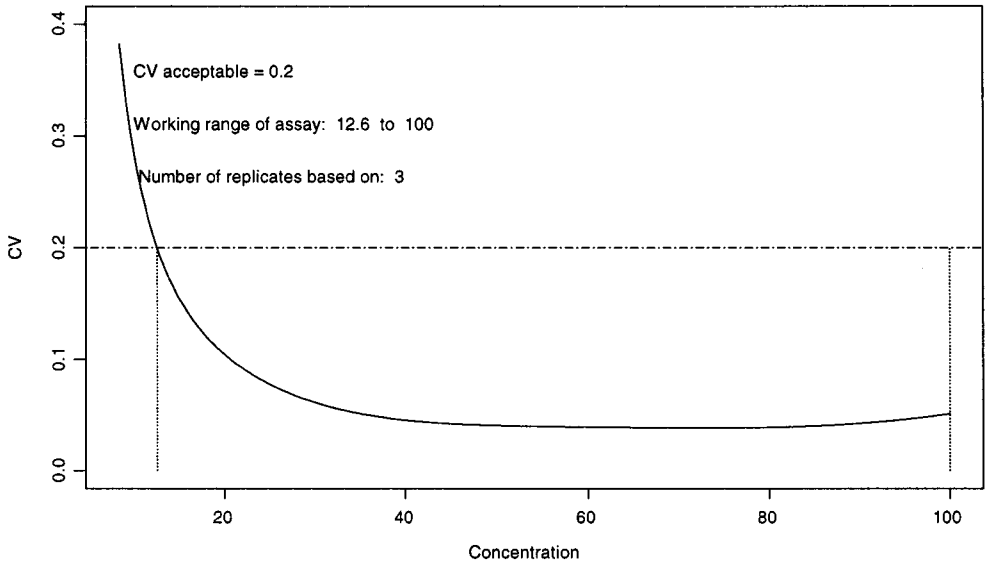

Figure 6. Precision profiles of the DNA assay.

use of user-defined CV (default = .2). The curve is a very useful addition to a calibrated curve since it informs the users of the usable range of the assay (Fig. 6). The S-plus command is simply

> precprof(fpl.dna, xlab=''Concentration'', main=''Precision profile: DNA Assay'')

The purpose of calibration is to predict the unknown concentrations for future samples. To illustrate the use of the fitted calibration curve, the data from spiked samples are used again. Table 5 shows the predicted concentrations together with the upper and lower 90% prediction limits. The prediction intervals were computed by inverting the approximate prediction interval of the response. More conservative Wald-type limits are also available from the output. The calibration output was generated using the S-Plus function *calib* as

> calib.dna ← calib(fpl.dna, spike.dna)
> plot(calib.dna)

The column spike.dna contains the average volumes of the spiked data, and fpl.dna is the object created earlier by the *fpl* function. Figure 7 presents the predicted concentration in a graph that shows the precision information as intervals. Also, the MDC of the assay is specified, and if the response of a sample is out of usable range, then it is marked as *x*.

Table 5

Predicted Concentration and Their Accuracy and Precision

Sample	Spike	Calibration Concentration	Percentage Recovery	Lower Calibration Limit	Upper Calibration Limit	Response
1	0	0.00	—	0.00	9.25	1.81[a]
2	0	0.00	—	0.00	9.06	−7.15[a]
3	10	10.58	105.8	4.36	15.05	340.72
4	50	50.26	100.5	46.83	53.89	3862.18
5	0	0.00	—	0.00	9.25	2.05[a]
6	10	11.54	115.4	6.19	15.85	396.71
7	50	47.92	95.8	44.59	51.42	3657.15

[a] Response was out of range.

