Multiple-dose design and bias-reducing methods for limiting dilution assays

L.W.G. Strijbosch

*Computer Applications Group,
Department of Business and Economics,
Tilburg University,
P.O. Box 90153,
NL-5000 LE Tilburg
The Netherlands*

R.J.M.M. Does

*Nederlandse Philips Bedrijven B.V.
Centre for Quantitative Methods HCM-711
P.O. Box 218
NL-5600 MD Eindhoven
The Netherlands*

W. Albers

*Department of Applied Mathematics
Twente University
P.O. Box 217
NL-7500 AE Enschede
The Netherlands*

This paper gives an overview of several (mostly recent) statistical contributions to the theory of Limiting and Serial Dilution Assays (LDA's, SDA's). A simple and useful method is presented for the setup of a design for an LDA or an SDA. This method is based on several user-supplied design parameters, consisting in the researcher's advance information and other parameters inherent to the particular problem. The commonly used Maximum Likelihood (ML) and Minimum Chi-square methods for the estimation of the unknown parameter in an LDA or an SDA are described and compared to several bias-reducing estimation methods, e.g. jackknife and bootstrap versions of the ML method. One particular jackknife version is recommended.

Key Words & Phrases: experimental design, maximum likelihood, minimum chi-square, jackknife, bootstrap, Monte Carlo comparison.

1. INTRODUCTION

Limiting and Serial Dilution Assays (LDA's, SDA's) are applicable in many areas of bioresearch, such as public hygiene, cell biology, immunology,

bacteriology and virology (see TASWELL, 1987, for a review). In general, the aim of such an analysis is to estimate either the frequency of a particular cell type among other but similar cells (LDA), or the density of particular cells or microorganisms per unit volume of solution, such as bacteria and viruses capable of (un)limited proliferation or having infective properties for certain cells (SDA). LDA's, SDA's and other related assays known as fermentation tube tests, coliform density tests, etc., can all be treated by the same statistical theory, though the underlying biological processes and laboratory techniques differ. Hereinafter the terminology of LDA will prevail. The unknown parameter is commonly estimated by using the "single-hit Poisson model" with binary (positive or negative) data produced by samples taken from different dilutions. The assumptions underlying this model are well-known (see FINNEY, 1978, p. 425, TASWELL, 1981, MILLER, 1982, LEFKOVITS and WALDMANN, 1984) and will be described briefly. Fig. 1 illustrates the setup of a limiting dilution experiment.

Figure 1. The setup of a limiting dilution experiment. The number of dilutions is denoted by m ; x_i is the (mean) number of cells tested in a replicate culture of dilution i ; ϕx_i is the (mean) number of cells of the specific type contained in that culture; ϕ is the parameter to be estimated.

A test preparation contains numerous cells of which an unknown proportion ϕ has a certain property, for example, immunocompetence. From this test preparation, m different dilutions are prepared. Then, from dilution i , n_i replicate cultures are taken such that the numbers of cells in the replicate cultures of dilution i are independent Poisson distributed variables with the same mean number x_i of cells tested. A fraction ϕ of those cells has the intended property. Ideally, the test preparation is diluted until the replicate cultures of the most extreme dilution do not contain any relevant cells: the particular cell is "limiting" until extinction. The use of the Poisson distribution for the number of cells in a sample is justified by the assumptions that the total number of cells in the test preparation is sufficiently large, the suspension has been stirred well before taking any sample or dilution and the sample volume of the cells is

extremely small when compared to the total volume of the sample. A further assumption is that a positive response is obtained for a replicate culture if and only if at least one cell of the specific type is present ("single-hit"). ϕx_i is the expected number of relevant cells for a culture in group i while $\exp(-\phi x_i)$ is the probability of a negative response. The biometrical model can be formulated as follows:

$$\{Y_{ij}\}_{i=1,\dots,m; j=1,\dots,n_i}, \quad (1)$$

where the Y_{ij} are independent Bernoulli-distributed random variables with $P(Y_{ij}=0)=1-P(Y_{ij}=1)=\exp(-\phi x_i)$, $i=1,\dots,m$, $j=1,\dots,n_i$. A negative response for a replicate culture is denoted by zero. Thus the dose-response model $P=\exp(-\phi x)$ (single-hit Poisson model) is used to describe the relation between the mean number of cells tested per replicate culture and the fraction of negatively responding cultures in a group. It is clear that at least some culture groups should be "informative" ($0 < P < 1$, see section 2.2) for values of ϕ in a certain range. This means that advance information about ϕ must be included in the design. No simple and widely accepted methods providing design characteristics on the basis of the researcher's criteria and prior information about the unknown parameter can be found in existing literature.

Many different estimation methods are reported and used for the analysis of data from LDA's and SDA's (see LOYER, 1981, TASWELL, 1981). Some of these (e.g. the least squares (LS) method), are statistically far from ideal but easily executed, while others (like the maximum likelihood (ML) and minimum chi-square (MC) methods) are statistically more appropriate but much less easy to apply. However, with the present widespread availability of (micro-)computers and software, ease of computation is no longer a sufficiently valid argument for preferring the simpler statistical methods. Since McGRADY (1915) described an estimation method that determines an estimator $\hat{\phi}$ ("the most probable number") maximizing the probability of obtaining the specific arrangement of positive and negative replicate cultures observed, the maximum likelihood method has been used by many authors. However, after the simulation study of TASWELL (1981), who compared the maximum likelihood method and the minimum chi-square method and found that the last method should be preferred, a controversy arose with respect to the use of those methods in dilution analysis (see FAZEKAS DE ST. GROTH 1982).

Statisticians can contribute to the execution of an LDA in at least three ways. In the first place they can provide validity tests that detect deviations from the single-hit Poisson model. TASWELL (1984) discusses principles of bioassay validity tests. However, validity tests will not be discussed here. Secondly, they can help the researcher construct an experimental design which will take advantage of existing prior information. And thirdly, they can advise on the estimation methods to be used. The latter two topics form the subject of this study.

