Bootstrap Confidence Intervals for Concentration Parameters in Dilution Assays

by

Sabyasachi Basu and Rudy Guerra
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

Robert Read
Animal Resources Center
University of Texas Southwestern Medical Center
Dallas, Texas 75235

Technical Report No. SMU/DS/TR-281

Bootstrap Confidence Intervals for Concentration Parameters in Dilution Assays

Sabyasachi Basu and **Rudy Guerra**
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

Robert Read
Animal Resources Center
University of Texas Southwestern Medical Center
Dallas, Texas 75235

July 20, 1995

Abstract

Many quantitations in science are performed with tests that are capable of detecting at least one target entity in a sample preparation. In such cases, dilution assays can be conducted to estimate the concentration per unit volume of the target substance in the sample. Commonly used confidence intervals for the concentration parameter involve inverting a hypothesis test and have been tabulated for limited designs. Loss of data points, as often occurs in practice, or sample limitations make it difficult to adhere to those designs for which intervals are readily available. In addition, it may be a nontrivial task to implement the common interval methods to accommodate the realized data structure. The bootstrap is a flexible, easily implemented procedure for finding approximate confidence intervals. The bias-corrected version of the bootstrap percentile method is shown through simulations to provide good coverage with relatively short widths in a variety of designs. An application to AIDS research which motivated this work is also presented.

Keywords: product binomial model; maximum likelihood; bias-correction; simulation; AIDS.

1 Introduction

Many quantitations in microbiology, medicine, and other sciences are performed with tests capable of detecting one or more target units (e.g., molecules, infectious units) in a sample preparation, but cannot differentiate readily between different numbers of units at this detection threshold. Such tests give binary data: negative (-), indicating the absence of target substance in the test sample, or positive (+), indicating the presence of at least one target entity in the test sample. For example, we see that one or more bacteria will produce evidence of growth in a culture tube. While one may attempt to estimate numbers in the original sample by comparing the rate of growth to known numbers of a standard bacteria, such an approach suffers from the assumption that all bacteria grow at the same rate. The same can be said for growth of viruses, particularly HIV and other lentiviruses (Myers et al., 1994) and detection of single copies of genetic sequence with polymerase chain reaction (PCR) (Simon et al., 1990; Simmonds et al., 1993).

When such a binary test is available, it is well known (Finney, 1978) that a dilution assay can be performed to estimate the concentration, C, per unit volume of the target substance in a sample preparation. Briefly, a portion of the sample preparation is diluted into D preparations, usually each with a different concentration of the original sample. At each dilution, n_d replicate, unit volume samples are tested for the target

unit resulting in as many binary responses. It is possible to estimate the concentration C by selecting appropriate dilution factors. For comprehensive discussion of dilution assays see, e.g., Finney (1978) or more recently Myers et al. (1994).

Formal analyses of dilution assays have traditionally focussed on point estimation of C with early references dating back to McGrady (1915) and Fisher (1922). More recently, attention has focussed on other statistical aspects regarding the estimation of C. Strijbosch, Does, and Albers (1990) provide a good review with an emphasis on experimental design and quantification of statistical bias in certain estimators of C.

This work concerns confidence interval estimation for C, the concentration parameter. Most intervals to date involve inverting a hypothesis test and have been tabulated for limited configurations of experimental design parameters and nominal confidence levels (e.g., Woodward, 1957; Marth, 1978; APHA; 1976). Different experimental designs or confidence levels require different tables. Even if one follows an experimental design for which intervals are available, the loss of a single data point - which frequently occurs in practice - will require different tables. This is especially true when the sample size is small or it is unclear how to impute "missing" values.

Moreover, to generate a new set of confidence intervals based on inverting a test may require an impractical amount of computer memory, even for moderately sized

designs. In some situations it may be possible to analytically invert statistical tests to yield the intervals. For example, a likelihood ratio (LR) based interval may be obtained - in simple cases - by using the fact that the LR test statistic usually follows an asymptotic chi-square distribution. This approach in the present situation, however, may lead to spurious conclusions since the sample sizes $\sum n_d$ with which we work are typically less than twenty. For general discussion on inverting statistical tests to obtain confidence intervals, especially using LR statistics, see Kalbfleisch (1985).

A more efficient approach to the problem can be achieved by bootstrap methods (Efron, 1979; Efron and Tibshirani, 1993). These procedures can often be implemented with compact computer programs requiring relatively little memory and time to run. In this paper we propose parametric bootstrap intervals for C. Section 2 reviews a standard product binomial model for dilution assay outcomes and discusses current methods of interval construction. The main contribution to the analysis of dilution experiments is made in section 3 where we propose parametric bootstrap confidence intervals based on the product binomial model. Sections 4 and 5 report results from simulation studies and a practical application to AIDS research which motivated this work, respectively. A summary of our findings is given in section 6.

