Chapter 6

Limiting Dilution Assays for the Separation, Characterization, and Quantitation of Biologically Active Particles and Their Clonal Progeny

CARL TASWELL

Department of Biostatistics, School of Public Health, Harvard University, Boston, Massachusetts 02115

I. Introduction

Since the beginning of this century, limiting dilution assays (LDAs)[1] have been used to quantitate a wide variety of biologically active particles (BAPs) including bacteria (Phelps, 1908), protozoa (Cunningham, 1915), viruses (Clark, 1927), tumor cells (Hewitt, 1958), immunocompetent cells (Makinodan and Albright, 1962), and neurocompetent cells (Barbarese *et al.*, 1983). LDAs can also be used to quantify the effectiveness of purification and depletion procedures (Taswell *et al.*, 1979) and to separate and characterize BAPs and their clonal progeny (Taswell *et al.*, 1980). Recent articles by Taswell (1981, 1984a,b) presented basic principles of LDAs, reviewed existing methods, and introduced new methods for the problems of model discrimination, parameter estimation, and design optimization. This article attempts to collect in one publication all methods of statistical analysis relevant to LDAs and to present them in a unified manner with a common terminology and notation.

Throughout most of their history, LDAs have been known as dilution assays, serial dilution assays, dilution series, dilution tests, fermentation tube tests, coliform density tests, etc., and limiting dilution analysis as the dilution method, dilution series method, fermentation tube technique, multiple tube method, multitube fermentation method, etc. It was only relatively recently that immunologists began using the newer terms "lim-

[1] Abbreviations: LDA, Limiting dilution assay; SHPM, single-hit Poisson model; BAP, biologically active or assayable particle; BIP, biologically inactive particle; POP, physicochemically observable particle; CTL-P, cytolytic T lymphocyte precursor; MLC, mixed leukocyte culture; LS, least squares; WM, weighted mean; ML, maximum likelihood; MC, minimum chi-squared; V, variance; CI, confidence interval; CV, coefficient of variation; df, degrees of freedom; pdf, probability density function; and all other mathematical and statistical notation as explained in the text.

CELL SEPARATION: METHODS AND SELECTED APPLICATIONS, VOL. 4

iting dilution assays" (Kennedy *et al.*, 1966) and "limiting dilution analysis" (Groves *et al.*, 1970). Most bacteriologists, virologists, public health officials, and sanitary engineers still use the older terms (Wilson, 1983; Greenberg *et al.*, 1985). Of all the different names for these bioassays, the term "limiting dilution assay (LDA)" is the most descriptive, the most general, and therefore the most appropriate for the collection of assays as a class. The words "limiting dilution" emphasize two important and related aspects of this class: (1) the assays are based on a process of dilution of the dose to extinction of the response, and (2) this process requires that only the BAP to be quantitated is diluted to these limiting doses while all other culture system constituents are provided at saturating (nonlimiting) doses.

Assuming that certain fundamental hypotheses (Section I,B) are validated for each case, the same methods of statistical analysis apply to all LDAs regardless of the kind of BAP diluted in liquid suspension. Indeed, these methods also apply to procedures used to quantitate BAPs that are not suspended in liquid. Botanists, ecologists, and foresters quantitate plants on tracts of land; their observational studies analogous to LDAs are called stocked-quadrat surveys (Blackman, 1935; Swindel, 1983). Agricultural and veterinary scientists quantitate viruliferous insects in a vector population capable of transmitting viral, bacterial, fungal, or parasitic diseases to plant and animal hosts; their experimental procedures analogous to LDAs are apparently not known by any particular name (Thompson, 1962; Kerr, 1971). All of these LDAs and analogous procedures are dose–response assays that detect quantal responses and require dilution of the dose to extinction of the response. They must be distinguished from a related class of assays (such as plate, colony, plaque, and pock count assays) that detect quantitative responses and do not require dilution to extinction. Different methods of statistical analysis apply to this related but distinct class of assays (Fisher *et al.*, 1922; Roberts and Coote, 1965). These methods cannot be used for LDAs.

A. Limiting Dilution Assays (LDAs)

LDAs detect binary (positive or negative) responses generated by BAPs in individual *in vivo* or *in vitro* cultures within groups of replicate cultures that vary in the dose of the test preparation from which the BAPs are sampled. LDAs can be used to estimate the absolute number of BAPs [called the most probable number, MPN, or density of coliform organisms by bacteriologists (Phelps, 1908; Wilson, 1983)], the 50% endpoint on the dilution scale of the BAP test preparation [the dilution level at which the group of replicates is 50% positive and 50% negative (Reed and Muench, 1938; Worcester, 1954)], and the relative frequency of BAPs [called the immunocompetent cell frequency by immunologists (Taswell, 1981)]. These three parameter estimates are obtained from two subclasses of LDAs: subclass I consists of all LDAs that can be used to calculate absolute numbers and 50% endpoints but not relative frequencies, and subclass II consists of all LDAs that can be used to calculate all three parameter estimates. Absolute numbers are expressed as the number of BAPs per unit volume of test preparation, 50% endpoints as units on the dilution scale of the BAP test preparation, and relative frequencies as the proportion of BAPs within a mixture of biologically active and inactive particles (BAPs and BIPs). Dilution levels for 50% endpoints may be necessary for drug or antisera titration studies but not appropriate for any study where it is possible and meaningful to estimate the absolute number of BAPs (because a dilution level is clearly less informative than an absolute number). Therefore, they will not be considered further in this article.

The two subclasses can then be designated by their distinguishing parametric estimates as absolute number LDAs (subclass I) and relative frequency LDAs (subclass II). Both absolute number and relative frequency LDAs are biological assays for particles of a specific type defined by their functional activity and called biologically active or assayable particles (BAPs). Relative frequency LDAs, however, also incorporate an accompanying physicochemical assay for particles of a general type defined by their structural morphology or other physicochemical characteristics and called physicochemically observable particles (POPs). As an example, LDAs are used to measure the relative frequency of cytolytic T lymphocyte precursors (CTL-Ps) as the BAPs within a mixture of leukocytes as the POPs (Taswell *et al.*, 1979). In this example, the functional activity of the BAPs is defined as cell differentiation and proliferation producing a clone of cells that can kill target cells (assayed indirectly by ^{51}Cr release), while the structural morphology of the POPs is defined as standard leukocytic morphology (observed directly by light microscopy). In absolute number LDAs, the number of POPs (theoretically equal to the number of BAPs plus BIPs) is never known because any physicochemical assay that could conceivably be used to observe them is not performed due to impracticality or impossibility.

B. The Single-Hit Poisson Model (SHPM)

To analyze dose–response data from LDAs, it is necessary to validate a model incorporating two fundamental hypotheses: one for the provision of the dose and the other for the generation of the response. For the sampling of BAPs aliquoted to replicate cultures, first McCrady (1915) assumed a binomial distribution hypothesis and then Greenwood and Yule (1917) a Poisson distribution hypothesis. For the generation of a

positive response in the test subjects (*in vivo* or *in vitro* replicate cultures), all of the original investigators assumed implicitly (but never stated explicitly) the single-hit hypothesis: a positive response is generated in every test subject that receives at least one BAP. Subsequent investigators, including Iwaszkiewicz and Neyman (1931), Worcester (1954), Lefkovits and Waldmann (1979), and Taswell (1984b), considered other hypotheses for the response generation process such as multiple-hit, multiple-target, variable-subject, false-positive, and false-negative phenomena.

Thus, the dose–response model originally proposed by Greenwood and Yule (1917), and recently called the single-hit Poisson model (SHPM) by Taswell (1981), incorporates the Poisson distribution hypothesis for the provision of the dose and the single-hit hypothesis for the generation of the response. The mathematical formulation of the SHPM is

$$\theta_d = \exp(-\omega_d) = \exp(-\phi\lambda_d) \tag{1}$$

where θ_d is the negative response probability for each subject in the dth dose group of replicate subjects, ω_d is the unknown mean number of BAPs provided to each subject of the dth dose group, ϕ is the unknown parameter to be estimated, and λ_d is the known parameter for the dose level of the dth dose group. For absolute number and relative frequency LDAs, there are different interpretations for ϕ and λ_d in the identity $\omega_d = \phi\lambda_d$. For absolute number LDAs, ϕ is the unknown absolute number ($0 \leqq \phi \leqq \infty$) of BAPs in the test preparation; and λ_d is a known varying dilution factor for constant volume samples or a known varying volume for constant density samples of the test preparation provided to test subjects. For relative frequency LDAs, ϕ is the unknown relative frequency ($0 \leqq \phi \leq 1$) of BAPs in the test preparation; and λ_d is the known mean number of POPs from the test preparation provided to test subjects.

To estimate the unknown parameter ϕ with the SHPM, the theoretically predicted negative response probability θ_d must be approximated by the experimentally observed negative response fraction p_d according to the equation

$$r_d/n_d = p_d \approx \theta_d = \exp(-\phi\lambda_d) \tag{2}$$

where r_d and n_d are, respectively, the number of negatively responding subjects and the number of replicate subjects for the dth dose group. After a logarithmic transformation and then substitution with the linear regression variables X_d and Y_d and slope parameter β, Eq. (2) becomes

$$Y_d = \ln(r_d/n_d) \approx -\phi\lambda_d = -\phi X_d = \beta X_d$$

which shows that the SHPM is a member of the class of generalized linear models (McCullagh and Nelder, 1983).

C. Sample and Population LDAs

Due to the unstable nature of most biological preparations, it is generally impractical, if not impossible, to assay repeatedly the same biological sample. It is possible, however, to assay repeatedly the same biological population by performing, either simultaneously or serially, a sequence of LDAs on a sequence of biological samples with one assay for each sample. Two or more assays can be performed for each sample, but only simultaneously and not serially, because of the biological instability of the sample. If they are performed simultaneously, however, then the data from the multiple simultaneous assays can be combined and analyzed as if obtained from a single larger assay. Assays and samples are thus considered to be always in one-to-one correspondence with each other. With this convention of one-to-one correspondence, the assays/samples are indexed by a, and ϕ_a is interpreted with respect to the sample as a fixed parameter but with respect to the population as a realized value of a random variable Φ_a with a distribution with mean μ_Φ and variance σ^2_Φ.

LDAs can then be described as sample or population assays according to the number A of assays/samples in the sequence or subsequence under consideration. For sample LDAs, there is only one ($A = 1$) assay/sample. They are used to obtain sample estimates $\hat{\phi}$ of sample parameters ϕ. For population LDAs, there are two or more ($A \geqq 2$) assays/samples with indices a such that $1 \leqq a \leqq A$. They are used to obtain both sample estimates $\hat{\phi}_a$ of sample parameters ϕ_a and population estimates $\hat{\mu}_\Phi$ and $\hat{\sigma}^2_\Phi$ of population parameters μ_Φ and σ^2_Φ. LDAs can also be described as single-dose or multiple-dose assays according to the number D_a of dose groups in the ath assay. For single-dose LDAs, there is only one ($D_a = 1$) dose group. For multiple-dose LDAs, there are two or more ($D_a \geqq 2$) dose groups. The description of LDAs as sample or population assays and as single-dose or multiple-dose assays is helpful when discussing the selection of statistical methods that are appropriate for different situations or for different interpretations of the same situation. For example, the same LDA can be interpreted as a sample LDA or as one of a sequence of population LDAs. Both assay design and data analysis may require different methods according to the different interpretations.

II. Sample LDAs: SHPM Validity Tests

Principles of bioassay validity tests and the distinction between model-discrimination experiments and parameter-estimation assays were discussed recently by Taswell (1984b). References to the literature on SHPM

validity tests for LDAs can be found in the reviews by Taylor (1962), Finney (1964), and Taswell (1984b). These tests detect deviations from the SHPM for parameter-estimation assays with a known culture system where the SHPM is assumed to apply most of the time. They do not select or "prove" one of many alternative models for model-discrimination experiments with an unknown culture system where a dose–response model has not yet been demonstrated. Thus, they are used as assay-screening tests, enabling faulty assays to be discarded and replaced while indicating the possible nature of the defect in the faulty assay.

A. Methods

Data from LDAs are validity tested for goodness of fit with the SHPM by calculating χ^2 statistics and confidence intervals (CI) for the slope parameter β from the generalized linear models 1 and 2

$$\ln \hat{\phi}_d = \alpha_1 + \beta_1 \lambda_d \tag{3}$$

$$\hat{\phi}_d = \alpha_2 + \beta_2 / \lambda_d \tag{4}$$

to determine whether β_1 and β_2 differ significantly from zero. These models regress $Y_d = \ln \hat{\phi}_d$ on $X_d = \lambda_d$ and $Y_d = \hat{\phi}_d$ on $X_d = 1/\lambda_d$ by weighted least squares where

$$\hat{\phi}_d = -(\ln p_d)/\lambda_d$$

is the dth single-dose estimate of ϕ. This generalized linear modeling approach to SHPM validity testing for LDAs was developed by Armitage and Spicer (1956), Gart and Weiss (1967), and Taswell (1984b). Consult Gart and Weiss (1967) for the derivation of model 1 and Taswell (1984b) for the derivation of model 2.