In many applications of dilution analysis the assays are very laborious and costly, and, in some circumstances, also unrepeatable. In these cases it is of

vital interest to carefully base an experimental design on available advance information and to choose an estimation method minimizing bias and/or mean square error. Recent research has been performed on design problems (LOYER 1981, TASWELL, 1987, STRIJBOSCH, BUURMAN, DOES et al., 1987, STRIJBOSCH, DOES and BUURMAN, 1988 and STRIJBOSCH, DOES and ALBERS, 1990). And, Monte Carlo studies have been made on the choice of the statistical procedure to be used (SALAMA, KOCH and TOLLEY, 1978, LOYER, 1981, TASWELL, 1981, STRIJBOSCH, BUURMAN, DOES et al., 1987, DOES, STRIJBOSCH and ALBERS, 1988, STRIJBOSCH and DOES, 1988, and COBB, CYR, SCHMEHL and BANK, 1989). Note that there is a difficulty in comparing the results of these Monte Carlo experiments because of the absence of generally accepted design methods; most authors used different experimental designs when generating simulation results. The controversy between TASWELL (1981) and FAZEKAS DE ST. GROTH (1982) emerged partly as a consequence of this kind of difficulty.

2. EXPERIMENTAL DESIGN

2.1. Introduction

The experimental design in dilution analysis concerns the choice of the number m of dilutions, the number n_i of replicates in dilution i , and the (mean) number x_i of cells/organisms to be used per replicate in dilution i ($i = 1, \dots, m$). Assay designs are generally set up starting from a particular prior estimate of the unknown parameter. However, depending on the type of experiment, prior knowledge of the unknown parameter varies significantly. In spite of the fact that dilution assays have been used since the beginning of this century, the development of design methods has been somewhat limited. The design suggestions made by FINNEY (1978, p. 435) for dilution series can be used in practice although ready-to-use formulae are not given. LOYER (1981, p. 53) proposes a complicated procedure for the determination of an experimental design which is not suitable for general use (see STRIJBOSCH, 1989, pp. 23-25). Without giving any design formulae, TASWELL (1987) suggests that "the goal of optimization methods should be to minimize the error of the estimates subject to the constraint of a chosen level of the noninformative assay probability" which is defined as the probability of obtaining either all responses negative or all responses positive.

2.2. The construction of an experimental design

The design of the experiment is dependent on a point estimate $\hat{\phi}$ or (more general) an interval estimate $[\phi_1, \phi_2]$ representing the researcher's boundaries for possible values of ϕ . The design problem is split into two parts. In the first part m and $\{x_i\}_{i=1, \dots, m}$ are determined. The vector containing the response of one individual replicate culture per dilution can be considered as the unit of observation. In the second part the sample size, i.e. $n_i = n$, $i = 1, \dots, m$, is chosen. There are no theoretical objections to choose $n = 1$ and adjust m

accordingly. However, in order to keep the laboratory procedures tractable, it is necessary to work with a limited number of different dilutions.

An important criterium in design considerations is the Fisher information defined by $I(\phi) = \text{var}_{\phi}(\partial \log L(\phi) / \partial \phi)$ where $\log L(\phi)$ denotes the log likelihood function (see (7)):

$$I(\phi) = \sum_{i=1}^m n_i x_i^2 / (e^{\phi x_i} - 1). \quad (2)$$

Single-dose designs, (i.e. $m=1$) are discussed by many authors, including FISHER (1922), FINNEY (1978, p. 435) and FAZEKAS DE ST. GROTH (1982). FISHER (1922) calculated that (when $n \rightarrow \infty$) the optimal dose $x = 1.59/\phi$ where ϕ is the unknown frequency, which corresponds to a fraction of 0.203 negative responders. For this value the Fisher information is maximized. This suggests that selecting $x = 1.59/\phi$, where ϕ represents the researcher's best prior guess for ϕ , would provide the most efficient single-dose design. FAZEKAS DE ST. GROTH (1982) stated that the experiment is efficient when the fraction of negative cultures is between 0.10 and 0.37; this range comprises levels above 90% of the maximum possible value of the Fisher information. A suggestion of FINNEY (1978, p. 428) is based on the same principles. However, if the prior value ϕ deviates too much from ϕ , all the replicates might be either negative or positive. In such cases the experiment obviously does not yield any relevant information; nevertheless, one-sided confidence intervals for the unknown parameter can be determined (see STRIJBOSCH, DOES and BUURMAN, 1988).

When a useful prior guess is not available, FINNEY (1978, p. 437) states that the experimenter must base his design on a lower and an upper bound for ϕ . His design proposals, however, lack an indication of the number of dilutions which should be used. STRIJBOSCH, DOES and BUURMAN (1987) develop in a heuristic way a design method which generalizes the efficient single-dose design to a design that can be used when the advance information on ϕ has the nature of a uniform prior distribution on $[\phi_1, \phi_2]$. This method assumes that the true value ϕ is between these bounds. Only fractions of negative responders between certain values P_1 and P_2 ($P_1 < P_2$) are considered "sufficiently informative" by the researcher. The aim is to have, per experiment (on the average), d dilutions that yield sufficiently informative fractions. Then, using the formulae (3)-(6), the number of dilutions m , the mean number of cells x_1 in the replicate cultures of the first dilution, and finally the dilution factor c can be determined:

$$m = d + \text{ENTIER}(\log(\phi_2/\phi_1)/\log(c)), \quad (3)$$

$$x_1 = \{(\log P_1)(\log P_2)c^{1-m}/(\phi_1\phi_2)\}^{1/2}, \quad (4)$$

$$c = \{(\log P_1)/(\log P_2)\}^{1/d}, \quad (5)$$

$$x_i = x_1 c^{i-1}, \quad i = 2, \dots, m. \quad (6)$$

('ENTIER' means 'the integer part of' and $\log$ is the natural logarithm function). STRIJBOSCH, BUURMAN, DOES et al., (1987) prove that this procedure yields a design with the property that for each value $\phi \in [\phi_1, \phi_2]$ exactly d values

