

Arbeiten aus dem Paul-Ehrlich-Institut
11th International Paul-Ehrlich-Seminar, Bad Homburg 2005
Verlag Chmielorz

Analyzing of ID₅₀EAL Data for the Standardization of German Cockroach Allergen Extracts in the US

Robert James, Herman Mitchell, Peter J. Gergen, Peyton A. Eggleston, Jay E. Slater¹

Introduction

The ID₅₀EAL² method has been used for over 20 years as an *in vivo* estimate of overall allergen potency [1], and continues to be an accepted standard technique. In this method, allergenic extracts are evaluated in subjects maximally reactive to the respective reference concentrates. Each subject is tested with serial 3-fold dilutions of the reference extract. After 15 minutes, the sum of the longest and midpoint orthogonal diameters of erythema (ΣE) is determined at each dilution, and the log dose producing a 50 mm ΣE response (D_{50}) is calculated [2]. The mean D_{50} from these highly allergic individuals is used to determine the D_{50} for the extract. Extracts that produce similar D_{50} responses can be considered bioequivalent and are assigned similar units, the bioequivalent allergy unit (BAU). Because the modal D_{50} of a series of extracts was 14 (a ³⁻¹⁴ or 1:4.8 million dilution), extracts with a mean D_{50} of 14 were arbitrarily assigned the value of 100,000 BAU/mL [3]. Recently, the number of subjects required to generate significant potency data was reassessed [4].

In previous work, the raw ID₅₀EAL data were analyzed using linear regression of a log-linear plot, even though the data tended to be sigmoidal. Acceptance criteria included a Pearson correlation coefficient ($r \geq 0.92$) [5]. A practical consequence of this has been that many subjects' data are excluded, increasing the likelihood of selection bias. In this study, we attempted to overcome this limitation by assessing other methods of analyzing the ID₅₀EAL data.

¹ Drs. Gergen and Slater are US government employees. The opinions expressed in this paper are those of the authors, and do not represent the positions of the National Institutes of Health, the US Food and Drug Administration, or the US Department of Health and Human Services.

² Intradermal Dilution for 50 mm Sum of Erythema Determines Bioequivalent Allergy Units

Methods

The data analyzed were drawn from the Cockroach Allergen Standardization Evaluation of the Inner City Asthma Consortium, performed under IND 11319 (preliminary data reported at [6]). In brief, 61 otherwise healthy adults with asthma or allergy, and with a positive prick skin test to German cockroach allergen extract ($\Sigma E \geq 30$ mm at 15 min), were recruited in Baltimore, Chicago, Denver and Washington, DC. Three US commercial glycerinated German cockroach allergen extracts (heretofore referred to as extracts A, B and C) were tested concurrently on each subject.

Seventeen serial, three-fold dilutions of each extract were prepared by research pharmacists at each institution, using sterile saline (0.9%) containing human albumin (0.03%) and phenol (0.4%) (w/v; Hollister-Stier). All injections were 0.2 mL, performed intradermally on the back, using a 27g short-beveled needle; ΣE was determined at exactly 15 min. Subjects were initially injected with histamine control (0.1 mg/mL base; ALK-Abelló) and diluent, along with the 3^{-17} dilutions of each allergen extract. If the ΣE s for histamine and the diluent control were ≥ 35 mm and < 5 mm, respectively, testing was continued. For serial testing using the cockroach allergen extracts, subjects were advanced three three-fold dilutions if there was no response to the more dilute extract injection. On the other hand, for positive (≥ 5 mm) ^{TME} responses, the subjects were advanced one three-fold dilution. The objective was to establish a dose-response curve for each extract in each subject, with ^{TME} responses from 0 to 125 mm, containing at least four valid data points that bracketed 50 mm. The D_{50} s were calculated by: 1) linear regression, 2) nonlinear regression to a four-parameter logistic equation, and 3) linear interpolation.

Both the linear regression and nonlinear regression methods regressed the ΣE s over their three-fold dilution numbers and then used inverse prediction to estimate the D_{50} . The regression curves and their D_{50} s were estimated separately for each subject, after which the mean D_{50} s were determined for each extract by each analytical technique.