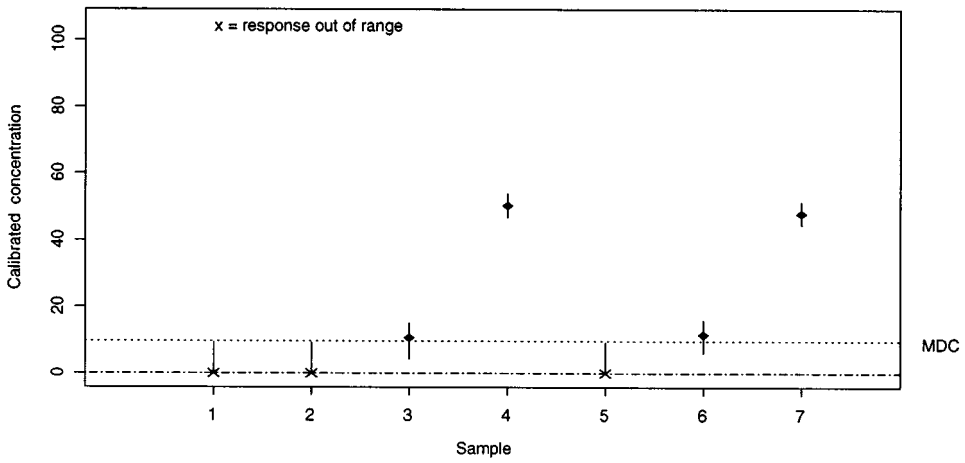

Figure 7. Calibrated sample concentrations with prediction intervals.

The statistical calculations in the *fpl* function were computed under the constant variance assumption across the assay range. When an assay exhibits the nonconstant, variance, the variance of the response generally increases with the mean, and the S-Plus function *fplvfe.pom* is available. The variance was estimated using the power of the mean (POM) approach, that is, the variance was modeled as powers of the means. The studentized residual plot in Fig. 8 was generated using the *diagplot* function with the ''*stud*'' option and shows the patterns of the residuals.

> diagplot(fpl.dna, ''stud'')

The increasing variance does not look like a problem for the current data, however. When a goal of a calibration study is estimation of several unknown concentrations, the POM approach does not account for simultaneous inference, and an alternative approach was suggested by Robinson-Cox (5). The S-Plus function for the improvement, *multi.calib*, is also available from Statlib. This approach generally provides wider confidence intervals than the POM alternative.

CONCLUSION

Calibration of the chemiluminescent hybridization assay for residual DNA in recombinant protein drugs was illustrated using logit-log and four-parameter logistic models. The logit-log model provides a simple implementation, but the four-parameter logistic model, with proper statistical performance measures, is a better alter-

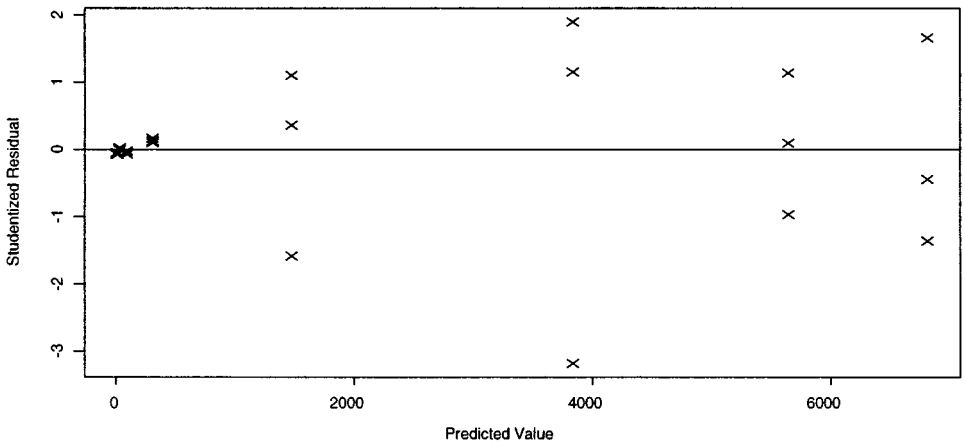

Figure 8. Studentized residual plot of the calibrated DNA assay.

native. Readily available statistical software S-Plus and a free library function *fpl* were used to illustrate the process.

REFERENCES

1. K. R. Lee and P. R McAllister, Helping analytical scientists apply statistics, Drug Dev. Ind. Pharm., 22(9&10), 891–908 (1996).
2. M. A. O'Connel, B. A. Belanger, and P. D. Haaland, Calibration and assay development using the four-parameter logistic model, Chemometrics Intelligent Lab. Sys., 20, 97–114 (1993).
3. Mathsoft, Incorporated, *S-Plus 4.5 for Windows*, Author, Seattle, WA, 1997.
4. D. Rodbard, Statistical estimation of minimum detectable concentration (''sensitivity'') for radioligand assay, Anal. Chem., 90, 1–12 (1978).
5. J. F. Robinson-Cox, Multiple estimation of concentrations in immunoassay using logistic models, J. Immunol. Meth., 186, 79–88 (1995).