of $\{x_i\}_{i=1,\dots,m}$ satisfy the condition $P_1 \leq \exp(-\phi x_i) < P_2$. When demanding a geometrical series and a condition of symmetry (e.g. the median of $\{\log(x_i)\}_{i=1,\dots,m}$ is not dependent on d), (3)-(6) is unique. Furthermore, with fixed $[\phi_1, \phi_2]$, each geometrical series $\{x_i\}_{i=1,\dots,m}$ uniquely corresponds with values for d, P_1 and P_2 . Note that the efficient single-dose design is obtained from (4) and (5) with $P_1 = P_2 = 0.203$, $d = 1$ and $\phi_1 = \phi_2 = \phi$. Several refinements may improve an already chosen design. Due to the presence of the ENTIER function, formula (3) allows to some extent the widening of the interval $[\phi_1, \phi_2]$, or, alternatively, the reduction of c without changing m . Furthermore, formula (5) allows a shift of the range $[P_1, P_2]$ without changing c and m . It is easily understood that these options can polish a design in order to obtain better statistical properties, or to approach the requirements of the researcher to a higher degree. When the design parameters ϕ_1, ϕ_2, d, P_1 and P_2 are fixed, the sample size can be determined on the basis of a required (user-supplied) coefficient of variation. The reader is referred to STRIJBOSCH, DOES and BUURMAN (1980) for more details on possible refinements, the determination of the sample size and a description of the computer program DESIGN which will expediate the setup of a design in the laboratory.

As is described by STRIJBOSCH, DOES and ALBERS (1990), the basic design method (3)-(6) can be generalized for some other related and widely used biological models, such as the logistic regression model which is most common in bioassay.

3. CONVENTIONAL ESTIMATION METHODS

In this section two conventional statistical procedures, i.e. the Maximum Likelihood (ML) method and the Minimum Chi-square (MC) method, for the evaluation of the experimental data in dilution analysis will be described. Most of the literature published on dilution series has been devoted to methods for estimating the unknown parameter. Recent references to this literature are FINNEY (1978), SALAMA, KOCH and TOLLEY (1978), TASWELL (1981), LOYER (1981), FAZEKAS DE ST. GROTH (1982), LOYER and HAMILTON (1984) and TASWELL (1987) among others.

TASWELL (1981) compared four statistical procedures for the estimation of the unknown parameter in dilution series with Monte Carlo simulation. These methods are the Least Squares (LS) method, the Weighted Mean (WM) method, the ML method and the MC method. He found that the four estimators can be ranked in the following order by decreasing bias and mean square error: LS, WM, ML and MC. There is no doubt that the last two methods are preferable to the other two. However, regarding the comparison of the ML and MC method, the meaning of the conclusions of TASWELL (1981) cannot be fully understood, since it is not entirely clear which experimental designs he used in the simulations. The properties of estimation methods are dependent on the experimental design used. STRIJBOSCH (1989, pp. 87, 88) argues that simulating the way TASWELL (1981) did favours MC.

LOYER (1981) compared the expected values and exact mean square errors of

the ML estimator with those of some alternative estimators which are not asymptotically efficient and should not generally be used when adequate computer facilities are available. As the considered estimators seem to be most useful for very small designs (e.g. $m = 3$, $n = 5$), they are not considered here. The main interest here pertains to LDA where designs are usually larger.

FAZEKAS DE ST. GROTH (1982) advanced several arguments for the use of ML methods instead of MC methods. He proposed the ML estimation procedure as set out in detail in FINNEY (1978).

3.1. The maximum likelihood (ML) method

Since the ML method has desirable properties such as asymptotic sufficiency, efficiency and consistency (c.f. KENDALL and STUART, 1973), it is clear that this method in dilution series is advocated by many authors (c.f. MATHER, 1949, PETO, 1953, FAZEKAS DE ST. GROTH, 1982 and BUURMAN, DAEMEN, GROENEWEGEN et al., 1983). One of the shortcomings of the ML method is that it leads to biased estimators.

From grounds of sufficiency it follows (c.f. (1)) that the relevant observations from an LDA consist of the independent binomial random variables R_i which are defined by $R_i = \sum_{j=1}^n (1 - Y_{ij})$, i.e. $R_i \sim \text{Bin}(n_i, \exp(-\phi x_i))$, $i = 1, \dots, m$. The vector $(R_1, \dots, R_m)$ will be denoted by R . Furthermore let $n_i - R_i$ be denoted by Q_i and $(Q_1, \dots, Q_m)$ by Q . The ML estimator $\hat{\phi}_{\text{ML}}$ is determined as the value of ϕ that maximizes

$$\log L(\phi) = \sum_{i=1}^m \sum_{j=1}^n \{ -(1 - Y_{ij})\phi x_i + Y_{ij} \log(1 - \exp(-\phi x_i)) \}$$

or

$$\log L(\phi) = \sum_{i=1}^m \{ -R_i \phi x_i + Q_i \log(1 - \exp(-\phi x_i)) \}, \quad (7)$$

where $L(\phi)$ denotes the relevant part of the likelihood function. The value of ϕ that maximizes $\log L(\phi)$ is usually calculated by Newton's method of iterative approximation. A serious problem might seem to arise with the ML estimation procedure. If all R_i are 0, which event occurs with an (extremely) small but positive probability, then $\hat{\phi}_{\text{ML}} = \infty$, and hence $E\hat{\phi}_{\text{ML}} = \infty$, thus leading to an infinite, rather than an asymptotically negligible, bias. This undesirable phenomenon can be eliminated by the following simple modification which will be tacitly assumed hereafter: if all R_i indeed happen to be 0, then modify $\hat{\phi}_{\text{ML}}$ by replacing an arbitrary Y_{ij} , say one corresponding with the smallest x_i , by 0. DOES, STRIJBOSCH and ALBERS (1988) demonstrate by way of example for the case $m = 1$ that this simple modification suffices to reduce the bias from infinity to the desired order n_1^{-1} .