Linear regression was determined by standard techniques using the SAS statistical software package (v 9.1, SAS, Inc., Cary, NC). For nonlinear regression, the four-parameter logistic equation was fit individually to the data from each subject using Proc Nlin from SAS. The four-parameter logistic equation is as follows:

$$\Sigma(\text{erythema}) = \min + \frac{\max - \min}{1 + \left(\frac{EC50}{conc} \right)^p}$$

where: $conc = 3^{-\text{serial_dilution}}$

and $conc$ = the concentration of extract; serial_dilution = the 3-fold serial dilution number of the extract; $\Sigma(\text{erythema})$ = the sum of the minor and major erythema diameters (ΣE); $\min$ = minimum erythema diameter possible at $conc = 0$; $\max$ = max-

imum erythema diameter possible at $\text{conc} = \infty$; EC_{50} = extract concentration at which half of the change (max-min) in ΣE is expected; D = Hill coefficient or slope factor. (When $D=1$, the equation reduces to the three-parameter logistic equation.)

Proc Nlin fits the four-parameter logistic to each individual's data using the Gauss-Newton method. Since initial estimates can affect the final model parameter fit to the data, models were fit separately using three different initial values for each of the four parameters for a total of $3^4 (= 81)$ permutations of initial parameter values combinations used to fit the model to data of each individual subject. The initial values used to fit the models are listed below.

Parameter	Initial values	units
min	2, 15, 40	mm
max	10, 125, 500	mm
EC_{50}	1E-10, 5E-7, 1E-4	fraction of undiluted extract concentration
D	0.2, 1, 3	(none)

The best model fit among the 81 permutations of initial values for each subject was selected as the model fit with the lowest Sum of Squares Error.

Once the four-parameter logistic equation has been fit to the data, the concentration ($\text{conc}_{50\text{mm}}$) at which the logistic crosses the 50mm ΣE line can be calculated using the formula:

$$\text{conc}_{50\text{mm}} = \text{EC}_{50} \cdot \left(\frac{\text{max} - \text{min}}{50 - \text{min}} - 1 \right)^{-\left(\frac{1}{D}\right)}$$

Note that if the max parameter of the logistic is less than 50 mm the calculation of the D_{50} becomes impossible. The D_{50} is calculated from the $\text{conc}_{50\text{mm}}$ as:

$$D_{50} = - \left(\frac{\log(\text{conc}_{50\text{mm}})}{\log 3} \right)$$

The linear interpolation method of determining D_{50} used the serial three-fold dilutions immediately before and after the 50 mm ΣE threshold was reached. Then, through linear interpolation, the three-fold dilution required to achieve the 50 mm ΣE threshold was estimated. This otherwise simple approach cannot be used when the dose response curve fails to cross the 50 mm threshold, and becomes difficult when the curve crosses the 50 mm line more than once. The selection criteria for estimating the D_{50} for an extract when there are multiple crossings of the 50mm ΣE curve within a subject follow (please refer to Figure 1 for illustrative hypothetical examples).