1. *Validity of the Assay*

a. The Chi-Squared Slope Statistic. Consider the weighted least-squares regression of Y_d on X_d for the linear model

$$Y_d = \alpha + \beta X_d$$

with weights $w_d = 1/\hat{V}(Y_d)$ where $\hat{V}$ is the estimate of the variance V. Then a $(1 - \alpha)100\%$ confidence interval (CI) for each Y_d is given by $(1 - \alpha)100\%$ CI$(Y_d) = Y_d \pm z_{\alpha/2}/w_d^{1/2}$. Define the weighted means

$$\bar{Y}_w = \sum w_d Y_d \Big/ \sum w_d, \qquad \bar{X}_w = \sum w_d X_d \Big/ \sum w_d$$

with sums over d for $1 \leq d \leq D$. Define the deviations from the sample means $y_d = Y_d - \bar{Y}_w$ and $x_d = X_d - \bar{X}_w$. Then estimates for the slope β, its variance $V(\beta)$, and the intercept α are given by

$$\hat{\beta} = \sum w_d x_d y_d \Big/ \sum w_d x_d^2$$

$$\hat{V}(\hat{\beta}) = \left(\sum w_d y_d^2 - \hat{\beta} \sum w_d x_d y_d\right) \Big/ \left[(D - 2) \sum w_d x_d^2\right]$$

$$\hat{\alpha} = \bar{Y}_w - \hat{\beta} \bar{X}_w$$

with the test statistics

$$\chi^2_{\text{slope}} = \left(\sum w_d x_d y_d\right)^2 \Big/ \sum w_d x_d^2 = \hat{\beta} \sum w_d x_d y_d$$

$$t = \hat{\beta} \Big/ \left[\hat{V}(\hat{\beta})\right]^{1/2}$$

observing the χ^2 distribution with 1 degree of freedom (df) and the t distribution with $D - 2$ df, respectively, for the null hypothesis that $\beta = 0$. A $(1 - \alpha)100\%$ CI for β is given by

$$(1 - \alpha)100\% \text{ CI}(\beta) = \hat{\beta} \pm t_{\alpha/2}[\hat{V}(\hat{\beta})]^{1/2}$$

where $t_{\alpha/2}$ is the two-tailed α level significant value of t with $D - 2$ df; P values for test statistics can be calculated using the algorithms reviewed by Maindonald (1984).

b. Model 1. Set $Y_d = \ln \hat{\phi}_d$, $X_d = \lambda_d$, and

$$w_d = n_d p_d (\ln p_d)^2 / (1 - p_d) \tag{5}$$

excluding dose groups with $p_d = 0$ or $p_d = 1$ because they result in $w_d = 0$ or in indeterminate forms. Perform weighted least-squares regression and calculate the CI for the slope β, and the χ^2_{slope} statistic $\chi^2_{\text{model 1}}$ under the null hypothesis that $\beta_1 = 0$. A negative value for $\hat{\beta}_1$ with a significant P value from the 1 df $\chi^2_{\text{model 1}}$ test suggests the presence of a dose–response effect consistent with a variable number (dependent on λ_d) of false negatives. A positive value for $\hat{\beta}_1$ with a significant P value from the 1 df $\chi^2_{\text{model 1}}$ test suggests the presence of a dose–response effect consistent with a multihit and/or multitarget response-generation process.

c. Model 2. Set $Y_d = \hat{\phi}_d$, $X_d = 1/\lambda_d$, and

$$w_d = n_d p_d \lambda_d^2 / (1 - p_d) \tag{6}$$

excluding dose groups with $\lambda_d = 0$, $p_d = 0$, or $p_d = 1$ because they result in $w_d = 0$ or in indeterminate forms. Perform weighted least-squares regression and calculate the CI for the slope β_2 and the χ^2_{slope} statistic $\chi^2_{\text{model 2}}$ under the null hypothesis that $\beta_2 = 0$. A positive value for $\hat{\beta}_2$ with a significant P value from the 1 df $\chi^2_{\text{model 2}}$ test suggests the presence of a dose–response effect consistent with a constant number (independent of λ_d) of false positives.

2. *Validity of the Estimate*

If the SHPM is not rejected by the assay validity tests (Section II,A,1), then a sample estimator such as $\hat{\phi}_{MC}$, $\hat{\phi}_{ML}$, or $\hat{\phi}_{WM}$ based on the SHPM

should be used to calculate the desired estimate of the unknown sample parameter ϕ (Section III,A). As a test of the validity of this final estimate $\hat{\phi}$, set

$$\hat{p}_d = \exp(-\hat{\phi}\lambda_d)$$

and substitute $\hat{p}_d$ for p_d in Eqs. (5) and (6) for the weights w_d. Calculate the χ^2_{slope} statistics otherwise as above for the assay validity tests. Choose the statistic $\chi^2_{model\ 1}$ or $\chi^2_{model\ 2}$, whichever is larger with a corresponding P value that is smaller, and report this statistic

$$\chi^2_{slope} = \max\{\chi^2_{model\ 1}, \chi^2_{model\ 2}\}$$

and its P value as the statistic for the final estimate validity test. Note that assay validity tests check goodness of fit between the model and the assay data whereas estimate validity tests check goodness of fit between the model, the assay data, and the final estimate.

B. Examples

Table I presents data (adapted from Fig. 1 of Taswell *et al.*, 1979) from relative frequency LDAs for determinations of CTL-P frequencies in normal murine spleen and thymus cell test preparations (C57BL/6 responding cells cultured with DBA/2 stimulating cells and assayed against P815 target cells). Table II presents assay validity test statistics and their P values for the spleen and thymus data from Table I. Since the SHPM is not rejected, the assays are considered valid. By the principle of parsimonious use of parameters, the estimate to be calculated should be an estimate derived from the single-parameter SHPM [Eq. (1)] and not from the multiple-parameter models [Eqs. (3) and (4)].

C. Selection of a Test

One df χ^2 regression slope tests for the SHPM (Section II,A; Taswell, 1984b) are more powerful than $D - 1$ df χ^2 general goodness-of-fit tests for the SHPM (Taswell, 1981). They discriminate against several alternative models whereas the likelihood ratio test derived by Cox (1962) discriminates against only one alternative model. They are more efficient and versatile than the validity tests published decades ago (cf. reviews cited above). Finally, as tests of linear regression slope parameters, they are readily amenable to graphical presentation for investigators who so desire, though they should draw their conclusions from values of test statistics and not appearances of graphs. Since the χ^2_{slope} tests are the most practical yet fully efficient validity tests for the SHPM, they should be used in preference to other tests.

TABLE I
Data from Sample LDAs for the Determination of Murine CTL-P Frequencies[a]

Assay	d	λ_d	r_d/n_d	p_d	$\hat{\phi}_d \times 10^4$
Spleen	1	250	17/24	0.708	13.79
	2	500	13/24	0.542	12.26
	3	750	6/24	0.250	18.48
Thymus	1	100	32/33	0.970	3.08
	2	500	28/33	0.848	3.29
	3	1000	24/33	0.727	3.18
	4	1500	21/33	0.636	3.01
	5	2000	14/33	0.424	4.29

[a]Refer to Section II,B for a discussion of this table. d, Index of dose group; λ_d, known mean number of POPs (which are total leukocytes in these assays); r_d, number of negatively responding cultures; n_d, number of replicate cultures; p_d, fraction of negatively responding cultures; $\hat{\phi}_d$, single-dose estimate of the relative frequency ϕ of BAPs (which are CTL-Ps in these assays). For a discussion of basic terms, refer to Section I,A for absolute number and relative frequency LDAs and for BAPs and POPs, to Section I,B for the SHPM, and to Section I,C for sample and population LDAs and the corresponding sample and population parameters and estimates.

TABLE II
Assay Validity Test Statistics for LDA Data from Table I[a]

Assay	Model	$\hat{\beta}$	95% CI (β)	χ^2_{slope}	P
Spleen	1	0.7×10^{-3}	-7.1–8.6×10^{-3}	0.67	0.41
	2	-0.1	-3.2–3.0	0.19	0.67
Thymus	1	2.0×10^{-4}	-1.7–5.7×10^{-4}	0.56	0.45
	2	-0.7×10^{-2}	-6.6–5.3×10^{-2}	0.04	0.84

[a]Refer to Section II,B for a discussion of this table; and to Section II,A,1 for an explanation of the generalized linear modeling approach to assay validity tests for the SHPM. $\hat{\beta}$ and 95% CI (β), estimates of the slope β of the regression line from alternative generalized linear models 1 and 2; χ^2_{slope} and P, a 1 df χ^2 statistic and its P value which test the significance of the deviation of the slope β from 0. When the SHPM is valid, β should theoretically equal 0.

III. Sample LDAs: Sample Estimators

Most of the literature published on LDAs since their introduction almost 80 years ago has been devoted to a diverse variety of data reduction methods for estimating the sample parameters ϕ and CI(ϕ). References to this literature can be found in reviews by Halvorson and Ziegler (1933), Eisenhart and Wilson (1943), Finney (1952), Taylor (1962), Cornell and Speckman (1967), and Loyer and Hamilton (1984). All of these published estimators can be classified as graphical, tabular, or computational and as design restricted or design unrestricted. An estimator is design restricted if it can be used only for assays with specified designs usually limited to small numbers of dose groups and replicates and/or equal numbers of replicates or to constant linear or exponential intervals between the dose levels.

A. Methods

Weighted mean (WM), maximum likelihood (ML), and minimum chi-squared (MC) estimators are design-unrestricted computational methods for calculating the sample estimate $\hat{\phi}$. Weighted averaging for LDAs was used by Barkworth and Irwin (1938), but formulas for the weights were not specified. These equations were derived and published by Taswell (1981). Likelihood maximization for LDAs was developed by Fisher (1922), Halvorson and Ziegler (1933), and Finney (1951). It was implemented by Peto (1953) in its present version with solution of the log likelihood (ln L) equation by Newton's method of iterative approximation. Chi-squared minimization for LDAs was proposed and implemented by Taswell (1981). Consult this latter reference for derivations of the formulas and explanations of the calculations for these three estimators. The noniterative WM estimate $\hat{\phi}_{WM}$ is used as the starting value for approximations to the iterative ML and MC estimates $\hat{\phi}_{ML}$ and $\hat{\phi}_{MC}$. The $(1 - \alpha)100\%$ CI for ϕ is calculated as $(1 - \alpha)100\%\ \text{CI}(\phi) = \hat{\phi} \pm z_{\alpha/2}[\hat{V}(\hat{\phi}|\phi)]^{1/2}$, where $\hat{V}(\hat{\phi}|\phi)$ is the estimated conditional variance of the sample estimator given the sample.

1. *Weighted Mean (WM)*

The WM estimate $\hat{\phi}_{WM}$ and its variance $\hat{V}(\hat{\phi}_{WM}|\phi)$ are calculated as

$$\hat{\phi}_{WM} = \sum w_d \phi_d \Big/ \sum w_d$$
$$\hat{V}(\hat{\phi}|\phi) = 1 \Big/ \sum w_d$$

where the weight w_d is calculated as in Eq. (6). Since doses with $\lambda_d = 0$ or $p_d = 0$ result in $w_d = 0$ and those with $p_d = 1$ result in indeterminate forms, they are excluded from the calculations.