Evaluated at $\hat{\phi}_{\text{ML}}$, the second derivative of (7) provides an approximation of $-I(\phi)$. An estimator of the variance of $\hat{\phi}_{\text{ML}}$ is given by

$$S^2(\hat{\phi}_{\text{ML}}) = \frac{-1}{\partial^2 \log L(\phi) / \partial \phi^2} \Big|_{\hat{\phi}_{\text{ML}}}.$$

3.2. The minimum chi-square (MC) method

Large sample theory shows that the difference between ML estimators and MC estimators tends to disappear when the number of observations is increased. Furthermore, ML and MC estimators have some important properties in common such as consistency, asymptotic normality of distribution and asymptotic efficiency (see RAO, 1957). The use of MC estimators as compared to ML estimators in bioassays has been a matter of discussion (see BERKSON, 1980, FAZEKAS DE ST. GROTH, 1982, MANTEL, 1985). BERKSON (1980) states that the use of MC estimators is preferable in cases where it is not certain that the ML estimation procedure is best. However, in the discussion of that paper this point of view is criticized.

The MC estimator ϕ_{MC} is determined as the value of ϕ that minimizes

$$\chi^2(\phi) = \sum_{i=1}^m \frac{(R_i - n_i \exp(-\phi x_i))^2}{n_i \exp(-\phi x_i) (1 - \exp(-\phi x_i))}. \quad (8)$$

Since the expected value of the MC estimator is infinite, $\hat{\phi}_{MC}$ has been adapted in the same manner as $\hat{\phi}_{ML}$ has. An estimator of the variance of $\hat{\phi}_{MC}$ is given by

$$S^2(\hat{\phi}_{MC}) = \frac{2}{\partial^2 \chi^2(\phi) / \partial \phi^2} \Big|_{\hat{\phi}_{MC}}.$$

4. BIAS-REDUCING ESTIMATION METHODS

Modifying the ML estimator according to the method of section 3.1 has the advantage that for all experimental results an estimate can be determined, while the bias is reduced to the desired order. However, the resulting estimator is still positively biased. Thus methods which can reduce the bias of this modified ML estimator further are called for. The positive bias of the ML estimator (some authors exclude the event of all positive responses from their computations, others, like SALAMA, KOCH and TOLLEY, 1978, used a special modification) is well known by researchers in the field.

Resampling plans such as the jackknife and the bootstrap provide estimators of bias and variance for an extremely wide class of statistics. In this section these methods are discussed and applied to dilution analysis. Although both application to the ML estimator and to the MC estimator could be performed, only application to the ML estimator will be elaborated on in this section. The main reasons for this choice are the general objections against MC as pointed out by MANTEL (1985), the fact that a more developed theory for the ML method is available, and the restriction of available computer time for the inherently time-consuming Monte Carlo comparisons. In most applications, jackknife and bootstrap estimators can be constructed in different ways. When applied to dilution analysis, three versions of jackknifed ML estimators are worthwhile to study, whereas four bootstrapped ML estimators are interesting. Finally, methods to reduce the bias of the ML estimator devised by SALAMA,

KOCH and TOLLEY (1978) are described.

4.1. The jackknife (three versions)

The concept of the jackknife method was originally proposed by QUENOUILLE (1949). The name "jackknife" is due to Tukey. QUENOUILLE (1949) introduced a technique for reducing the bias of a serial correlation estimator based on splitting the sample into two parts. QUENOUILLE (1956) explored the general applicability of this concept when splitting the sample into g parts of size h each. MILLER (1974) gives an introduction and a historical survey. Here the "leaving-one-out" jackknife is considered where g equals the sample size and $h = 1$.

Let $X_1, \dots, X_N$ be a sample of independent and identically distributed (iid) random variables. Suppose that one wishes to estimate a parameter Θ based on the sample of size N . Consider an estimator $T_N = T_N(X_1, \dots, X_N)$ of the parameter Θ . Let $T_{N-1}(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_N)$ be the corresponding estimator based on the sample of size $(N-1)$ where observation k has been deleted. Then the k -th jackknifed pseudo-value is defined as

$$T_{Nk}^J = NT_N - (N-1)T_{N-1}(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_N), \quad k = 1, \dots, N. \quad (9)$$

The jackknife estimator T_N^J defined by

$$T_N^J = N^{-1} \sum_{k=1}^N T_{Nk}^J \quad (10)$$

in many cases reduces the bias (see EFRON, 1982). The reason behind this is: if

$$E_{\Theta} T_N = \Theta + a_1 N^{-1} + a_2 N^{-2} + o(N^{-3}), \text{ as } N \rightarrow \infty,$$

for some constants a_1 and a_2 , then

$$E_{\Theta} T_N^J = \Theta + o(N^{-2}), \text{ as } N \rightarrow \infty.$$

Thus the jackknife procedure reduces the bias by eliminating the first order bias term. TUKEY (1958) suggested that the jackknife method could also be used to estimate the variance and obtain robust interval estimation. The jackknife estimator of variance is given by

$$S_{\Theta}^2(T_N^J) = [\sum_{k=1}^N (T_{Nk}^J)^2 - (\sum_{k=1}^N T_{Nk}^J)^2 / N] / (N(N-1)).$$

In the case of dilution series, the pseudo-values in (9) can be obtained in three different ways. When $n_i = n$, $i = 1, \dots, m$, the biometrical model (1) is a matrix with columns that are iid random vectors. As jackknife estimators are in general determined from iid variables, the natural way to jackknife is to drop one column from (1) at a time. Thus, it is necessary that the individual elements of (1) are known. The columns of this matrix provide the $X_1, \dots, X_N$ variables in this case. This procedure produces the jackknife estimator ϕ_{Jc} . The bias-reducing effect of the jackknife in the iid case is a consequence of general theory of jackknife methods. However, this jackknife version violates a certain symmetry inherent to the experiment, because one could permute the entries in any row in (1) as the Y_{ij} are iid within rows without altering the $\hat{\phi}_{ML}$. This