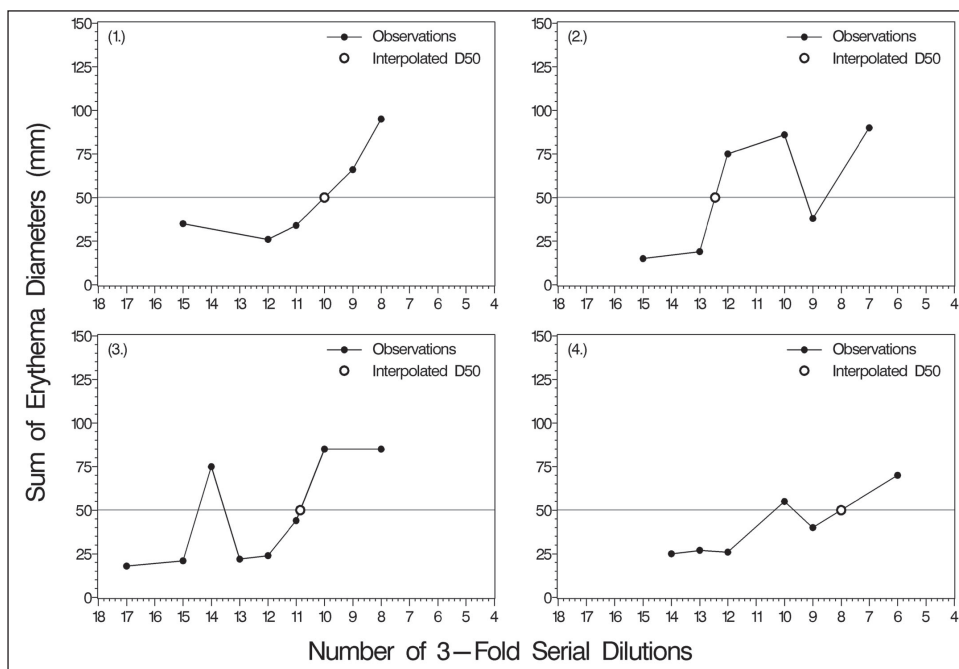


Fig. 1. Hypothetical examples illustrating the rules for determining the D_{50} using the linear interpolation method. Frame 1 illustrates the simplest situation, in which a graded increase in ΣE occurs, and the serial three-fold dilutions that bracket the point at which the ΣE crosses 50 mm are used to estimate the D_{50} value. Frames 2-4 illustrate the choice of D_{50} when there are erratic data with multiple crossings of the 50 mm ΣE line. In frame 2, the first 50 mm crossing is chosen as the D_{50} since the two consecutive data points immediately following the 50 mm crossing remain above 50 mm. In frame 3, the second 50 mm crossing is chosen as the D_{50} since the first crossing fails to stay above 50 mm for two consecutive data points. In frame 4, the data collection stopped before a second data point beyond the second 50 mm crossing was determined. Here, it is assumed that had a final additional data point beyond the crossing been collected, it probably would have been above the 50 mm line. Thus, the final crossing was chosen as the D_{50} .

1. Use the lowest interpolated D_{50} value at which the curve crosses the 50 mm line provided that ΣE stays above 50 mm for at least the next two greater injection doses.
2. If criteria #1 does not apply and the ΣE for the last and highest skin test concentration actually performed is greater than 50 mm, assume that any injections at higher concentrations (had they been performed) would have remained above 50mm and use this last crossing for calculating D_{50} .
3. If neither of these criteria applies, then the D_{50} cannot be calculated.

Results

Regardless of the analytical method used to calculate D_{50} , not all of the skin testing data could be used. In some instances, in spite of a positive screening prick test, intradermal testing failed to elicit measurable erythema responses. In others,

responses were obtained but were erratic. For these reasons, out of 61 subjects, the A extract D_{50} s for 12 subjects, the B extract D_{50} s of seven subjects, and the C extract D_{50} s of five subjects could not be estimated using any of the three analytic methods.

Figures 2, 3, and 4 show the D_{50} estimates of data from selected subjects. Figure 2 plots selected subject data where the D_{50} calculations are similar among the three analytic methods; Figure 3 illustrates subject data where the D_{50} determinations dif-