2. *Maximum Likelihood (ML)*

The ML estimate $\hat{\phi}_{ML}$ is calculated as the value of ϕ that maximizes the log likelihood

$$\ln L = \sum \left[\ln \binom{n_d}{r_d} - r_d \phi \lambda_d + (n_d - r_d) \ln(1 - e^{-\phi \lambda_d}) \right]$$

by Newton's method of iterative approximation

$$\hat{\phi}_{i+1} = \hat{\phi}_i - \frac{(\partial \ln L / \partial \phi)}{(\partial^2 \ln L / \partial \phi^2)}\Big|_{\hat{\phi}_i}$$

where $\hat{\phi}_i$ is the ith iterative $\hat{\phi}_{ML}$ estimate of ϕ and

$$\frac{\partial \ln L}{\partial \phi}\Big|_{\hat{\phi}_i} = \sum \left[-r_d \lambda_d + \frac{(n_d - r_d)\lambda_d e^{-\phi \lambda_d}}{(1 - e^{-\phi \lambda_d})} \right]\Big|_{\hat{\phi}_i}$$

and

$$\frac{\partial^2 \ln L}{\partial \phi^2}\Big|_{\hat{\phi}_i} = \sum \left[\frac{-(n_d - r_d)\lambda_d^2 e^{-\phi \lambda_d}}{(1 - e^{-\phi \lambda_d})^2} \right]\Big|_{\hat{\phi}_i}$$

are, respectively, the first and second partial derivatives of ln L with respect to ϕ evaluated at $\hat{\phi}_i$. $\hat{V}(\hat{\phi}_{ML}|\phi)$ is calculated as

$$\hat{V}(\hat{\phi}_{ML}|\phi) = \frac{-1}{(\partial^2 \ln L / \partial \phi^2)}\Big|_{\hat{\phi}_{ML}}$$

the negative reciprocal of the second derivative evaluated at $\hat{\phi}_{ML}$. All data are included in the calculations.

3. *Minimum Chi-Squared (MC)*

The MC estimate $\hat{\phi}_{MC}$ is calculated as the value of ϕ that minimizes Pearson's chi-squared

$$\chi^2 = \sum \left[\frac{(r_d - n_d e^{-\phi \lambda_d})^2}{n_d e^{-\phi \lambda_d}(1 - e^{-\phi \lambda_d})} \right]$$

by Newton's method of iterative approximation

$$\hat{\phi}_{i+1} = \hat{\phi}_i - \frac{(\partial \chi^2 / \partial \phi)}{(\partial^2 \chi^2 / \partial \phi^2)}\Big|_{\hat{\phi}_i}$$

where $\hat{\phi}_i$ is the ith iterative $\hat{\phi}_{MC}$ estimate of ϕ and

$$\left.\frac{\partial \chi^2}{\partial \phi}\right|_{\hat{\phi}_i} = \sum \left[\frac{n_d \lambda_d e^{-\phi\lambda_d}(2r_d - n_d) + r_d^2 \lambda_d (e^{\phi\lambda_d} - 2)}{n_d(1 - e^{-\phi\lambda_d})^2}\right]\Bigg|_{\hat{\phi}_i}$$

and

$$\left.\frac{\partial^2 \chi^2}{\partial \phi^2}\right|_{\hat{\phi}_i} =$$

$$\sum \left[\frac{n_d \lambda_d^2 (n_d - 2r_d)(e^{\phi\lambda_d} - e^{-3\phi\lambda_d}) + r_d^2 \lambda_d^2 (e^{\phi\lambda_d} - 4 + 7e^{-\phi\lambda_d} - 4e^{-2\phi\lambda_d})}{n_d(1 - e^{-\phi\lambda_d})^4}\right]\Bigg|_{\hat{\phi}_i}$$

are, respectively, the first and second partial derivatives of χ^2 with respect to ϕ evaluated at $\hat{\phi}_i$. $\hat{V}(\hat{\phi}_{MC}|\phi)$ is calculated as

$$\hat{V}(\hat{\phi}_{MC}|\phi) = \left.\frac{2}{(\partial^2 \chi^2/\partial \phi^2)}\right|_{\hat{\phi}_{MC}}$$

twice the reciprocal of the second derivative evaluated at $\hat{\phi}_{MC}$. All data are included in the calculations.

B. Examples

Table III presents frequency estimates and estimate validity test statistics for the spleen and thymus data from Table I. All three of the sample

TABLE III
Parameter Estimates and Estimate Validity Test Statistics for LDA Data from Table I[a]

Assay	Estimator	$\hat{\phi} \times 10^4$	$\hat{V}(\hat{\phi}\|\phi) \times 10^9$	95% CI (ϕ) $\times 10^4$	χ^2_{slope}	P
Spleen	WM	14.47	65.64	9.44–19.49	0.650	0.420
	ML	15.01	66.96	9.93–20.08	0.658	0.417
	MC	14.92	68.64	9.79–20.06	0.657	0.418
Thymus	WM	3.423	2.688	2.406–4.439	0.491	0.484
	ML	3.506	2.739	2.480–4.532	0.498	0.480
	MC	3.511	2.796	2.475–4.547	0.498	0.480

[a] Refer to Section III,B for a discussion of this table; to Section III,A for an explanation of the sample estimators (WM, ML, and MC) used to calculate the sample parameter estimates [$\hat{\phi}$, $\hat{V}(\hat{\phi}|\phi)$, and 95% CI (ϕ)], and to Section II,A,2 for the estimate validity test statistics (χ^2_{slope} and P) for the SHPM.

estimates ($\hat{\phi}_{WM}$, $\hat{\phi}_{ML}$, and $\hat{\phi}_{MC}$) approximate one another for each sample (spleen and thymus). Furthermore, the estimate validity test statistics indicate good fit between the SHPM, the assay data, and each of the three estimates. However, convergence and acceptance (failure to reject by SHPM validity testing) of the three estimates do not always occur as in these examples of "good" assays.

C. Selection of an Estimator

Graphical methods are design unrestricted with limited accuracy and precision, while tabular methods are design restricted with limited applicability. Most of the older computational methods (Cornell and Speckman, 1967) are design restricted. They were developed to simplify calculations before the computer revolution of the past decade eliminated concerns for computational complexity. The newer computational methods (Section III,A; Taswell, 1981) are design unrestricted. These methods ($\hat{\phi}_{MC}$, $\hat{\phi}_{ML}$, $\hat{\phi}_{WM}$) offer the greatest accuracy and precision and the widest applicability. Using Monte Carlo experiments to evaluate their sampling properties, Taswell (1981) ranked these three estimators as well as the least-squares estimator $\hat{\phi}_{LS}$ in increasing order of mean squared error as $\hat{\phi}_{MC} < \hat{\phi}_{ML} < \hat{\phi}_{WM} \ll \hat{\phi}_{LS}$ and advocated use of $\hat{\phi}_{MC}$ as the best estimator with the smallest mean squared error for assay designs with a small number (2–5) of dose-groups and a large number (≥ 24) of replicates. In several recent review articles, these results have been cited incorrectly (Miller, 1982), criticized colorfully (Fazekas de St. Groth, 1982), and declared irrelevant (Lefkovits and Waldmann, 1984).

Miller (1982, p. 222) claimed incorrectly that "Porter and Berry (1963) have worked out a statistically valid procedure for determining frequencies from limiting dilution data" and that "Taswell (1981) ... came to the conclusion that the Porter and Berry procedure was the best method." Porter and Berry (1963) acknowledged Finney (1952) and used a method originally published by him (Finney, 1951). If given an eponym, it could be referred to as Finney's estimator but not "the Porter and Berry procedure." It is a method that solves the ML equation with iterative approximation based on a log–log transformation devised by Mather (1949). Peto's (1953) estimator is a method that solves the ML equation with Newton's iterative approximation based on derivatives. Taswell (1981) investigated Peto's estimator but not Finney's estimator. However, Finney's and Peto's estimators are both ML estimators and can be denoted $\hat{\phi}_{ML}$. If calculated accurately, they should and do produce the same results (Taswell, unpublished results). As stated above, Taswell (1981)

concluded that the MC estimator $\hat{\phi}_{MC}$ is the best method and not $\hat{\phi}_{ML}$ nor "the Porter and Berry procedure."

Using emotionally provocative language when referring to Berkson (1980) and Taswell (1981), Fazekas de St. Groth (1982) criticized Taswell's recommendation for $\hat{\phi}_{MC}$ rather than $\hat{\phi}_{ML}$ as the best estimator. Taswell's preference relied on Monte Carlo experiments and the experimental statistics arguments of Berkson, whereas Fazekas de St. Groth's criticism relied only on the theoretical mathematics arguments of the discussants debating Berkson. In particular, Fazekas de St. Groth (1982, p. R21) extensively quoted Pfanzagl (1980) but repeatedly omitted the word "logit" from the phrase "minimum logit chi-square estimator." Furthermore, he neglected to mention that Berkson's (1980, p. 466) bioassay example and Pfanzagl's discussion of it were devoted to the two-parameter logistic model and not the one-parameter exponential model (which is what the SHPM is). Berkson (1980, p. 458) clearly stated his belief, however, that sometimes the ML estimator and sometimes one of the several different MC estimators is better and therefore that every problem should be investigated individually. Fazekas de St. Groth also failed to discuss Berkson's reply to the criticism of his use of the mean squared error as an estimation criterion for biased estimators. This criticism was the substance of the quotation from Pfanzagl. Berkson's (1980, pp. 462, 486, and 487) reply was to challenge his critics to derive bias-corrected estimators that can be computed and compared by Monte Carlo experiments. Biased and bias-corrected estimators, alternative estimation criteria, and the MC/ML controversy have been discussed further by Amemiya (1980), Rao (1981), Harris and Kanji (1983), and Keating and Mason (1985). Until the results of new Monte Carlo experiments with new estimators (whether biased or bias corrected) and/or with a new estimation criterion (that is an alternative to the mean squared error and that is demonstrated to be relevant) are published for the problem of estimating parameters from LDAs, those reported by Taswell (1981) remain the only results currently available that compare design-unrestricted MC and ML estimators. As far as these results are concerned, investigators may decide for themselves whether they prefer realistic small sample experimental computer simulations or idealized infinitely large sample theoretical mathematical derivations.

In support of their use of $\hat{\phi}_{LS}$, Lefkovits and Waldmann (1984, p. 267) offered their opinion that it does not matter which method is used to estimate the unknown parameter because the parameter estimate represents a roughly approximate assessment of a biological activity and not an absolutely precise determination of a physical constant. Although it is certainly true that physicochemical parameters can be measured more

accurately and precisely than biological parameters, this fact does not justify avoiding measuring biological parameters as accurately and precisely as possible. For LDA parameter estimates, "as accurately and precisely as possible" means maximizing information obtained with limited amounts of time, test subjects and other materials, and labor and other expenses according to some cost–benefit analysis. According to Taswell's (1981) Monte Carlo results, the use of $\hat{\phi}_{MC}$ (or $\hat{\phi}_{ML}$ or even $\hat{\phi}_{WM}$) rather than the use of $\hat{\phi}_{LS}$ should provide the benefit of at least a several-fold increase and as much as an order-of-magnitude increase in accuracy and precision (equivalent to a comparable decrease in materials, labor, etc.) for the cost of at most some additional lines of programming on a calculator or microcomputer. Hopefully, the possibility that LDAs may soon be used as tools in the routine diagnosis and treatment of human disease (Martin and Hansen, 1985) should convince most investigators that any cost–benefit analysis should favor use of the estimator with the greatest accuracy and precision.

Future investigations that explore the relative merits and demerits of various estimators should compare their performance with regard to both point and interval estimates of the parameters, and with regard to the assay design (which may be optimized for either validity test statistics or parameter estimates as explained in Section V). To demonstrate convincingly in a Monte Carlo experiment that an estimator is superior for a given set of assumed parameter values and assay-design constants, it would be necessary to prove that it has a lower coefficient of variation corresponding not only to its point estimates but also to its interval estimates. (A coefficient of variation corresponding to the point estimates can be defined by taking the ratio of the mean squared error for all of the estimates to the assumed value for the parameter. A coefficient of variation corresponding to the interval estimates can be defined by taking the mean of the individual ratios of each estimate's standard error to the estimate's value.) Thus, if for two different estimators the magnitude of the difference between point estimates is small relative to the magnitude of the interval estimates, it would not be possible to claim that one estimator is "truly" superior to the other in any practically meaningful way. However, if both estimators produced similar point estimates and one produced smaller interval estimates than the other, then it would be reasonable to prefer the use of one rather than the other. The ability of estimators to produce correct interval estimates could be tested by incorporating prior distributions for the assumed parameters in the Monte Carlo experiments. Incorporating prior distributions into the simulation would also provide a means of automatically examining an estimator's performance over a range of parameter values.

IV. Population LDAs

Most LDAs are performed serially in time on a sequence of samples drawn from the same population. As discussed in Section I,C, these assays can be interpreted as sample LDAs or as population LDAs and should be analyzed appropriately according to the interpretation. When interpreted as population LDAs, the sample parameters ϕ_a for each assay/sample in the sequence are considered to be the realized (but unknown) values of the random variables Φ_a distributed according to the prior probability density function (pdf) $\pi(\Phi_a = \phi_a|\mu_\Phi, \sigma_\Phi^2)$ with population parameters μ_Φ for the mean and σ_Φ^2 for the variance of the distribution of Φ_a. The realized value ϕ_a of the random variable Φ_a is then considered an unknown fixed parameter ϕ_a that, together with the known assay-design constants λ_{ad} and n_{ad}, determines the observed values r_{ad} of the random variables R_{ad} according to the conditional pdf $f_{ad}(R_{ad} = r_{ad}|\Phi_a = \phi_a)$. Finally, the data set $\{\lambda_{ad}, n_{ad}, r_{ad}|1 \leqq a \leqq A, 1 \leqq d \leqq D_a\}$ is used to calculate validity test statistics and the parameter estimates $\{\phi_a|1 \leqq a \leqq A\}$ for the samples, and $\hat{\mu}_\Phi$ and $\hat{\sigma}_\Phi^2$ for the population.