would alter the estimator $\hat{\phi}_{Jc}$ which is a troubling lack of uniqueness. Another disadvantage of this estimator is that it can only be applied when there is an equal number of replicates in each culture group, i.e., $n_i = n$ for $i = 1, \dots, m$. Therefore two other jackknife versions are considered. One is based on jackknifing the rows of the matrix (1), i.e., the jackknife procedure is based on $R_1, \dots, R_m$ which provide the $X_1, \dots, X_N$ variables in that case. Thus the jackknife estimator $\hat{\phi}_{Jr}$ is obtained. The other jackknife method is based on the elements of the matrix (1), i.e. the procedure is based on $Y_{11}, \dots, Y_{mn}$ providing the $X_1, \dots, X_N$ variables in that case. Thus the jackknife estimator $\hat{\phi}_{Je}$ is obtained. The latter two versions are non-iid cases and are obtained by deleting either one row at a time or one element at a time, respectively. Most of the theoretical work regarding resampling methods like the jackknife and the bootstrap thus far has been for the iid case. Resampling methods justifiable in the iid case may not work in more complex situations. In the Appendix it is indicated why the first order bias term of the ML estimator disappears in these non-iid cases as well (cf. QUENOUILLE 1956). Note that all three jackknife methods are asymptotic methods. The one based on the columns demands $n \rightarrow \infty$, the one based on the rows demands $m \rightarrow \infty$ and, finally, the one based on the elements demands that $\sum_{i=1}^m n_i \rightarrow \infty$. For the determination of $\hat{\phi}_{Je}$ it is not necessary that the individual elements of $\{Y_{ij}\}_{i=1, \dots, m; j=1, \dots, n}$ be known; it is sufficient to know $\underline{R}$.

4.2. The bootstrap (four versions)

There seems to be no general rule to choose between the jackknife and the bootstrap for bias reduction and variance estimation. However, since examples are known for which the application of the bootstrap in more complex situations leads to inappropriate estimators (see WU 1986), it is interesting to compare bootstrap methods with jackknife methods.

The bootstrap method has been introduced by EFRON (1979). Let $X_1, \dots, X_N$ be a sample of iid random variables with common cumulative distribution function F . Suppose again that $T_N = T_N(X_1, \dots, X_N)$ is an estimator of the parameter Θ . Furthermore, let F_N be the empirical distribution function defined by $F_N(x) = \# \{i: X_i \leq x, 1 \leq i \leq N\} / N$ and write $E_G(T_N)$ for the expectation of the statistic T_N computed based upon a random sample of size N from G and similarly for $\text{Var}_G(T_N)$, then the bootstrap bias and variance estimators for T_N are given as $[E_{F_N}(T_N) - T_N]$ and $\text{Var}_{F_N}(T_N)$, respectively. The usual way to evaluate these estimators is by drawing a large number of random samples ("bootstrap samples") from F_N . The ordinary bootstrap approach consists of drawing M bootstrap samples $(X^{(k)}_1, \dots, X^{(k)}_N)_{k=1, \dots, M}$ of size N from F_N (see EFRON, 1982). Such a sample is obtained by computer generated random sampling with replacement. Repeated resampling from F_N results in an approximation $\bar{T}_N^B$ for $E_{F_N}(T_N)$:

$$\bar{T}_N^B = M^{-1} \sum_{k=1}^M T_{Nk}^B,$$

where $T_{Nk}^B = T_N(X_1^{(k)}, \dots, X_N^{(k)})$, $k = 1, \dots, M$. This estimator has a bias in estimating T_N which approximates the bias of T_N in estimating Θ . Thus, $B_N(\Theta)$ is used as an approximation of the bootstrap estimator of bias $[E_{F_N}(T_N) - T_N]$ and defined by

$$B_N(\Theta) = \bar{T}_N^B - T_N.$$

This leads to the bias-corrected "bootstrap estimator" T_N^B for Θ

$$T_N^B = T_N - B_N(\Theta) = 2T_N - \bar{T}_N^B.$$

The bootstrap estimator of variance $\text{Var}_{F_N}(T_N)$ is approximated by

$$S_{\Theta}^2(T_N^B) = [\sum_{k=1}^M (T_{Nk}^B)^2 - (\sum_{k=1}^M T_{Nk}^B)^2 / M] / (M-1).$$

As discussed before, there are three different ways to define a set $(X_1, \dots, X_N)$ from the biometrical model (1), thus resulting in the bootstrap estimators $\hat{\phi}_{Bc}$, $\hat{\phi}_{Br}$ and $\hat{\phi}_{Be}$. A fourth bootstrap version can be constructed by joining individual samples from each row of $\{Y_{ij}\}$ resulting in a bootstrap estimator which will be denoted by $\hat{\phi}_{Bx}$. A major disadvantage of the x -version is that, when $(R_i = 0 \vee R_i = n_i)$, $i = 1, \dots, m$, all bootstrap samples are equal, so that an estimator of the variance cannot be evaluated. An alternative way to construct a bootstrap estimator could be the parametric bootstrap method as suggested by FREEDMAN (1981).

4.3. Other bias-reducing methods

As the distribution of the random variable R is specified, it is possible to investigate the stochastic expansion of $\hat{\phi}_{ML}$ and base a corrected ML estimator on the corresponding properties. Using Taylor expansions and implicit function theorems, SALAMA, KOCH and TOLLEY (1978), SK&T in short, showed the existence of functions $H(Q, \hat{\phi}_{ML})$ such that the estimator $\hat{\phi}_S$ defined by

$$\hat{\phi}_S = \hat{\phi}_{ML} - H(\bar{Q}, \hat{\phi}_{ML})$$

satisfies

$$E(\hat{\phi}_S) = \phi + \sum_{i=1}^m O(n_i^{-2}),$$

thus removing the first order bias term. Unfortunately, some formulae given by SK&T show typing errors. The correct formulae are given by STRIBOSCH and DOES (1988). SK&T used a special modification of the ML estimator in case all Q_i equal n_i , $i = 1, \dots, m$: $\hat{\phi}_S$ is taken as the maximum over all realizations of the experiment for which at least one of the Q_i is less than n_i . This modification is essentially the same as the modification of the ML estimator described earlier in 4.2.

The paper of SK&T does not provide an estimator for the standard error of $\hat{\phi}_S$ based on the same principles. This, of course, is a drawback of the proposed method, since corresponding confidence interval estimates cannot be determined.