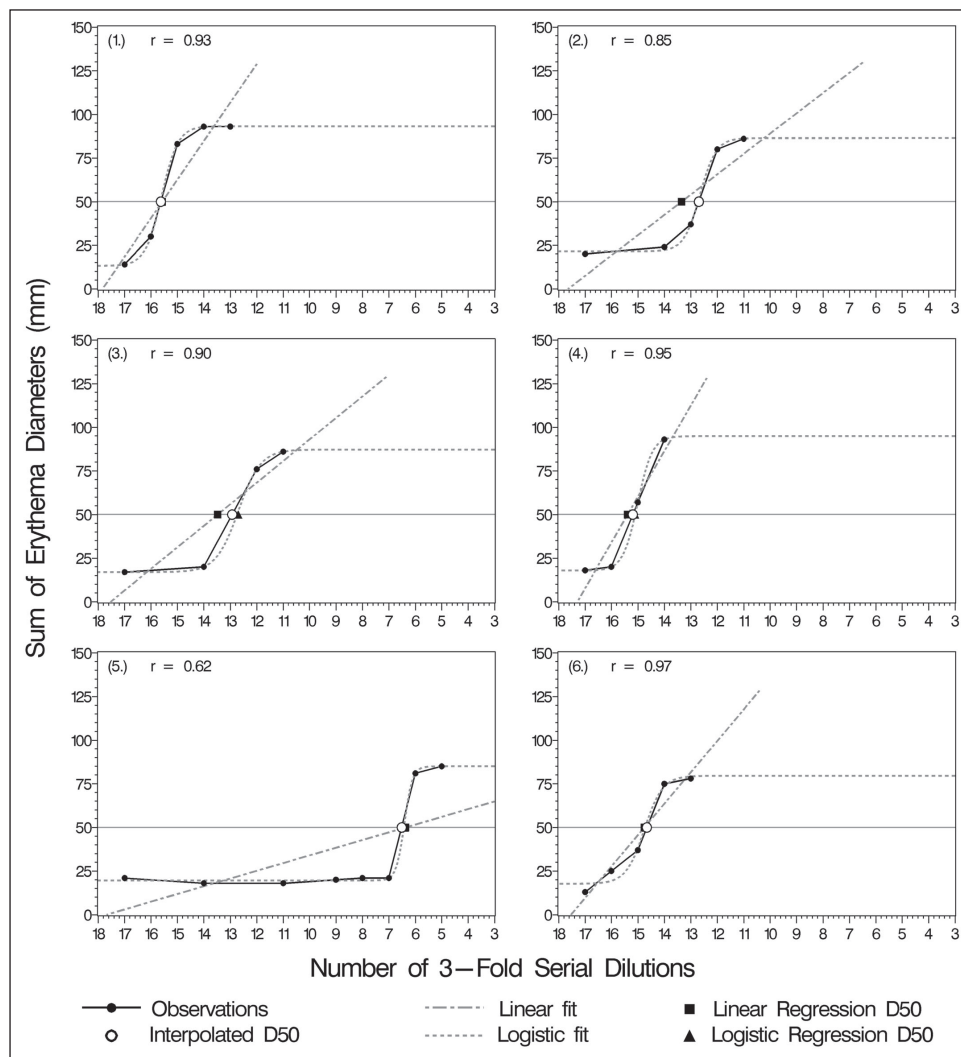


Fig. 2. Selected experimental data examples where the D_{50} determinations are similar among the three methodologies. The Pearson correlation coefficient, r , is included on the plots, since the accepted linear regression method only includes D_{50} values from data with $r \geq 0.92$. If the $r \geq 0.92$ inclusion criterion is applied, the linear regression D_{50} s in frames 2, 3 and 5 are not considered valid.

Fig. 3. Selected experimental data where the D_{50} determinations differ among the three methodologies. The Pearson correlation coefficient, r , is included on the plots, since the accepted linear regression method only includes D_{50} values from data with $r \geq 0.92$. If the $r \geq 0.92$ inclusion criterion is applied, the linear regression D_{50} s in frames 3 through 8 are not considered valid.