A. Validity Tests

For sample LDAs, it is necessary to validate a model for the conditional pdf $f_{ad}(r_{ad}|\phi_a)$. For population LDAs, it is necessary to validate both a model for the prior pdf $\pi(\phi_a|\mu_\Phi, \sigma_\Phi^2)$ and a model for the conditional pdf $f_{ad}(r_{ad}|\phi_a)$. Alternatively for population LDAs, it may be possible to validate a multistage model for the marginal pdf $g_{ad}(r_{ad}|\mu_\Phi, \sigma_\Phi^2)$. The conditional pdf f_{ad} has been assumed to be the binomial distribution

$$f_{ad}(r_{ad}|\phi_a) = \binom{n_{ad}}{r_{ad}}\theta_{ad}^{r_{ad}}(1 - \theta_{ad})^{n_{ad}-r_{ad}}$$

with the binomial parameter θ_{ad} set equal to the negative response probability $\theta_{ad} = \exp(-\phi_a\lambda_{ad})$ according to the SHPM ever since it was first proposed by Greenwood and Yule (1917). The prior pdf π was assumed to be the log-normal distribution by Thomas (1955) but has not yet been confirmed by other investigators. Other possibilities for the prior include the normal distribution, the beta distribution

$$\pi(\phi_a|\mu_\Phi, \sigma_\Phi^2) = \frac{\phi_a^{\alpha-1}(1 - \phi_a)^{\beta-1}}{B(\alpha,\beta)} \tag{7}$$

with

$$\alpha = \mu_\Phi[\mu_\Phi(1 - \mu_\Phi)/\sigma_\Phi^2 - 1]$$
$$\beta = (1 - \mu_\Phi)[\mu_\Phi(1 - \mu_\Phi)/\sigma_\Phi^2 - 1]$$

for relative frequency LDAs where $0 \leqq \phi_a \leqq 1$ and $B(\alpha,\beta)$ is the beta function, and the gamma distribution

$$\pi(\phi_a|\mu_\Phi, \sigma_\Phi^2) = \frac{\phi_a^{\gamma-1}e^{-\phi_a/\delta}}{\delta^\gamma\Gamma(\gamma)}$$

with

$$\gamma = \mu_\Phi^2/\sigma_\Phi^2$$
$$\delta = \sigma_\Phi^2/\mu_\Phi$$

for absolute number LDAs where $0 \leqq \phi_a \leqq \infty$ and $\Gamma(\gamma)$ is the gamma function. Possibilities for the marginal pdf g_{ad} include the various combinations of one of the priors with the conditional, for example,

$$g_{ad}(r_{ad}|\mu_\Phi, \sigma_\Phi^2) = \int_0^1 \binom{n_{ad}}{r_{ad}} e^{-r_{ad}\phi_a\lambda_{ad}}(1 - e^{-\phi_a\lambda_{ad}})^{n_{ad}-r_{ad}} \times \frac{\phi_a^{\alpha-1}(1 - \phi_a)^{\beta-1}}{B(\alpha, \beta)}\, d\phi_a \tag{8}$$

an extension of the beta-binomial distribution [which uses a beta distribution, Eq. (7), as the prior]. Validity tests for the models corresponding to the conditional pdf f_{ad} have been developed for sample LDAs (Section II) but not yet for population LDAs (cf. discussion by Taswell, 1984b). Validity tests for the models corresponding to the prior pdf π and the marginal pdf g_{ad} have not yet been developed for population LDAs. Considering each assay as one of a sequence of similar population LDAs rather than as an isolated sample LDA unrelated to the others in the sequence provides a conceptual framework that should permit the development of efficient sequential methods for validity testing.

B. Population Estimators

Porter and Berry (1963) calculated the population estimate $\hat{\mu}_\Phi$ by pooling A individual sample data sets $\{\lambda_{ad}, n_{ad}, r_{ad}|1 \leqq d \leqq D_a\}$ into one combined population data set $\{\lambda_{ad}, n_{ad}, r_{ad}|1 \leqq a \leqq A, 1 \leqq d \leqq D_a\}$ and then analyzing this combined data set with the sample estimator $\hat{\phi}_{ML}$. Taswell *et al.* (1979) calculated $\hat{\mu}_\Phi$ by averaging the set of A individual sample estimates $\{\hat{\phi}_a|1 \leqq a \leqq A\}$ with the unweighted arithmetic mean. Using Monte Carlo experiments, Taswell (1981) compared these two approaches: (1) sample estimator analysis of the pooled data versus (2) averaging the individual sample estimates, for each of the four sample estimators $\hat{\phi}_{MC}$, $\hat{\phi}_{ML}$, $\hat{\phi}_{WM}$, and $\hat{\phi}_{LS}$. He obtained a smaller mean squared error for the first approach relative to the second for each of the four estimators and advocated $\hat{\phi}_{MC}$ analysis of the pooled data as the best

method overall. However, these Monte Carlo experiments actually simulated multiple assays for the same sample rather than multiple samples for the same population. Thus, the Monte Carlo results apply to sample estimates, not population estimates, and justify the assertion that data from multiple assays for the same sample can and should be pooled in order to establish the convention of one-to-one correspondence between assays and samples as explained in Section I,C. Nevertheless, it should be clear that any population estimator consisting of a method that first pools data from different samples and then analyzes the pooled data with a sample estimator is invalid because it fails to account for the biological variation between samples. More precisely, the invalidity of this approach derives from the facts that (1) sample estimators are based on the SHPM, (2) the SHPM applies only within each sample and not between different samples, and therefore (3) sample estimators are valid only for individual sample data and not for combined population data.

The terminology and notation of Section IV,A further clarify the distinction between sample estimators and population estimators. Sample estimators for the sample estimand ϕ from sample LDAs are based only on the conditional pdf f and do not use any information from other assays/samples. Population estimators for the sample estimands $\{\phi_a | 1 \leqq a \leqq A\}$ from population LDAs are based on both the prior pdf π and conditional together and do use information from other assays/samples in the sequence. Population estimators for the population estimands μ_Φ and σ^2_Φ are based either on the marginal pdf g or on both the prior and conditional together. Those based on the marginal calculate only population estimates $\hat{\mu}_\Phi$ and $\hat{\sigma}^2_\Phi$, whereas those based on the prior and conditional calculate both sample estimates $\{\hat{\phi}_a | 1 \leqq a \leqq A\}$ and population estimates $\hat{\mu}_\Phi$ and $\hat{\sigma}^2_\Phi$. Population estimators based on the marginal require specification of the prior in order to formulate the marginal as, for example, in Eq. (8) where the prior has been specified as a beta distribution [Eq. (7)]. An example of a population estimator based on this marginal is the solution for α and β that maximizes the log likelihood equation

$$\ln L(\alpha, \beta) = \sum_{a=1}^{A} \ln \int_0^1 \prod_{d=1}^{D_a} \left[\binom{n_{ad}}{r_{ad}} e^{-r_{ad}\phi_a\lambda_{ad}} (1 - e^{-\phi_a\lambda_{ad}})^{n_{ad}-r_{ad}} \right] \times \frac{\phi_a^{\alpha-1}(1-\phi_a)^{\beta-1}}{B(\alpha, \beta)} \, d\phi_a$$

with the estimates $\hat{\alpha}$ and $\hat{\beta}$ then used to calculate the population estimates

$$\hat{\mu}_\Phi = \hat{\alpha}/(\hat{\alpha} + \hat{\beta})$$

$$\hat{\sigma}^2_\Phi = \hat{\alpha}\hat{\beta}/[(\hat{\alpha} + \hat{\beta})^2(\hat{\alpha} + \hat{\beta} + 1)]$$

Population estimators based on the prior and conditional may or may not require specification of the prior. Examples include parametric empirical Bayes estimators (Morris, 1983) that specify a prior, nonparametric empirical Bayes estimators (Robbins, 1983) that do not specify a prior, and nonparametric maximum likelihood estimators (Laird, 1978) that combine aspects of both approaches.

Population estimators based on an unspecified prior (together with the conditional) possess the clear advantage that no assumptions need be made about the nature or form of the prior distribution. This approach may be the most appropriate at a time when insufficient data have been accumulated and tested for selection of one of several alternative prior distributions such as the normal, log-normal, beta, and gamma distributions, etc. The simplest class of population estimators based on an unspecified prior is the class of moments estimators of which the simplest is the unweighted arithmetic mean. Although theoretically superior (see above) to the pooling used by Porter and Berry (1963), there are several major problems with the unweighted averaging used by Taswell *et al.* (1979). First, the calculated mean is not weighted, so it is inefficient. Second, the calculated variance of the mean is the estimate $\hat{V}(\hat{\mu}_\Phi)$ and not the estimate $\hat{\sigma}^2_\Phi$. To characterize a population, the pair $\{\hat{\mu}_\Phi, \hat{\sigma}^2_\Phi\}$ is required and not the pair $\{\hat{\mu}_\Phi, \hat{V}(\hat{\mu}_\Phi)\}$. Ideally, the quadruple $\{\hat{\mu}_\Phi, \hat{V}(\hat{\mu}_\Phi), \hat{\sigma}^2_\Phi, \hat{V}(\hat{\sigma}^2_\Phi)\}$ would be desired. Third, the calculated population estimates are obtained from sample estimates in a way that does not adequately incorporate the concept of the prior distribution. Thomas (1955) recognized that

$$V(\hat{\phi}_a) = V(\hat{\phi}_a|\phi_a) + V(\phi_a) = V(\hat{\phi}_a|\phi_a) + \sigma^2_\Phi \tag{9}$$

the variance $V(\hat{\phi}_a)$ of the sample estimator equals the sum of the statistical variance $V(\hat{\phi}_a|\phi_a)$ of the sample estimator given the sample and the biological variance $V(\phi_a) = \sigma^2_\Phi$ of the samples within the population. Every population estimator must incorporate the concept represented by Eq. (9) in some way in order to be theoretically valid. These three problems invalidate use of the unweighted arithmetic mean as a population estimator.

A population estimator based on an unspecified prior and incorporating Eq. (9) is the estimator that calculates the population estimates by the method of weighted moments from the sample estimates according to the following algorithm:

1. Set $\hat{\sigma}^2_\Phi = 0$
2. Set $w_a = 1/[\hat{V}(\hat{\phi}_a|\phi_a) + \hat{\sigma}^2_\Phi]$
3. Calculate $\hat{\mu}_\Phi = \Sigma w_a \hat{\phi}_a / \Sigma w_a$ and $\hat{V}(\hat{\mu}_\Phi) = 1/\Sigma w_a$

4. Calculate $\hat{\sigma}^2_\Phi = \max\{0, \hat{V}(\hat{\mu}_\Phi) + \Sigma[(\hat{\phi}_a - \hat{\mu}_\Phi)^2 - \hat{V}(\hat{\phi}_a|\phi_a)]/A\}$
5. Repeat steps 2–4 until each estimate converges
6. Report the triple $\{\hat{\mu}_\Phi, \hat{V}(\hat{\mu}_\Phi), \hat{\sigma}^2_\Phi\}$

where $\hat{\phi}_a$ and $\hat{V}(\hat{\phi}_a|\phi_a)$ are sample estimates such as $\hat{\phi}_{MC}$ and $\hat{V}(\hat{\phi}_{MC}|\phi)$ as calculated in Section III,A. Use of this method is demonstrated with examples in Section VI. This method could be enhanced further by simultaneously calculating the A sample estimates $\hat{\phi}_a$ with nonparametric empirical Bayes methods although this refinement would require retaining all of the data $\{\lambda_{ad}, n_{ad}, r_{ad} | 1 \leqq a \leqq A, 1 \leqq d \leqq D_a\}$ rather than just the sample estimates $\{\hat{\phi}_a, \hat{V}(\hat{\phi}_a|\phi_a) | 1 \leqq a \leqq A\}$. More sophisticated methods integrating assay design should also be developed. As $a \to \infty$ and information accumulates about μ_Φ and σ^2_Φ, current estimates of these population parameters based on the previous $a - 1$ assays should be used to estimate each new sample parameter ϕ_a by empirical Bayes and other methods as well as to improve the efficiency of the assay design for the estimation of both the new sample parameter ϕ_a and the new estimates of the population parameters μ_Φ and σ^2_Φ based on a rather than $a - 1$ assays (Section V).