4.4. Confidence intervals

As the ML and MC estimators are asymptotically normally distributed, it is common practice to use as two-sided 95% confidence intervals $\hat{\phi} \pm 1.96s$ where $\hat{\phi}$ is the point estimate and s the estimate of the standard error corresponding to the method used. The same procedure applies to the jackknife and the bootstrap versions of the ML estimator. The proportion in which the calculated confidence bounds contain the true value of the parameter will be investigated. Other, probably better but more complicated, ways to obtain jackknife and bootstrap confidence limits are described by HINKLEY (1977) and HALL (1988), respectively. These methods are not investigated in this study.

5. MONTE CARLO RESULTS

Research on estimators in dilution analysis requires a design method that can be used as a frame of reference. The design method (3)-(6) seems to be an acceptable candidate for use in general dilution assays and in Monte Carlo comparisons since it incorporates the researcher's criteria. Moreover, a means of examining an estimator's performance over a range of parameter values is automatically provided. Thus the design parameters $\phi_1, \phi_2, d, P_1, P_2$ and $n_i, i = 1, \dots, m$, have to be chosen. Though very many different combinations of these parameters are possible and meaningful, some acceptable limitations for a Monte Carlo study can be made. All simulations are performed with an equal number of replicates $n_i = n, i = 1, \dots, m$. With different $n_i, i = 1, \dots, m$, the 'column' versions of the jackknife and bootstrap method could not be executed. Moreover, when the other design parameters are chosen carefully, there is no reason to depart from this choice. Restriction to relatively small values of n seems to be sensible for three reasons. In the first place, it turns out that differences between estimation methods are especially interesting for these values. Secondly, the Monte Carlo experiments use a great deal of CPU-time. And, thirdly, in SDA smaller values of n are commonly used (e.g. $n = 3$, or $n = 6$). As for the choice of ϕ_1 and ϕ_2 , it should first be noted that only the quotient ϕ_2/ϕ_1 is relevant and not the individual values of ϕ_1 and ϕ_2 . This is both a consequence of the experimental design (according to formulae (3)-(6)) and the biological model used in dilution analysis. After consultation of researchers in the field of LDA, a reasonable choice for this quotient appears to be 10. Thus, the following experimental designs are used in the simulations: $\phi_1 = 0.001, \phi_2 = 0.01, P_1 = 0.15, P_2 = 0.70, d = 2$ (leading to $m = 4$) and $d = 3$ (leading to $m = 7$). For the case $d = 2$, the simulation program has been executed with $n = 6, 12$ and 18 , and, for the case $d = 3$, with $n = 6$ and 12 . Simulation results have been obtained for 19 equidistant values of ϕ within the interval $[\phi_1, \phi_2]$ starting with $\phi = \phi_1$ and ending with $\phi = \phi_2$. The number of generated samples for each combination of $\phi_1, \phi_2, P_1, P_2, d, n$ and ϕ was 1,000. As EFRON and TIBSHIRANI (1986) indicated that a number of bootstrap replications in the range 50 to 200 is quite adequate in most situations and taking into consideration the limitations on computer time, the number M of

bootstrap samples was chosen to be 100.

Many criteria could be used to compare the performance of the estimation methods. However, restriction has been made to the following three criteria: the mean relative bias (MRB) determined as the mean bias relative to the true value of ϕ , the coefficient of variation (CV) determined as the square root of the mean square error relative to the true value of ϕ , and the realized confidence level (RCL) determined as the proportion of times that the calculated 95% confidence interval includes the true value of ϕ . These criteria are calculated as follows:

$$\begin{aligned}\text{MRB} &= \text{MB}/\phi, \text{ where } \text{MB} = \sum_{t=1}^T (\hat{\phi}_t - \phi)/T, \\ \text{CV} &= (\text{MSE})^{1/2}/\phi, \text{ where } \text{MSE} = \sum_{t=1}^T (\hat{\phi}_t - \phi)^2/T, \\ \text{RCL} &= \# \{t: \phi \in [\hat{\phi}_t - 1.960s_t, \hat{\phi}_t + 1.960s_t], 1 \leq t \leq T\} / T.\end{aligned}\quad (11)$$

For Monte Carlo replicate t ($1 \leq t \leq 1,000$), $\hat{\phi}_t$ and s_t denote the point estimate and the estimate of the standard error, respectively. The number T of nonmissing estimates is usually 1,000.

5.1. Results

Only the results obtained for the smallest ($d=2, n=6$) and the largest ($d=3, n=12$) design are presented. It turns out that the Jc and the Je estimators have comparable statistical properties. The same is true for the Bc , Be and Bx estimators. Moreover, since the 'element' versions of these estimators have a wider applicability, the Jc , Bc and Bx estimators are omitted from the Figures. SK&T gave two alternatives for a bias-reducing estimator. Again, since the great likeness, only one, indicated by ϕ_{S2} (see STRIJBOSCH, 1989) is presented.

The values of the MRB and the CV by the simulation program for the ML estimator, the jackknife versions Jr and Je , the bootstrap versions Br and Be , the $S2$ and the MC estimators when using the smallest design are presented graphically in the Figures 2a-3b. In order to get rid of disturbing random variation in the original data, we used a spline interpolation option (I=SM50) when producing the plots with the SAS package.

Figure 2a. Mean Relative Bias of the ML, J_r , J_e , Br , Be , S_2 and MC estimators. The design parameters are $\phi_1=0.001$, $\phi_2=0.01$, $P_1=0.15$, $P_2=0.70$, $d=2$ (leading to $m=4$) and $n=6$.

Figure 2b. Mean Relative Bias of the ML, J_r , J_e , Br , Be , S_2 and MC estimators. The design parameters are $\phi_1=0.001$, $\phi_2=0.01$, $P_1=0.15$, $P_2=0.70$, $d=3$ (leading to $m=7$) and $n=12$.

Figure 3a. Coefficient of Variation of the ML, Jr , Je , Br , Be , $S2$ and MC estimators. Design as in Fig. 2a.