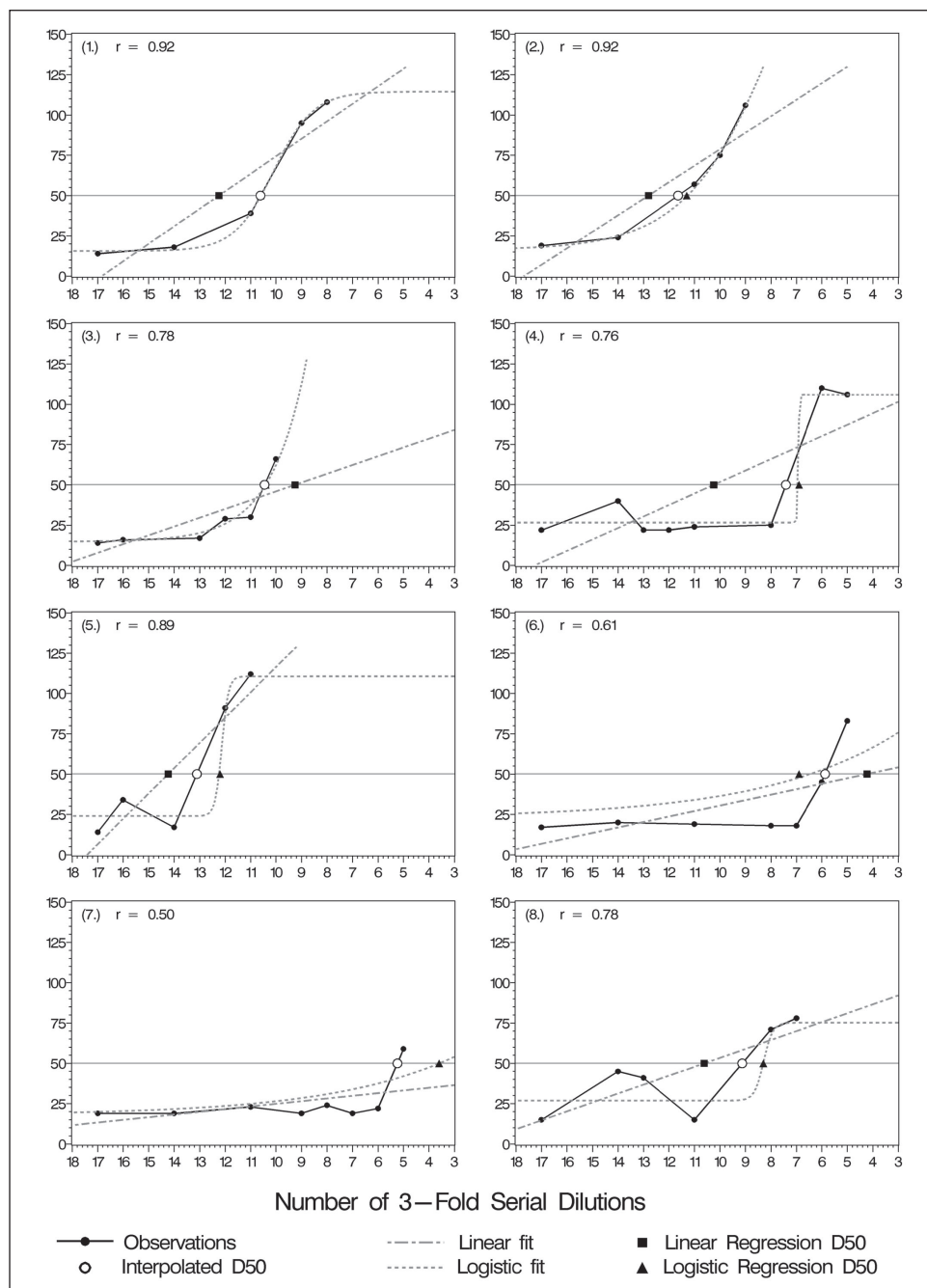
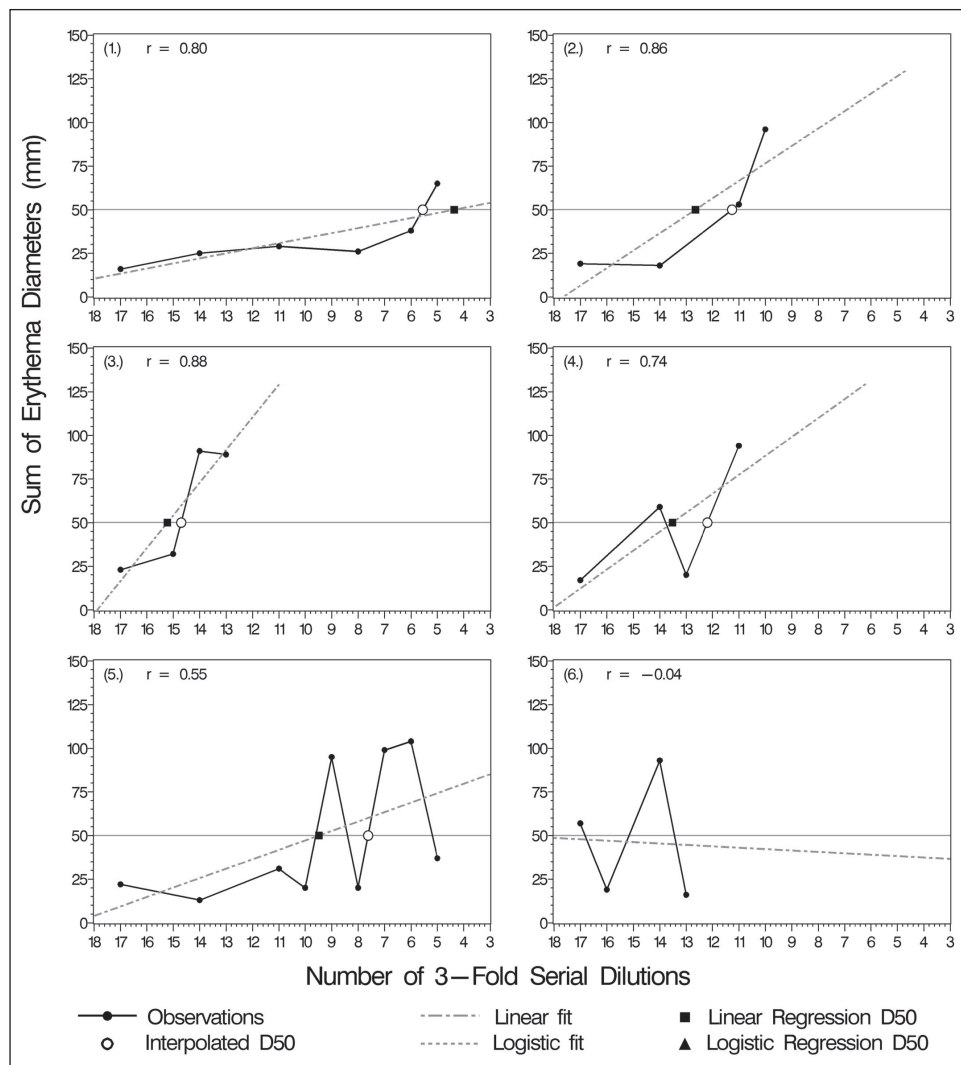