V. Assay Design

Analysis of data from both sample and population LDAs produces validity test statistics and parameter estimates as results. Design optimization criteria should be used in a manner that permits the choice of a design most appropriate for the result desired. Multiple-dose assay designs optimal for validity test statistics are not optimal for parameter estimates. Single-dose assay designs optimal for parameter estimates are not optimal, in fact not even useful at all, for validity test statistics. Thus, an additional criterion will be necessary to shift the optimal design in a sequential analysis of population LDAs from one emphasizing model discrimination to one emphasizing parameter estimation (Hill *et al.*, 1968; Borth, 1975).

A. Validity Tests

As discussed in Section IV,A, validity tests are currently available only for the conditional pdf f in sample LDAs. Design optimization has not yet been investigated but could be patterned after that done by Chambers and Cox (1967). In general, however, these validity tests (especially those in Section II,A) increase in efficiency and power as the design constants D and $\{n_d | 1 \leqq d \leqq D\}$ increase in size. Validity tests are not yet available for the prior pdf π together with the conditional pdf f or for the marginal pdf g in population LDAs. Since these validity tests could be developed as sequential hypothesis tests, design optimization would be an integral part of the sequential analysis.

B. Sample and Population Estimators

Design optimization has been investigated extensively for sample estimators and recently for population estimators. An optimal dose was recommended for the first time by Stein (1919) and calculated by Fisher (1928) to be $\omega^* = 1.6$ corresponding to $\theta^* = 0.20$ where ω^* is the optimal number of BAPs and θ^* is the optimal negative response probability for the single-dose sample assay. Subsequently, what became known as the Fisher information I was plotted as a function of ω by Fisher (1935) and as a function of θ by Bartlett (1935). Since then, Fisher's results have been obtained by more than a dozen other authors, including Fazekas de St. Groth (1982), who plotted I [equal to w in Eq. (5) with p replaced by $\theta = \exp(-\omega)$] as a function of both ω and θ. Single-dose sample LDAs at Fisher's optimal dose may be the most efficient for estimating the sample parameter ϕ, but they are not practical because they require prior knowledge of the parameter ϕ in order to design the assay so that $\lambda = 1.6/\phi$. Obviously, though, if ϕ is known in advance, there is no need to perform the assay.

This logical contradiction does not occur for population LDAs where prior knowledge, available from previous assays in the sequence, can be used to estimate μ_Φ and σ^2_Φ, approximate the prior distribution of Φ_a, and design the next assay in the sequence so that the probability of $\lambda_a = \omega^*_a/\phi_a$ is maximized. This task can be accomplished by choosing λ_a equal to the current estimate $\hat{\mu}_\Lambda = \mu^*_\Omega/\hat{\mu}_\Phi$ where $\hat{\mu}_\Phi$ is the current estimate of μ_Φ (calculated from the previous $a - 1$ assays) and μ^*_Ω is the optimal number of BAPs for the single-dose population assays. μ^*_Ω depends on both μ_Φ and σ^2_Φ (or $CV_\Phi = \sigma_\Phi/\mu_\Phi$), which are approximated with the current estimates $\hat{\mu}_\Phi$ and $\hat{\sigma}^2_\Phi$. μ^*_Ω is calculated according to an asymptotic analysis and tabulated with corresponding values of μ^*_Θ and μ^*_Ψ, respectively, the negative response and noninformative assay probabilities for the single-dose population assays (Taswell, 1984a). The noninformative assay probability,

$$\psi_a = \prod_{d=1}^{D_a} e^{-n_{ad}\phi_a\lambda_{ad}} + \prod_{d=1}^{D_a} (1 - e^{-\phi_a\lambda_{ad}})^{n_{ad}}$$

which reduces to

$$\psi_a = e^{-n_a\phi_a\lambda_a} + (1 - e^{-\phi_a\lambda_a})^{n_a}$$

for single-dose assays, is defined as the probability that the test subjects in the assay respond all positively or all negatively. Once λ_a has been chosen, n_a must be chosen sufficiently large in size so that ψ_a attains a satisfactorily low level in probability in order to avoid assays that are "noninformative" with $r_a = 0$ or $r_a = n_a$.

For $\mu_\Phi \leqq 0.1$, Taswell (1984a) obtained increasing values for μ_Ω^* within the range $1.6 \leqq \mu_\Omega^* \leqq 2.6$ with corresponding decreasing values for μ_Θ^* within the range $0.20 \geqq \mu_\Theta^* \geqq 0.14$ as a function of increasing values of CV_Φ within the range $0 \leqq CV_\Phi \leqq 0.5$. For $CV_\Phi = 0.5$, at $\mu_\Omega = 1.8$ ($\mu_\Theta = 0.23$) with 95% and at $\mu_\Omega = 1.2$ ($\mu_\Theta = 0.35$) with 80% of the efficiency of $\mu_\Omega^* = 2.6$ ($\mu_\Theta^* = 0.14$), he obtained decreasing values of μ_Ψ evaluated at μ_Ω or $\mu_\Psi(\mu_\Omega)$ within the ranges $0.04 \leqq \mu_\Psi(\mu_\Omega = 1.8) \leqq 0.003$ and $0.005 \leqq \mu_\Psi(\mu_\Omega = 1.2) \leqq 0.00004$, respectively, as a function of increasing values of n_a within the range $40 \leqq n_a \leqq 200$. Since $CV_\Phi = 0.5$ possibly represents a variance typical of some stable biological populations (Section VI,B), these results suggest that single-dose population assays could be performed when sufficient prior information is available and model discrimination is no longer an issue. This analysis assumed a beta distribution [Eq. (7)] as the prior pdf π and used minimization of the Cramer–Rao minimum variance σ_{CR}^2 (equivalent to maximization of the Fisher information I) to find the optimal dose. Future analyses should explore other prior distributions and other design optimization criteria such as extensions of I and σ_{CR}^2 (Gart, 1959; Ferreira, 1981). However, whether applied to single- or multiple-dose assay designs, the goal of optimization methods should be to minimize the error of the estimators subject to the constraint of a chosen fixed level of the noninformative assay probability.

VI. Comparative Experiments

LDAs can be used to compare natural populations, experimentally treated populations, and the effectiveness of purification and depletion procedures (Taswell *et al.*, 1979). These comparisons require estimation of the differences between and/or ratios of the BAP frequencies of the populations being compared. Simple methods are described in Section VI,A for calculating confidence intervals and test statistics for these differences and ratios. More complicated methods based on sequential analysis should be developed in the future. Sequential hypothesis testing will enable the investigator to terminate the experiment at the earliest possible time (smallest number of sequential assays) for a given level of significance. How to interpret hypothesis tests was discussed recently by Salsburg (1985), who satirized the abuse of P values in an amusing article entitled "The Religion of Statistics as Practiced in Medical Journals." He suggested a reform from sole use of P values to greater use of multiple confidence intervals (in particular, a set of three: the 50% CI, 80% CI, and 99% CI) with decisions based on the CI appropriate for the risks entailed as a consequence of each decision.

A. Methods

When comparing populations, it is necessary to decide whether the samples from the populations are independent or paired. In most experiments, there are two independent sequences $\{\phi_a | 1 \leqq a \leqq A\}$ and $\{\phi_b | 1 \leqq b \leqq B\}$ of samples from two different populations A and B. However, if there is a single sequence $\{\phi_a | 1 \leqq a \leqq A\}$ from the population A where each sample ϕ_a is divided in half and the two halves $\phi_{a,1}$ and $\phi_{a,2}$ are subjected to two different treatments (including "control and experimental" and "before and after"), then the sample halves should be considered paired. Paired (half) samples $\{\phi_{a,1},\phi_{a,2} | 1 \leqq a \leqq A\}$ permit estimation of the sample differences D_a and ratios M_a and the population difference μ_D and ratio μ_M. Independent samples $\{\phi_a,\phi_b | 1 \leqq a \leqq A, 1 \leqq b \leqq B\}$ permit estimation of the population difference $\mu_{\Phi,A-B}$ and ratio $\mu_{\Phi,A/B}$. Confidence intervals for these parameters can then be compared and P values for test statistics can be calculated using the algorithms reviewed by Maindonald (1984).

1. *Confidence Intervals and Tests for Differences*

a. Paired Samples. Let $\phi_{a,1}$ and $\phi_{a,2}$ ($1 \leqq a \leqq A$) denote the paired halves of the ath sample subjected to the two different treatments. Calculate the sample estimates $\hat{\phi}_{a,1}$, $\hat{V}(\hat{\phi}_{a,1}|\phi_{a,1})$, $\hat{\phi}_{a,2}$, and $\hat{V}(\hat{\phi}_{a,2}|\phi_{a,2})$ as in Section III,A. Estimate the sample differences D_a and population difference μ_D and their variances as

$$\hat{D}_a = \hat{\phi}_{a,1} - \hat{\phi}_{a,2}$$

$$\hat{V}(\hat{D}_a|\phi_a) = \hat{V}(\hat{\phi}_{a,1}|\phi_{a,1}) + \hat{V}(\hat{\phi}_{a,2}|\phi_{a,2})$$

$$w_a = 1/\hat{V}(\hat{D}_a|\phi_a)$$

$$\hat{\mu}_D = \sum w_a \hat{D}_a \Big/ \sum w_a$$

$$\hat{V}(\hat{\mu}_D) = 1 \Big/ \sum w_a$$

Then calculate a CI and test statistic as

$$(1 - \alpha)100\%\ \mathrm{CI}(\mu_D) - \hat{\mu}_D \pm z_{\alpha/2}[\hat{V}(\hat{\mu}_D)]^{1/2}$$

$$z = \hat{\mu}_D/[\hat{V}(\hat{\mu}_D)]^{1/2}$$

where z observes the standard normal distribution in a two-tailed test for the null hypothesis that $\mu_D = 0$ (or $\phi_{a,1} = \phi_{a,2}$).

b. Independent Samples. Let $\{\phi_a | 1 \leqq a \leqq A\}$ and $\{\phi_b | 1 \leqq b \leqq B\}$ denote independent samples from two different populations A and B. Calculate the population estimates $\hat{\mu}_{\Phi,A}$, $\hat{V}(\hat{\mu}_{\Phi,A})$, $\hat{\mu}_{\Phi,B}$, and $\hat{V}(\hat{\mu}_{\Phi,B})$ as in Section IV,B by the method of weighted moments. Then estimate $\mu_{\Phi,A-B}$ and its variance and calculate a CI and test statistic as

$$\hat{\mu}_{\Phi,A-B} = \hat{\mu}_{\Phi,A} - \hat{\mu}_{\Phi,B}$$

$$\hat{V}(\hat{\mu}_{\Phi A-B}) = \hat{V}(\hat{\mu}_{\Phi,A}) + \hat{V}(\hat{\mu}_{\Phi,B})$$

$$(1 - \alpha)100\%\ \mathrm{CI}(\mu_{\Phi,A-B}) = \hat{\mu}_{\Phi,A-B} \pm z_{\alpha/2}[\hat{V}(\hat{\mu}_{\Phi,A-B})]^{1/2}$$

$$z = \hat{\mu}_{\Phi,A-B}/[\hat{V}(\hat{\mu}_{\Phi,A-B})]^{1/2}$$

where z observes the standard normal distribution in a two-tailed test for the null hypothesis that $\mu_{\phi,A-B} = 0$ (or $\mu_{\phi,A} = \mu_{\phi,B}$).