Figure 3b. Coefficient of Variation of the ML, Jr , Je , Br , Be , $S2$ and MC estimators. Design as in Fig. 2b.

The MRB of the ML estimator varies from 3 to about 12% in the smallest design. The Je estimator reduce the bias significantly to values around zero, whereas the Jr estimator shows a different behaviour. This can be explained as follows. In the first place, it is intuitively clear that Je is capable of reducing the bias more than Jr , since the higher order bias terms of the Jr and Je estimators are $O(N^{-r})$, $r > 1$ where $N = m$ and $\sum_{i=1}^m n_i$, respectively. It is gratifying to find that the Monte Carlo experiments confirm this statement. Secondly, as compared to the Je estimator, the Jr estimator is more frequently based on modified ML estimators, while the influence of the modification on the

jackknife estimator is generally greater for the Jr version than for the other jackknife versions (see STRIJBOSCH, 1989, p. 75). As can be expected from the theory, the bootstrap estimators are strongly related with the corresponding jackknife estimators. As m and n increase, the jackknife estimators coincide increasingly, as well as the bootstrap estimators, until they are practically indiscernible in the largest design while still reducing the MRB over the full ϕ -range.

When comparing the seven estimators on the basis of the MRB and CV, it becomes clear that in general the Je , Be and $S2$ estimators should be preferred. An attractive property of the jackknife and the bootstrap is that they also produce the variance estimate used to determine proper confidence bounds for ϕ . A disadvantage of the $S2$ estimation method is the lack of an estimator for the variance. From a computational point of view, it is obvious that the jackknife is more attractive than the bootstrap, certainly for the larger designs. The designs which are commonly used for LDA are sufficiently large to justify the general advocacy of the Je estimator for these assays.

Finally, as concerns criterium (11), Figures 4a-b make clear that for most experimental situations the MC estimator of variance produces too small 95% confidence intervals. The ML estimator of variance produces the best confidence intervals. The Je and Be estimators of variance produce on the average 95% confidence intervals which are slightly too small.

Figure 4a. Realized Confidence Levels for the ML, *Je*, *Be* and MC methods. The design parameters are $\phi_1=0.001$, $\phi_2=0.01$, $P_1=0.15$, $P_2=0.70$, $d=2$ (leading to $m=4$) and $n=6$.

Figure 4b. Realized Confidence Levels for the ML, *Je*, *Be* and MC methods. The design parameters are $\phi_1=0.001$, $\phi_2=0.01$, $P_1=0.15$, $P_2=0.70$, $d=3$ (leading to $m=7$) and $n=12$.

However, using percentile points of the corresponding Student distribution (cf. HINKLEY, 1977) instead of the standard normal distribution when calculating the 95% confidence bounds could correct the *Je* confidence intervals satisfactorily. Accordingly, the percentile-*t* method described by HALL (1988) would probably result in confidence intervals with smaller coverage error.

ACKNOWLEDGEMENT

We would like to thank the referee for helpful suggestions.

APPENDIX

Let X_j be independent random variables with smooth (in θ) density functions f_j depending on the same parameter Θ , $j = 1, \dots, N$. Note that this is true in the situation of section 4.1. Denote the first, second and third derivative of $\log f_j(\Theta)$ with respect to Θ by U_j , V_j and W_j , respectively. From the stochastic expansion of the ML estimator $\hat{\Theta}_N$ for Θ it can be shown (see BRILLINGER, 1964. CHIBISOV, 1973) that

$$E_{\Theta}(\hat{\Theta}_N - \Theta) = \frac{\sum_{j=1}^N a_j}{(\sum_{j=1}^N b_j)^2} + O(N^{-3/2}), \text{ as } N \rightarrow \infty,$$

where

$$a_j = E_{\Theta} U_j V_j + \frac{1}{2} E_{\Theta} W_j \text{ and } b_j = E_{\Theta} U_j^2, \quad j = 1, \dots, N.$$

Note that $b_j \geq 0$ for $j = 1, \dots, N$. Let $\hat{\Theta}_N^{(k)}$ denotes the ML estimator for Θ based on $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_N$, then its expectation centered around Θ satisfies

$$E_{\Theta}(\hat{\Theta}_N^{(k)} - \Theta) \doteq \frac{\sum_{j \neq k} a_j}{(\sum_{j \neq k} b_j)^2}, \quad (\text{A.1})$$

where $\doteq$ denotes that higher order terms have been neglected. If $b_1, \dots, b_N$ satisfy the condition

$$\lim_{N \rightarrow \infty} \frac{\max_{1 \leq j \leq N} b_j}{\sum_{j=1}^N b_j} = 0, \quad (\text{A.2})$$

then (A.1) can be approximated by

$$\frac{\sum_{j=1}^N a_j}{(\sum_{j=1}^N b_j)^2} \left(1 + \frac{2b_k}{\sum_{j=1}^N b_j}\right) - \frac{a_k}{(\sum_{j=1}^N b_j)^2}. \quad (\text{A.3})$$

By neglecting higher order terms it follows from (A.1), (A.2) and (A.3) that the k -th jackknifed pseudo-value (see (9)) satisfies

$$E_{\Theta} T'_{Nk} \doteq \Theta + \frac{\sum_{j=1}^N a_j}{(\sum_{j=1}^N b_j)^2} \left[1 - N \frac{2b_k}{\sum_{j=1}^N b_j}\right] + N \frac{a_k}{(\sum_{j=1}^N b_j)^2} \quad (\text{A.4})$$

and hence it follows from (A.4) that the jackknife estimator T'_N (see (10)) satisfies

$$E_{\Theta} T'_N = E_{\Theta} (N^{-1} \sum_{k=1}^N T'_{Nk}) \doteq \Theta$$

which shows that the first order bias term disappears.