Fig. 4. Selected experimental data where the logistic regressions failed to fit the four-parameter logistic model to the data, and the D_{50} s using the logistic regression method could not be determined. The Pearson correlation coefficient, r , is included on the plots, since the accepted linear regression method only includes D_{50} values from data with $r \geq 0.92$. Note that in frame 6, D_{50} could not be determined by any of the three methods.



fer among the three analytic methods; and Figure 4 shows selected subject data where the logistic regression models failed to fit the data, preventing any logistic estimation of the D_{50} . As expected, when the D_{50} s were estimable, the curves used to estimate the D_{50} s tended to be sigmoidal in shape (e.g. Figure 2: frames 1, 2, 3, 5, 6; Figure 3: frames 1 and 4). Other curves, if not sigmoidal, were at least consistent with the possibility that they may have become sigmoidal had the data been collected further to higher concentrations (lower dilutions).

When the data curves appeared sigmoidal with the higher data points well balanced against the lower data points on the curve, all three analytic methods tended to produce very similar D_{50} values (for example, Figure 2). However, when the upper part of the sigmoidal curve appeared truncated (for example, Figure 3: frames 2, 3, 6, 7; Figure 4: frames 1 and 2), the linear regression curves tended not to fit the data well and under- or over-estimated the D_{50} concentrations.

The figures show that the 4-parameter logistic equation to fit the data well and its D_{50} estimates tended to be in general agreement with the linear interpolation D_{50} values. However, the logistic equation often failed to fit the data if the logistic shape was not well defined (Figure 4). Thus, only 128 D_{50} s out of 183 extract data samples (three extracts for each of the sixty-one subjects) were estimable using the nonlinear logistic method.