2. *Confidence Intervals and Tests for Ratios*

a. Paired Samples. Estimate the sample ratios M_a and population ratio μ_M and their variances as

$$\hat{M}_a = \hat{\phi}_{a,1}/\hat{\phi}_{a,2}$$

$$\ln \hat{M}_a = \ln \hat{\phi}_{a,1} - \ln \hat{\phi}_{a,2}$$

$$\hat{V}[\ln(\hat{M}_a)|\phi_a] = \hat{V}(\hat{\phi}_{a,1}|\phi_a)/\hat{\phi}^2_{a,1} + \hat{V}(\hat{\phi}_{a,2}|\phi_a)/\hat{\phi}^2_{a,2}$$

$$w_a = 1/\hat{V}[\ln(\hat{M}_a)|\phi_a]$$

$$\hat{\mu}_{\ln M} = \sum w_a(\ln \hat{M}_a)/\sum w_a$$

$$\hat{V}(\hat{\mu}_{\ln M}) = 1/\sum w_a$$

$$\hat{\mu}_M = \exp(\hat{\mu}_{\ln M})$$

Then calculate a CI and test statistic as

$$(1 - \alpha)100\%\ \mathrm{CI}(\mu_M) = \exp\{\hat{\mu}_{\ln M} \pm z_{\alpha/2}[\hat{V}(\hat{\mu}_{\ln M})]^{1/2}\}$$

$$z = \hat{\mu}_{\ln M}/[\hat{V}(\hat{\mu}_{\ln M})]^{1/2}$$

where z observes the standard normal distribution in a two-tailed test for the null hypothesis that $\mu_M = 1$ (or $\phi_{a,1} = \phi_{a,2}$).

b. Independent Samples. Estimate $\mu_{\Phi,A/B}$ and its variance as

$$\hat{\mu}_{\Phi,A/B} = \hat{\mu}_{\Phi,A}/\hat{\mu}_{\Phi,B}$$

$$\ln \hat{\mu}_{\Phi,A/B} = \ln \hat{\mu}_{\Phi,A} - \ln \hat{\mu}_{\Phi,B}$$

$$\hat{V}(\ln \hat{\mu}_{\Phi,A/B}) = \hat{V}(\hat{\mu}_{\Phi,A})/\hat{\mu}^2_{\Phi,A} + \hat{V}(\hat{\mu}_{\Phi,B})/\hat{\mu}^2_{\Phi,B}$$

Then calculate a CI and test statistic as

$$(1 - \alpha)100\%\ \mathrm{CI}(\hat{\mu}_{\Phi,A/B}) = \exp\{\ln \hat{\mu}_{\Phi,A/B} \pm z_{\alpha/2}[\hat{V}(\ln \hat{\mu}_{\Phi,A/B})]^{1/2}\}$$

$$z = \ln(\hat{\mu}_{\Phi,A/B})/[\hat{V}(\ln \hat{\mu}_{\Phi,A/B})]^{1/2}$$

where z observes the standard normal distribution in a two-tailed test for the null hypothesis that $\mu_{\Phi,A/B} = 1$ (or $\mu_{\Phi,A} = \mu_{\Phi,B}$).

B. Examples

Table IV presents parameter estimates from population LDAs identical to the sample LDAs for normal spleen and thymus test preparations described in Sections II,B and III,B. The purified spleen test preparations consisted of normal spleen test preparations that were passed over nylon wool columns. Population estimates were calculated by the method of weighted moments as in Section IV,B from sample estimates calculated by the method of minimum chi-squared as in Section III,A,3. Results from these assays were previously published by Taswell *et al.* (1979). However, estimates of $\hat{\sigma}^2_\Phi$ and $\mathrm{C\hat{V}}_\Phi = \hat{\sigma}_\Phi/\hat{\mu}_\Phi$ were not calculated in their analysis of the data. In these examples $\hat{\sigma}^2_\Phi$ was always greater than $\hat{V}(\hat{\mu}_\Phi)$, and $\mathrm{C\hat{V}}_\Phi$ ranged from 0.5 to 1.0, suggesting that there is large variation between samples of each population. Tables V and VI compare two populations by examining their differences and ratios. Samples were prepared and split in half with one half subjected to experimental treatment (a purification procedure) and the other half not. Thus, the half samples

TABLE IV
PARAMETER ESTIMATES FROM POPULATION LDAS FOR THE DETERMINATION OF MURINE CTL-P FREQUENCIES[a]

Assay sequence	A	Range $(\hat{\phi}_a) \times 10^3$	$\hat{\mu}_\Phi \times 10^3$	$\hat{V}(\hat{\mu}_\Phi) \times 10^7$	$\hat{\sigma}^2_\Phi \times 10^7$	$\mathrm{C\hat{V}}_\Phi$
Normal spleen	12	0.2–6.3	2.2	2.1	22.8	0.70
Purified spleen	5	0.8–15.6	6.2	77.7	360.8	0.98
Normal thymus	5	0.4–1.0	0.7	0.2	1.0	0.46

[a] Refer to Section VI,B for a discussion of this table, to Section IV,B for an explanation of the population estimators (weighted moments) used to calculate the population parameter estimates [$\hat{\mu}_\Phi$, $\hat{V}(\hat{\mu}_\Phi)$, $\hat{\sigma}^2_\Phi$, and $\mathrm{C\hat{V}}_\Phi$], and to Section I,C for basic terms and notation including the number A of assays/samples in the sequence and the estimate $\hat{\phi}_a$ of the ath sample parameter ϕ_a in the sequence. Note the distinction between σ^2_Φ and $V(\mu_\Phi)$. σ^2_Φ is the variance of the samples whereas $V(\mu_\Phi)$ is the variance of the mean of the samples.

TABLE V
SAMPLE ESTIMATES FOR MURINE CTL-P FREQUENCIES FROM SAMPLES ANALYZED AS BOTH PAIRED AND INDEPENDENT SEQUENCES FOR THE COMPARISON OF POPULATION ESTIMATES IN TABLE VI[a]

Samples		Purified spleen		Normal spleen	
Paired	a	$\hat{\phi}_{a,1}$	$\hat{V}(\hat{\phi}_{a,1}\|\phi_{a,1})$	$\hat{\phi}_{a,2}$	$\hat{V}(\hat{\phi}_{a,2}\|\phi_{a,2})$
	1	1.564×10^{-2}	1.713×10^{-5}	6.282×10^{-3}	1.482×10^{-6}
	2	1.867×10^{-3}	1.759×10^{-7}	7.963×10^{-4}	3.475×10^{-8}
	3	4.725×10^{-3}	5.824×10^{-7}	2.980×10^{-3}	2.407×10^{-7}
	4	8.447×10^{-4}	3.121×10^{-8}	2.280×10^{-4}	6.791×10^{-9}
Independent	a or b	$\hat{\phi}_a$	$\hat{V}(\hat{\phi}_a\|\phi_a)$	$\hat{\phi}_b$	$\hat{V}(\hat{\phi}_b\|\phi_b)$

[a] Refer to Section VI,B for a discussion of this table and to Section VI,A for an explanation of terms and notation.

were true paired samples. However, they were analyzed as both paired and independent samples for the purpose of demonstrating the methods of analysis and the importance of distinguishing between paired and independent samples. Table V presents the sample estimates and Table VI the population estimates with test statistics. Improper analysis of the paired samples as independent samples wastes the information gathered by the experiment. Proper analysis of the paired samples results in dramatic rejection of the null hypothesis that the two populations are equivalent.

TABLE VI
POPULATION ESTIMATES FOR DIFFERENCES AND RATIOS OF MURINE CTL-P FREQUENCIES FROM TWO POPULATIONS WITH THE SAMPLE ESTIMATES IN TABLE V[a]

	Differences		Ratios	
	Paired samples ($\hat{\mu}_D$)	Independent samples ($\hat{\mu}_{\Phi,A\ B}$)	Paired samples ($\hat{\mu}_M$)	Independent samples ($\hat{\mu}_{\Phi,A/B}$)
$\hat{\mu}$	0.740×10^{-3}	2.601×10^{-3}	2.129	2.083
95% CI (μ)	$0.4–1.1 \times 10^{-3}$	$-4.5–9.7 \times 10^{-3}$	1.6–2.9	0.4–11.8
z statistic	4.209	0.715	4.974	0.831
P value	2.6×10^{-5}	0.47	8.3×10^{-7}	0.41

[a] Refer to Section VI,B for a discussion of this table and to Section VI,A for an explanation of terms and notation.

VII. Clonal Analysis

LDAs can also be used to separate, characterize, and quantitate subpopulations of BAPs with different activities by clonal analysis (Taswell *et al.*, 1980). Throughout this article, the term population has referred to a collection of samples where each sample is an independent test preparation derived from one of a sequence of identical biological sources. For example, a sample may correspond to one animal and a population to a sequence of such animals of the same species, strain, age, sex, etc. In this section and in Section VIII, however, the term population refers to the collection of BAPs (e.g., cells) in the test preparation. Albeit confusing, both usages are nevertheless consistent with common parlance. Thus, given a test preparation consisting of a population of BAPs with multiple activities, the population may be (1) homogeneous with each BAP possessing multiple activities, (2) heterogeneous with each BAP possessing a single activity, or (3) a complex combination of homogeneous and heterogeneous subpopulations.

In a typical clonal analysis, POPs are administered to test cultures at dose levels sufficiently low to ensure low BAP dose levels and high monoclonality probability. After the expansion of precursors into clones, individual cultures are split into two or more fractions, tested for the presence of two or more different activities, and assigned the appropriate culture activity phenotype. In the case of a BAP population tested for two different activities A and B, there are four possible culture activity phenotypes with both activities (A^+B^+, A^+B^-, A^-B^+, and A^-B^-) and four with one activity without regard to the other (A^+, A^-, B^+, and B^-). Phenotype distributions and BAP frequencies are then estimated. Test statistics for the independent association of phenotypes and monoclonality probabilities for each culture are then estimated based on the phenotype distributions and BAP frequencies, respectively. Statistical clonal analysis can be confirmed by biological clonal analysis wherein cultures presumed to be true clones (*i.e.*, monoclonal) are recultured or subcloned (Taswell *et al.*, 1980). Indeed, many LDA culture systems permit propagation and repeated subcloning of the original clones.

A. METHODS

Clonal analysis requires careful examination of cultures with the double-positive activity phenotype A^+B^+. The standard χ^2 test (Section VIII,A,1) can be performed on the observed phenotype distribution for the independent association of phenotypes in a 2×2 contingency table. Fisher's exact probability test for this problem was discussed by

Lefkovits and Waldmann (1979). The monoclonality probability κ that a positive response in a test culture is generated by a single BAP as the precursor of a true clone (Section VII,A,2) can be calculated as a function of the estimated BAP frequency $\hat{\phi}$ and the known POP dose level λ (Taswell *et al.*, 1980). Several theoretical plots of κ have been published. Lefkovits and Waldmann (1979) plotted κ as a function of the unknown BAP dose-level ω (which can be estimated by $\hat{\omega} = \hat{\phi}\lambda$) and Miller (1982) plotted κ as a function of the positive response probability $1 - \theta$ [which can be estimated by $1 - p = 1 - r/n$ or by $1 - \hat{p} = 1 - \exp(-\hat{\phi}\lambda)$].

1. *Chi-Squared Test for the Independent Association of Phenotypes*

Code the eight phenotypes A^+B^+, A^+B^-, A^-B^+, A^-B^-, A^+, A^-, B^+, and B^- by the indices ij with values 0 and 1 as 11, 10, 01, 00, 1., 0., .1, and .0. Define s_{ij} as the observed numbers of cultures with the phenotypes coded by the subscripts ij from a group of n cultures all at the same POP dose-level λ. Calculate the expected numbers $\hat{s}_{ij} = s_{i.}s_{.j}/n$ and the test statistic

$$\chi^2 = \sum_{i=0}^{1}\sum_{j=0}^{1} (s_{ij} - \hat{s}_{ij})^2/\hat{s}_{ij}$$

where χ^2 observes the χ^2 distribution with 1 df in an upper-tailed test for the null hypothesis that the phenotypes A^+, A^-, B^+, and B^- are associated independently.

2. *Monoclonality Probability for Positive Cultures*

Using the phenotype codes as explained in Section VII,A,1, define $s_{d,ij}$ as the observed numbers of cultures with the phenotypes coded by the subscripts ij from a dose group of n_d cultures at the POP dose-level λ_d. Define $r_{d,ij} = n_d - s_{d,ij}$ as the observed numbers of cultures that are "negative" relative to the coded phenotype. Define $\hat{\phi}_{ij}$ as the estimate of the ij-coded phenotypic BAP frequency calculated from the data set $\{\lambda_d, n_d, r_{d,ij} | 1 \le d \le D\}$ by a sample estimator as explained in Section II,A. Calculate

$$\hat{\kappa}_{d,ij} = \frac{\hat{\phi}_{ij}\lambda_d \exp(-\hat{\phi}_{ij}\lambda_d)}{1 - \exp(-\hat{\phi}_{ij}\lambda_d)} = \frac{\hat{\phi}_{ij}\lambda_d}{\exp(\hat{\phi}_{ij}\lambda_d) - 1}$$

as the monoclonality probability estimate based on the ij-coded phenotypic BAP frequency estimate $\hat{\phi}_{ij}$ for positive cultures generated at the POP dose-level λ_d.