REFERENCES

- BERKSON, J. (1980), Minimum chi-square, not maximum likelihood!, *Annals of Statistics* 8, 457-487.
- BRILLINGER, D.R. (1964), The asymptotic behaviour of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates, *Review of the International Statistical Institute* 32, 202-206.
- BUURMAN, W.A., DAEMEN, A.J.J.M., GROENEWEGEN, G., DOES, R.J.M.M., VAN DER LINDEN, C.J. and VEGT, P.A. (1983), Limiting dilution analysis of canine alloantigen-reactive *T* lymphocytes, *Transplantation* 35, 363-367.
- CHIBISOV, D.M. (1973), An asymptotic expansion for a class of estimators containing maximum likelihood estimators, *Theory of Probability and Its Applications* 18, 295-303.
- COBB, L., CYR, L., SCHMEHL, M.K. and BANK, H.L. (1989), Comparison of statistical methods for the analysis of limiting dilution assays, *In Vitro Cellular & Developmental Biology* 25, 76-81.
- DOES, R.J.M.M., STRIJBOOSCH, L.W.G. and ALBERS, W. (1988), Using jackknife methods for estimating the parameter in dilution series, *Biometrics* 44, 1093-1102.
- EFRON, B. (1979), Bootstrap methods: another look at the jackknife, *Annals of Statistics* 7, 1-26.
- EFRON, B. (1982), *The jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia.
- EFRON, B. and TIBSHIRANI, R. (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science* 1, 54-77.
- FAZEKAS DE ST. GROTH, S. (1982), The evaluation of limiting dilution assays, *Journal of Immunological Methods* 49, R11-R23.
- FINNEY, D.J. (1978), *Statistical Method in Biological Assay*, 3rd Edition, Academic Press, New York.
- FISHER, R.A. (1922), On the mathematical foundations of theoretical statistics, *Royal Society of the London Philosophical Transactions (Series A)* 222, 309-368.
- FREEDMAN, D.A. (1981), Bootstrapping regression models, *Annals of Statistics* 9, 1218-1228.
- HALL, P. (1988), Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics* 16, 927-953.
- HINKLEY, D.V. (1977), Jackknife confidence limits using Student *t* approximations, *Biometrika* 64, 21-28.
- LEFKOVITS, I. and WALDMANN, H. (1984), Limiting dilution analysis of the cells of immune system. I. The clonal basis of the immune response, *Immunology Today* 5, 265-268.
- LOYER, M.W. (1981), Using a serial dilution experiment to estimate the density of organisms, Unpublished Ph.D. Thesis, Montana State University.

- LOYER, M.W. and HAMILTON, M.A. (1984), Interval estimation of the density of organisms using a serial-dilution experiment, *Biometrics* 40, 907-916.
- MANTEL, N. (1985), Maximum likelihood vs. minimum chi-square, *Biometrics* 41, 777-783.
- MATHER, K. (1949), The analysis of extinction time data in bioassay, *Biometrics* 5, 127-143.
- MILLER, R.G. (1974), The jackknife - a review, *Biometrika* 61, 1-15.
- MILLER, R.G. (1982), Clonal analysis by limiting dilution. An overview, in: C.G. Fathman and F.W. Fitch, eds. *Isolation, Characterization, and Utilization of T Lymphocyte Clones*. Chap. 18, pp. 219-231, Academic Press, New York.
- PETO, S. (1953), A dose-response equation for the invasion of micro-organisms, *Biometrics* 9, 320-335.
- QUENOUILLE, M.H. (1949), Approximate tests of correlation in time series, *Journal of the Royal Statistical Society Series B* 11, 18-84.
- QUENOUILLE, M.H. (1956), Notes on bias in estimation, *Biometrika* 43, 353-360.
- RAO, C.R. (1957), Theory of the method of estimation by minimum chi-square, *Bulletin de l'Institut International de Statistique* 35, 25-32.
- SALAMA, I.A., KOCH, G.G. and TOLLEY, H.D. (1978), On the estimation of the most probable number in a serial dilution experiment, *Communications in Statistics - Theory and Methods* A7(13), 1267-1281.
- STRIJBOSCH, L.W.G. (1989), *Experimental Design and Statistical Evaluation of Limiting Dilution Assays*, Thesis, Haarlem.
- STRIJBOSCH, L.W.G., BUURMAN, W.A., DOES, R.J.M.M., ZINKEN, P.H. and GROENEWEGEN, G. (1987), Limiting dilution assays: experiment design and statistical analysis, *Journal of Immunological Methods* 97, 133-140.
- STRIJBOSCH, L.W.G. and DOES, R.J.M.M. (1988), Comparison of bias-reducing methods for estimating the parameter in dilution series, *Communications in Statistics - Simulation and Computation* B17(4), 1173-1190.
- STRIJBOSCH, L.W.G., DOES, R.J.M.M. and ALBERS, W. (1990), Design methods for some dose-response models, *Statistics in Medicine* 9, 1353-1363.
- STRIJBOSCH, L.W.G., DOES, R.J.M.M. and BUURMAN, W.A. (1987), Experimentele proefopzet en statistische analyse van verdunningsreeksen (in Dutch), *Kwantitatieve Methoden* 24, 149-160.
- STRIJBOSCH, L.W.G., DOES, R.J.M.M. and BUURMAN, W.A. (1988), Computer aided design and evaluation of limiting and serial dilution experiments, *International Journal of Bio-Medical Computing* 23, 279-290.
- TASWELL, C. (1981), Limiting dilution assays for the determination of immunocompetent cell frequencies I. Data analysis, *Journal of Immunology* 126, 1614-1619.
- TASWELL, C. (1984), Limiting dilution assays for the determination of immunocompetent cell frequencies III: Validity tests for the single-hit Poisson model, *Journal of Immunological Methods* 72, 29-40.
- TASWELL, C. (1987), Limiting dilution assays for the separation, characterization, and quantitation of biologically active particles and their clonal

- progency, in: Pretlow, T.G. and Pretlow, T.P. (eds), *Cell Separation: Methods and Selected Applications 4*, 109-145, Academic Press, New York.
- TUKEY, J.W. (1958), Bias and confidence in not quite large samples, abstract, *Annals of mathematical Statistics* 29, 614.
- WU, C.F.J. (1986), Jackknife, bootstrap and other resampling methods in regression analysis, *Annals of Statistics* 14, 1261-1295.

Received September 1989, Revised March 1990.