The mean D_{50} for each of the three analytic methods gave comparable results (Table 1). However, the standard deviations of the linear regression method were as much as 60% larger than the linear interpolation or nonlinear logistic regression methods. Since selection bias could possibly account for differences among the analytic methods, Table 2 is shown summarizing the D_{50} results from only those subjects and extracts whose D_{50} s were estimable using all three analytic methods. This avoids possible selection bias that may have otherwise resulted if the subset of subjects differed among the three analytic methods. The D_{50} means in Table 2 are in general agreement with those in Table 1. Again, the linear regression method of determining the D_{50} has greater variance than either the linear interpolation or nonlinear logistic regression analytic methods.

Our estimates of the D_{50} s using the linear regression methodology included all linear fits to the data. However, if we limit our linear regression D_{50} estimates to only those with Pearson correlation coefficients of $r \geq 0.92$, the means and standard devi-

extract	analytic method	n	calculated D_{50}		
			mean	std. dev.	std. error
A	linear regression	49	10.22	4.33	0.62
	linear interpolation	49	10.24	2.68	0.38
	logistic regression	43	10.13	2.66	0.41
B	linear regression	54	11.87	3.19	0.43
	linear interpolation	54	11.65	2.81	0.38
	logistic regression	43	11.29	2.60	0.40
C	linear regression	55	10.92	3.49	0.47
	linear interpolation	55	10.64	2.54	0.34
	logistic regression	42	10.69	2.33	0.36

Table 1. Summary statistics of the D_{50} s calculated using each of the three analytic methods. Data from all subjects and analytic methods in which any of the D_{50} s were estimable, including subjects with $r < 0.92$, have been included.

extract	analytic method	n	calculated D_{50}		
			mean	std. dev.	std. error
A	linear regression	43	10.30	4.09	0.62
	linear interpolation	43	10.24	2.65	0.40
	logistic regression	43	10.13	2.66	0.41
B	linear regression	43	11.70	3.04	0.46
	linear interpolation	43	11.41	2.61	0.40
	logistic regression	43	11.29	2.60	0.40
C	linear regression	42	11.18	2.77	0.43
	linear interpolation	42	10.72	2.30	0.35
	logistic regression	42	10.69	2.33	0.36

Table 2. Comparison of D_{50} analytic methods using only the subset of data from subjects and extracts whose D_{50} s were estimable using each of the three analytic methods. This avoids possible selection bias that could have resulted if a different subset of subjects were being compared among the three analytic methods.

extract	analytic method	n	calculated D_{50}		
			mean	std. dev.	std. error
A	linear regression	11	12.29	2.11	0.64
	linear interpolation	11	11.97	2.04	0.61
	logistic regression	11	11.81	2.13	0.64
B	linear regression	12	14.13	1.38	0.40
	linear interpolation	12	13.87	1.49	0.43
	logistic regression	12	13.69	1.55	0.45
C	linear regression	10	12.54	1.58	0.50
	linear interpolation	10	12.20	1.79	0.57
	logistic regression	10	11.99	2.00	0.63

Table 3. Comparison of analytic methods of determination of D_{50} comparing only subjects and extracts where the linear $r \geq 0.92$. Note that 40 out of 183 D_{50} evaluations had $r \geq 0.92$ when using linear regression. Of these 40, however, only 33 assessments also had D_{50} s that could be estimated using both the linear interpolation and nonlinear logistic regression analytic methods. To avoid bias from data missing for one analytic method but not for another, only subject/extract data that could estimate the D_{50} s from all three analytic methods were included in this table.

ations of the three analytic methods become equivalent (Table 3) with their mean D_{50} values differing ≤ 0.55 serial dilution units and standard deviations < 0.42 . Unfortunately, only 40 of the 183 D_{50} estimates had good enough fits to the data to meet the $r \geq 0.92$ acceptance criterion due to the sigmoidal shapes of our curves. Thus, limiting our estimates to only $r \geq 0.92$ would severely cut the number of acceptable D_{50} estimates to less than a quarter of our subjects.