B. Examples

Figure 1 displays the CTL anti-P815 and anti-AKRA activities of the individual test cultures from a single dose group of an LDA for the clonal analysis of a primary C57BL/6 anti-DBA/2 murine mixed leukocyte culture (MLC) cell population. Table VII presents data and parameter estimates from another LDA for an identical population (identified as population I in Table II of Taswell *et al.*, 1980). From this assay, the phenotype distribution $\{s_{1,11} = 1, s_{1,10} = 10, s_{1,01} = 3, s_{1,100} = 58\}$ for dose-group $d = 1$ with $\lambda_1 = 1$ results in $\chi^2 = 0.309$ with $P = 0.578$ thus failing to reject the

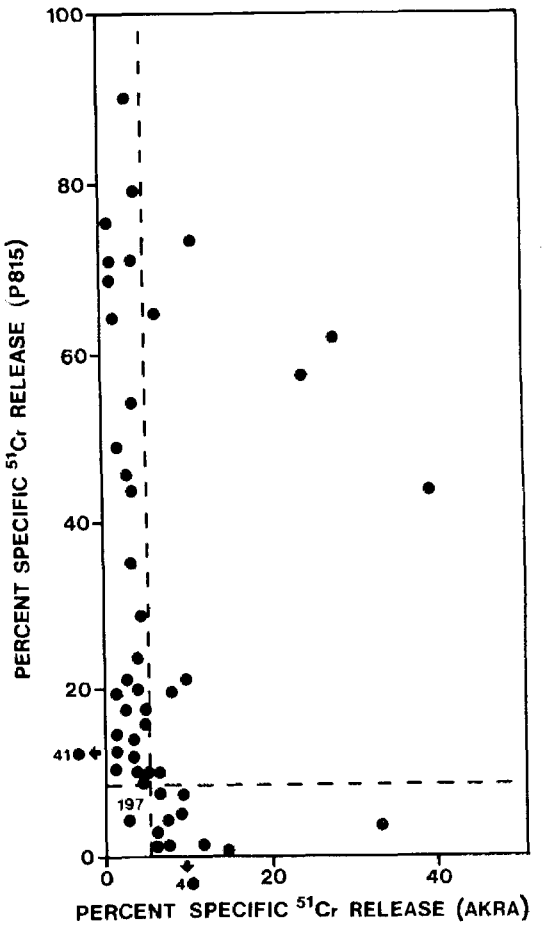

FIG. 1. CTL anti-P815 and anti-AKRA activities (phenotypes A and B, respectively) for individual test cultures from a single dose group of an LDA for the clonal analysis of a primary C57BL/6 anti-DBA/2 murine MLC cell population. Dotted lines indicate the minimum positive activity level: right upper quadrant, A^+B^+; left upper quadrant, A^+B^-; right lower quadrant, A^-B^+; left lower quadrant, A^-B^-. Reproduced from Fig. 2 of Taswell *et al.* (1980), *The Journal of Experimental Medicine* **151,** pp. 1372–1385, by copyright permission of The Rockefeller University Press. For discussion of this figure, refer to Section VII,B.

TABLE VII

DATA AND PARAMETER ESTIMATES FROM AN LDA FOR THE CLONAL ANALYSIS OF A MURINE MLC CELL POPULATION[a]

			$A^+(ij = 1.)$			$B^+(ij = .1)$			$A^+B^+(ij = 11)$			$A^+B^-(ij = 10)$			$A^-B^+(ij = 01)$			A^-B $(ij = 00)$
d	λ_d	n_d	$s_{d,1.}$	$r_{d,1.}$	$\hat{\kappa}_{d,1.}$	$s_{d,.1}$	$r_{d,.1}$	$\hat{\kappa}_{d,.1}$	$s_{d,11}$	$r_{d,11}$	$\hat{\kappa}_{d,11}$	$s_{d,10}$	$r_{d,10}$	$\hat{\kappa}_{d,10}$	$s_{d,01}$	$r_{d,01}$	$\hat{\kappa}_{d,01}$	$s_{d,00}$
1	1	72	11	61	0.940	4	68	0.982	1	71	0.985	10	62	0.959	3	69	0.979	58
2	3	24	10	14	0.827	2	22	0.946	2	22	0.956	8	16	0.881	0	24	0.938	14
3	6	24	8	16	0.676	1	23	0.894	1	23	0.914	7	17	0.772	0	24	0.878	16
4	9	24	16	8	0.546	8	16	0.844	8	16	0.872	8	16	0.673	0	24	0.821	8
			$\hat{\phi}_{1.} = 1.230 \times 10^{-1}$			$\hat{\phi}_{.1} = 3.658 \times 10^{-2}$			$\hat{\phi}_{11} = 2.966 \times 10^{-2}$			$\hat{\phi}_{10} = 8.290 \times 10^{-2}$			$\hat{\phi}_{01} = 4.256 \times 10^{-2}$			

[a] Refer to Section VII,B for a discussion of this table and to Section VII,A for an explanation of terms and notation. $s_{d,ij}$, Number of cultures that have the phenotype coded by the subscripts ij; $r_{d,ij}$, number of cultures that are "negative" relative to the coded phenotype; $\hat{\phi}_{ij}$, estimate of the ij-coded phenotypic BAP frequency; $\hat{\kappa}_{d,ij}$, monoclonality probability for positive cultures based on $\hat{\phi}_{ij}$.

null hypothesis that the phenotypes associate independently. However, the χ^2 test is not trustworthy when the cell frequencies in the 2×2 contingency table are less than ~5 as in this case. Nevertheless, the monoclonality probability estimates $\hat{\kappa}_{1,ij}$ equal or exceed 0.94 for all relevant ij. Therefore, it is likely that these positive cultures were true clones. This statistical clonal analysis was confirmed by biological clonal analysis, as shown in Fig. 2, which demonstrates clearly that subclones retained the same activity phenotypes as their parent clones.

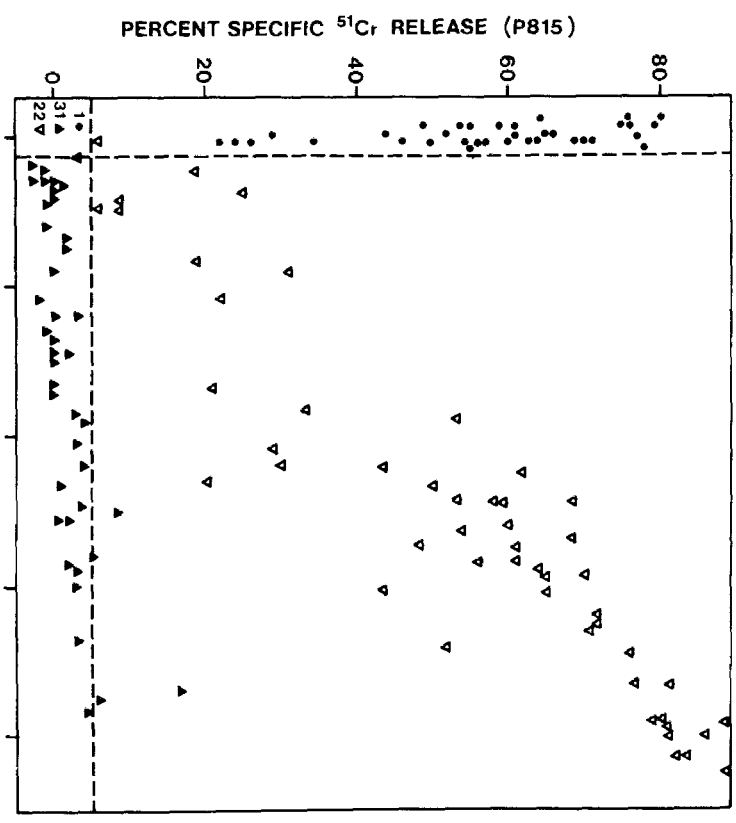

FIG. 2. Phenotypic activities for individual test cultures or "subclones" derived from parent cultures or "clones" that were originally established at $\lambda \sim 1$ with $\hat{\kappa} \geq 0.94$ according to the analysis in Table VII. Subcultures were derived from original parent cultures with activity phenotype: ▽, A^+B^+; ●, A^+B^-; ▲, A^-B^+. Reproduced from Fig. 3 of Taswell *et al.* (1980), *The Journal of Experimental Medicine* **151**, pp. 1372–1385, by copyright permission of The Rockefeller University Press. For discussion of this figure, refer to Section VII,B.

VIII. Partition Analysis

Finally, the principles of LDAs can be used to examine the interactions of different kinds of BAPs within heterogeneous populations. In this section as explained in Section VII, the term population refers to a collection of BAPs in a single sample (i.e., in a test preparation) and not to a collection of multiple samples. Thus, this examination, called partition analysis by Lefkovits and Waldmann (1984), presumably reveals the presence of one or more interacting pairs of BAPs that are diluted to limiting doses at different POP dose levels and are thus revealed at different points along the dose–response curve. Partition analysis with this interpretation of interacting pairs of BAPs has been developed by immunologists and is characterized by saw tooth or sinusoidal curves on plots of $Y_d = \ln p_d$ versus $X_d = \lambda_d$ (Waldmann, 1977; Corley *et al.*, 1978; Eichmann *et al.*, 1983). Various models have been proposed to explain these multiphasic curves (Lefkovits and Waldmann, 1979; Fey *et al.*, 1983), but their validity has not yet been established. In order to do so, (1) experiments should be performed so that there are many points (dose groups with dose-levels λ_d) for each presumed arcing segment and linear segment of the dose–response curve, and (2) the complete data set $\{\lambda_d, n_d, r_d | 1 \leq d \leq D\}$ for each experiment should be published in tables such as Table I so that any investigator can analyze the data for both model selection and parameter estimation. Segments of the curves have already been used to estimate frequencies based on the SHPM, but rules for determining inclusion of points within the segment to be analyzed have not been published. If the SHPM does apply to linear segments of these multiphasic curves, then plots corresponding to the validity tests of Section II,A should produce figures with displaced but parallel line segments giving the appearance of step functions. Although these experiments use the principles of LDAs, they are nevertheless experiments and not assays. Until these experiments become routine assays, the importance of publishing data sets of numbers in tables (rather than curves in figures) cannot be overemphasized. Publication of these data will enable all interested biologists and statisticians to examine this problem more carefully.

IX. Conclusion

LDAs were originally developed and have been most extensively used by public health officials and sanitary engineers for the examination of water supplies, sewage and waste water, and dairy products (Phelps, 1908; McCrady, 1915; Greenwood and Yule, 1917; Greenberg *et al.*, 1985; Richardson, 1985). As discussed in Section I, LDAs have also been used by investigators from many other biological and medical sciences. It is the immunologists, however, who have been responsible for renewing interest over the past decade in the continuing development of methods for the statistical analysis of LDA data. This renewed interest derives from the increased size of assays and conplexity of applications in immunology. Sanitary engineers typically use 1, 5, or 10 replicates for each of from 1 to 3 dose groups to determine, for example, whether the concentration of bacteria in drinking water does not exceed the maximum safe level. Immunologists, however, typically use, say, 24, 60, 192, or more replicates for each of from 3 to 6 or more dose groups to perform experimental comparisons, clonal analyses, and partition analyses (Sections VI, VII, and VIII, respectively) that are relatively much more complicated. Renewed interest in statistical research for LDAs also derives from advances in computers and statistics. Efficient statistical analysis of data from larger, more complicated assays would never have been practically feasible without the assistance of the powerful yet economical personal computers that have become available just within the past decade. Many new theories and methods have been developed in statistics over the past several decades, some that have been and some that have not yet been applied to LDAs, as discussed throughout this article. Certainly, much work remains to be done.

This article attempts to provide an outline of all statistical methods relevant to LDAs, reviewing past origins and recommending future directions. Apparently, it is the first such attempt to collect statistical work on LDAs from many diverse fields and to unify it with a common terminology and notation within a systematic treatment of validity tests, parameter estimators, and assay design for both sample and population LDAs. Hopefully, it will not be the last such attempt. The distinction between sample and population LDAs (Section I,C) and the boundary between the concepts of using sample LDAs to estimate sample parameters (Section III) and population LDAs to estimate both sample and population parameters (Section IV,B) should be explored further. These issues of parameter estimation should be investigated within a conceptual framework that fully integrates model discrimination and selection (Sections II and IV,A) and design optimization (Section V). The goal of this approach should be to extract maximum information from past assays in a sequence in order to obtain maximum information from future assays in the sequence. Furthermore, it should also be to estimate the biological variance $V(\phi_a) = \sigma^2_\Phi$ of the samples $\{\phi_a | 1 \leq a \leq A\}$ within the population in addition to the usual statistical variance $V(\hat{\phi}_a | \phi_a)$ of the sample estimator $\hat{\phi}_a$ given the sample ϕ_a. A method for estimating σ^2_Φ is introduced for the first time in

this article (Section IV,B), and examples are provided with estimates of the biological variance between samples within the same population for several different populations (Section VI). Estimation of σ_Φ^2 will enable investigators to better characterize their study populations by quantitating the biological variation between the BAP frequencies of test preparations from individual mice, patients, or other sources. LDAs have been used for almost a century now. They have proved to be valuable tools in the hands of biological and medical scientists for the separation, characterization, and quantitation of BAPs and their clonal progeny. Continuing development and proper use of new methods of statistical analysis for LDAs can only serve to enhance the power of these tools.