The linear interpolation method appeared to estimate the D_{50} s at least as well as the nonlinear regression method. In addition, using the linear interpolation method we were able to estimate the D_{50} s for every subject and extract whose Σ Es crossed and stayed above the 50 mm threshold, calculating 158 D_{50} s out of the 183 extract data samples. In contrast, using the nonlinear logistic and linear regression (with $r \geq 0.92$) analytic methods, only 128 and 40 of the D_{50} s, respectively, were estimable.

Discussion

The small number of acceptable D_{50} estimates (only 40 out of 183 assays) using the linear regression methodology differs from that reported previously with only 1 out of 22 assays failing to meet the $r \geq 0.92$ inclusion criterion [1]. This may be explained by the fact that sigmoidal curves are very linear in the neighborhood of their inflection points. In our test procedure, we carefully began testing for allergic responses at very low concentrations and then gradually increased the dose until the required response was observed. This is good practice in that it prevents overshooting the allergen concentration and missing the D_{50} estimate. For the linear regression estimates of the D_{50} to be as precise as the linear interpolation or nonlinear logistic regression analytic methods (as in Table 3), the linear portion of the sigmoidal response curve would need to be known a priori for each subject tested. Given the impracticality of prejudging the shape of the response curve, our approach of assaying at the lower nonlinear portions of the response curve is more reasonable.

Even when performed under carefully controlled circumstances, quantitative skin test responses to allergens can be difficult to measure with accuracy, and log-dose responses are not always linear or even sigmoidal. By using a relatively simple linear interpolation technique, which makes no assumption of a fixed mathematical relationship (or the linear portion thereof) between the log-dose and the skin test response, we could capture data from a greater number of study subjects, and our D_{50} estimates more closely approximate the mean value. The fewer missing values reduced the standard errors and the bias potential associated our analyses. Furthermore, the D_{50} values that we obtained using linear interpolation were very close to the ones obtained using the other analytical methods. Finally, the linear interpolation method does not require sophisticated software and may even be estimated using graph paper.

For these reasons – along with its intuitive and analytical simplicity – we have used the D_{50} values obtained with the linear interpolation technique for testing bioequivalence and all further analyses.

References

1. Turkeltaub PC, Rastogi SC, Baer H, Anderson MC, Norman PS. A standardized quantitative skin-test assay of allergen potency and stability: Studies on the allergen dose-response curve and effect of wheal, erythema, and patient selection on assay results. *J Allergy Clin Immunol* 1982; 70: 343-352.
2. Turkeltaub PC. Allergen vaccine unitage based on biological standardization: Clinical significance. In: *Allergens and Allergen Immunotherapy*. Lockey R, Bukantz SC (Eds). Marcel Dekker Inc, New York 1999; pp 321-340.

3. Turkeltaub PC. Biological standardization of allergenic extracts. *Allergol Immunopathol (Madr)* 1989; 17: 53-65.
4. Rabin RL, Slater JE, Lachenbruch P, Pastor RW. Sample size considerations for establishing clinical bioequivalence of allergen formulations. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 2003; 94: 24-33.
5. Methods of the allergenic products testing laboratory. Docket 94N-0012. Federal Register 11-23-1994.
6. Eggleston PA, Liu AH, Pongracic JA, Sarpong SB, James R, Gergen PJ, Mitchell H, Slater JE, for the Inner City Asthma Consortium. Creating a standardized cockroach extract. *J Allergy Clin Immunol* 2005; 115 : S73.

Discussion

Malling: I think that your consideration about using interpolation or linear regression in relation to the skin test is interesting. My experience, of course, is that if the technique is optimal you will always get a linear curve in a double logarithmic system. So, when curves are like the ones you showed, there is probably a technical problem or it is an inadequate system.

Slater: In a study like this, one of the major advantages of working with the Inner City Asthma Consortium was they have a really outstanding nursing staff at NIH. The nurses went to each of the sites and individually trained the testers. They were all qualified to do the skin testing protocol. But you are right. Despite best efforts I am sure that what we were observing was in part a technical problem.

Malling: I think that one should consider to exclude these values from the calculations.