Computer Software

Programs for all statistical methods detailed in this article have been developed for the Commodore Amiga, Apple Macintosh, IBM PC, and compatible personal computers. Contact the author for further information about the availability of this software.

Acknowledgments

The author was supported by an NIH Training Grant in Biostatistics, 5 T32 CA09337-06 from the National Cancer Institute, and a cooperative agreement with the SIAM Institute for Mathematics and Society and the Environmental Protection Agency.

References

Amemiya, T. (1980). The n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimator. *Ann. Stat.* **8,** 488–505.

Armitage, P., and Spicer, C. C. (1956). The detection of variation in host susceptibility in dilution counting experiments. *J. Hyg.* **54,** 401–414.

Barbarese, E., Pfeiffer, S. E., and Carson, J. H. (1983). Progenitors of oligodendrocytes: Limiting dilution analysis in fetal rat brain culture. *Dev. Biol.* **96,** 84–88.

Barkworth, H., and Irwin, J. O. (1938). Distribution of coliform organisms in milk and the accuracy of the presumptive coliform test. *J. Hyg.* **38,** 446–457.

Bartlett, M. S. (1935). A study by statistical methods of the distribution of species in grassland associations—Appendix. *Ann. Bot.* **49,** 775–777.

Berkson, J. (1980). Minimum chi-square, not maximum likelihood! *Ann. Stat.* **8,** 457–487.

Blackman, G. E. (1935). A study by statistical methods of the distribution of species in grassland associations. *Ann. Bot.* **49,** 749–777.

Borth, D. M. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. *J. R. Stat. Soc. Ser. B.* **37,** 77–87.

Chambers, E. A., and Cox, D. R. (1967). Discrimination between alternative binary response models. *Biometrika* **54,** 573–578.

Clark, H. (1927). On the titration of bacteriophage and the particulate hypothesis. *J. Gen. Physiol.* **11,** 71–81.

Corley, R. B., Kindred, B., and Lefkovits, I. (1978). Positive and negative allogeneic effects mediated by MLR-primed lymphocytes: Quantitation by limiting dilution analysis. *J. Immunol.* **121,** 1082–1089.

Cornell, R. G., and Speckman, J. A. (1967). Estimation for a simple exponential model. *Biometrics* **23,** 717–737.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. R. Stat. Soc. Ser. B* **24,** 406–424.

Cunningham, A. (1915). Studies on soil protozoa. *J. Agric. Sci.* **7,** 49–74.

Eichmann, K., Fey, K., Kuppers, R., Melchers, I., Simon, M. M., and Weltzien, H. U. (1983). Network regulation among T cells; Conclusions from limiting dilution experiments. *Springer Semin. Immunopathol.* **6,** 7–32.

Eisenhart, C., and Wilson, P. W. (1943). Statistical methods and control in bacteriology. *Bacteriol. Rev.* **7,** 57–137.

Fazekas de St. Groth, S. (1982). The evaluation of limiting dilution assays. *J. Immunol. Methods* **49,** R11–R23.

Ferreira, P. E. (1981). Extending Fisher's measure of information. *Biometrika* **68,** 695–698.

Fey, K., Melchers, I., and Eichmann, K. (1983). Quantitative studies on T cell diversity. IV. Mathematical analysis of multiple limiting populations of effector and suppressor T cells. *J. Exp. Med.* **158,** 40–52.

Finney, D. J. (1951). The estimation of bacterial densities from dilution series. *J. Hyg.* **49,** 26–35.

Finney, D. J. (1952, 1964, 1978). "Statistical Method in Biological Assay." 1st, 2nd, and 3rd Eds., Charles Griffen, London.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. London Ser. A* **222,** 309–368.

Fisher, R. A. (1928). "Statistical Methods for Research Workers," 2nd Ed. Oliver & Boyd, Edinburgh.

Fisher, R. A. (1935). "The Design of Experiments," 1st Ed. Oliver & Boyd, Edinburgh.

Fisher, R. A., Thornton, H. G., and MacKenzie, W. A. (1922). The accuracy of the plating method of estimating the density of bacterial populations. *Ann. Appl. Biol.* **9,** 325–359.

Gart, J. J. (1959). An extension of the Cramer–Rao inequality. *Ann. Math. Stat.* **30,** 367–380.

Gart, J. J., and Weiss, G. H. (1967). Graphically oriented tests for host variability in dilution experiments. *Biometrics* **23,** 269–284.

Greenberg, A. E., Trussell, R. R., Clesceri, L. S., and Franson, M. A. H., eds. (1985). "Standard Methods for the Examination of Water and Wastewater," 16th Ed. American Public Health Association, Washington, D.C.

Greenwood, M., and Yule, G. U. (1917). On the statistical interpretation of some bacteriological methods employed in water analysis. *J. Hyg.* **16,** 36–54.

Groves, D. L., Level, W. E., and Makinodan, T. (1970). A model for the interaction of cell types in the generation of hemolytic plaque-forming cells. *J. Immunol.* **104,** 148–165.

Halvorson, H. O., and Ziegler, N. R. (1933). Application of statistics to problems in bacteriology. I. A means of determining bacterial population by the dilution method. *J. Bacteriol.* **25,** 101–121.

Harris, R. R., and Kanji, G. K. (1983). On the use of minimum chi-square estimation. *Statistician* **32,** 379–394.

Hewitt, H. B. (1958). Studies of the dissemination and quantitative transplantation of a lymphocytic leukaemia of CBA mice. *Br. J. Cancer* **12,** 378–401.
Hill, W. J., Hunter, W. G., and Wichern, D. W. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics* **10,** 145–160.
Iwaszkiewicz, K., and Neyman, J. (1931). Counting virulent bacteria and particles of virus. *Acta Biol. Exp.* **6,** 101–142.
Keating, J. P., and Mason, R. L. (1985). Practical relevance of an alternative criterion in estimation. *Am. Stat.* **39,** 203–205.
Kennedy, J. C., Till, J. E., Siminovitch, L., and McCulloch, E. A. (1966). The proliferative capacity of antigen-sensitive precursors of hemolytic plaque-forming cells. *J. Immunol.* **96,** 973–980.
Kerr, J. D. (1971). The probability of disease transmission. *Biometrics* **27,** 219–222.
Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Am. Stat. Assoc.* **73,** 805–811.
Lefkovits, I., and Waldmann, H. (1979). "Limiting Dilution Analysis of Cells in the Immune System." Cambridge Univ. Press, London.
Lefkovits, I., and Waldmann, H. (1984). Limiting dilution analysis of the cells of immune system. I. The clonal basis of the immune response. *Immunol. Today* **5,** 265–268.
Loyer, M. W., and Hamilton, M. A. (1984). Interval estimation of the density of organisms using a serial-dilution experiment. *Biometrics* **40,** 907–916.
McCrady, M. H. (1915). The numerical interpretation of fermentation-tube results. *J. Infect. Dis.* **17,** 183–212.
McCullagh, P., and Nelder, J. A. (1983). "Generalized Linear Models." Chapman & Hall, London.
Maindonald, J. H. (1984). "Statistical Computation." Wiley, New York.
Makinodan, T., and Albright, J. F. (1962). Cellular variation during the immune response: One possible model of cellular differentiation. *J. Cell. Comp. Physiol.* **60**(Suppl. 1), 129–144.
Martin, P. J., and Hansen, J. A. (1985). Quantitative assays for detection of residual T cells of T-depleted human marrow. *Blood* **65,** 1134–1140.
Mather, K. (1949). The analysis of extinction time data in bioassay. *Biometrics* **5,** 127–143.
Miller, R. G. (1982). Clonal analysis by limiting dilution. An overview. *In* "Isolation, Characterization, and Utilization of T Lymphocyte Clones" (C. G. Fathman and F. W. Fitch, eds.), Chap. 18, pp. 219–231. Academic Press, New York.
Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Am. Stat. Assoc.* **78,** 47–65.
Peto, S. (1953). A dose–response equation for the invasion of micro-organisms. *Biometrics* **9,** 320–335.
Pfanzagl, J. (1980). Discussion of "Minimum chi-square, not maximum likelihood!" by J. Berkson. *Ann. Stat.* **8,** 479–481.
Phelps, E. B. (1908). A method for calculating the numbers of *B. coli* from the results of dilution tests. *Publ. Health Pap. Rep.* **33,** 9–13 (American Public Health Association, *Am. J. Publ. Hyg.*)
Porter, E. H., and Berry, R. J. (1963). The efficient design of transplantable tumor assays. *Br. J. Cancer* **17,** 583–595.
Rao, C. R. (1981). Some comments on the minimum mean square error as a criterion of estimation. *In* "Statistics and Related Topics" (M. Csörgö, D. A. Dawson, J. N. K. Rao, and A. K. Md. e. Salenh, eds.), pp. 123–143. North-Holland, Publ., Amsterdam.
Reed, L. J., and Muench, H. (1938). A simple method of estimating fifty per cent endpoints. *Am J. Hyg.* **27,** 493–497.

Richardson, G. H., ed. (1985). "Standard Methods for the Examination of Dairy Products," 15th ed. American Public Health Association, Washington, D.C.
Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Stat.* **11,** 713–723.
Roberts, E. A., and Coote, G. G. (1965). The estimation of concentration of viruses and bacteria from dilution counts. *Biometrics* **21,** 600–615.
Salsburg, D. S. (1985). The religion of statistics as practiced in medical journals. *Am. Stat.* **39,** 220–223.
Stein, M. F. (1919). The interpretation of *B. coli* test results on a numerical and comparative basis as applied to the investigation of water and sewage. *J. Bacteriol.* **4,** 243–265.
Swindel, B. F. (1983). Choice of size and number of quadrats to estimate density from frequency in Poisson and binomially dispersed populations. *Biometrics* **39,** 455–464.
Taswell, C. (1981). Limiting dilution assays for the determination of immunocompetent cell frequencies. I. Data analysis. *J. Immunol.* **126,** 1614–1619.
Taswell, C. (1984a). Limiting dilution assays for the determination of immunocompetent cells frequencies. II. Experimental design for single-dose assays. M. S. thesis, Department of Mathematics, New York University.
Taswell, C. (1984b). Limiing dilution assays for the determination of immunocompetent cell frequencies. III. Validity tests for the single-hit Poisson model. *J. Immunol. Methods* **72,** 29–40.
Taswell, C., MacDonald, H. R., and Cerottini, J. -C. (1979). Limiting dilution analysis of alloantigen-reactive T lymphocytes. II. Effect of cortisone and cyclophosphamide on cytolytic T lymphocyte precursor frequencies in the thymus. *Thymus* **1,** 119–131.
Taswell, C., MacDonald, H. R., and Cerottini, J. -C. (1980). Clonal analysis of cytolytic T lymphocyte specificity. I. Phenotypically distinct sets of clones as the cellular basis of cross-reactivity to alloantigens. *J. Exp. Med.* **151,** 1372–1385.
Taylor, J. (1962). The estimation of numbers of bacteria by tenfold dilution series. *J. Appl. Bacteriol.* **25,** 54–61.
Thomas, H. A. (1955). Statistical analysis of coliform data. *Sewage Ind. Wastes* **27,** 212–222.
Thompson, K. H. (1962). Estimation of the proportion of vectors in a natural population of insects. *Biometrics* **18,** 568–578.
Waldmann, H. (1977). Conditions determining the generation and expression of T helper cells. *Immunol. Rev.* **35,** 121–145.
Wilson, G. (1983). The bacteriology of air, water, and milk. *In* "Topley and Wilson's Principles of Bacteriology, Virology, and Immunity" (G. Wilson and H. M. Dick, eds.), 7th Ed., Vol. 1, Chap. 9, pp. 251–295. Williams & Wilkens, Baltimore.
Worcester, J. (1954). How many organisms? *Biometrics* **10,** 227–234.

Cell Separation

METHODS AND SELECTED APPLICATIONS

EDITED BY

THOMAS G. PRETLOW II AND THERESA P. PRETLOW

Institute of Pathology
Case Western Reserve University
Cleveland, Ohio

VOLUME 4

1987

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Orlando San Diego New York Austin
Boston London Sydney Tokyo Toronto