2 Overview of Confidence Intervals for Concentration C

In this section we review a model commonly employed for dilution experiments and its role in finding confidence intervals for the concentration C. Consider an experiment with D dilutions, n_d replicates at dilution $d, d = 1, 2, ..., D$. Replicate, unit volume samples are taken at dilution d such that the following may reasonably be assumed: (1) the probability that no target molecules are detected in a unit volume sample from any given replicate is (approximately) $q_d = exp(-Cu_d)$ where u_d is the dilution factor at dilution d; (2) the number X_d of positive replicates out of n_d follows a binomial distribution with probability of "success" $p_d = 1 - q_d$, and (3) $X_1, ..., X_D$ are independent random variables. The likelihood function over the D dilutions is therefore

$$L = \prod_{d=1}^{D} \frac{n_d!}{x_d!(n_d - x_d)!} p_d^{x_d} q_d^{(n_d - x_d)}. \tag{1}$$

Several important experimental steps are required for the product binomial model to reasonably hold in practice. Myers et al. (1994) provide a good discussion of these issues, as well as a self-contained presentation of the model and model parameter estimation from first principles of probability and statistics.

Using the product binomial model, several procedures for interval estimation of

C have been proposed. Loyer and Hamilton (1984) provide a contemporary review and on the basis of minimal sufficiency, expected width, and detection of improbable outcomes, recommend Sterne-type intervals (Sterne, 1954; Blyth and Still, 1983) over traditionally used intervals proposed by Woodward (1957) and deMan (1977). Standard, normal based intervals $\hat{C} \pm z_{1-\alpha/2} SE(\hat{C})$ motivated by maximum likelihood (ML) theory are considered by Strijbosch et al. (1990); $z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the Gaussian distribution. The estimate $\hat{C}$ is a bias-corrected version of the ML estimator and the standard error (SE) is calculated using either the jackknife or bootstrap. For large sample sizes, the standard interval using the jackknife for both bias-correction and SE provides reasonable coverage.

More recently, Myers et al. (1994) have suggested inverting exact likelihood ratio tests to find confidence intervals. Although their work emphasizes a methodology for constructing the LR intervals, they do cite unpublished results indicating a preference for the LR method. In particular, they compared their proposed intervals to Sterne intervals and the LR method was found to be better based on expected width.

To date, comparative studies of confidence intervals in this context have emphasized either expected length (e.g., Loyer and Hamilton, 1984) or coverage (e.g., Strijbosch et al., 1990). A simultaneous assessment of coverage and length is particularly important in this situation, however, because of the well known positive bias in

the ML estimator of C (Thomas, 1942; Strijbosch et al., 1990). Intuitively, intervals based on estimators with a large positive bias can be expected to provide inadequate coverage relative to the nominal level or yield excessively long intervals. Therefore, it is of interest to understand how the bias affects both operating characteristics of any given interval technique.

In this paper we present results on both coverage and length in a comparison among four types of intervals: two bootstrap methods, likelihood ratio, and normal based intervals. The latter intervals are discussed, as they should be in any comparative study, since they are widely used in practice. The LR intervals are also used a frame of reference since they have been reported to be the preferred technique (Myers et al., 1994).

3 Bootstrap Confidence Intervals

The bootstrap is a procedure whose main application is in estimating various measures of statistical uncertainty. One such standard application is estimating the standard error in an estimate of a parameter. The basic idea is to estimate the probability mechanism P generating the data by $\hat{P}$, next use $\hat{P}$ to generate more data and in turn another observed value of the estimate, and finally assess the variability in the original estimate by that found in many values obtained from $\hat{P}$.

To fix ideas, let $x_1, ..., x_n$ be a random sample from an unknown probability distribution F, and suppose we wish to estimate the standard error in an estimate $\hat{\theta}$ of an unknown parameter θ; e.g., the median of F. The standard nonparametric bootstrap estimates the standard error as follows. First estimate the probability mechanism F by F_n, the empirical distribution function assigning mass $1/n$ at each observed x_i. Second, obtain a bootstrap sample, $x_1^*, ..., x_n^*$, of n independent draws from F_n; and third, estimate the parameter using the bootstrap sample, say $\hat{\theta}^*$. The second and third steps are repeated a large number (B) of times, obtaining bootstrap replicates of the parameter $\hat{\theta}_b^*, b = 1, ..., B$. The sample standard deviation of the B bootstrap replicates is then an estimate of the standard error in $\hat{\theta}$.

The bootstrap is not limited to the standard error as a measure of uncertainty, nor does it require the empirical distribution function as an estimate of the probability mechanism. Indeed, the present application involves confidence intervals as a measure of statistical accuracy and the probability mechanism is estimated by the product-binomial model using an estimate of C - the so-called parameteric bootstrap. We see, then, that the bootstrap can be viewed as a general technology for estimating sampling distributions of statistics.

The product binomial model is a plausible probability mechanism for the outcomes in a dilution bioassay. Therefore, this model provides a means by which we

can generate many dilution outcomes, and in turn, estimate the sampling distribution of the ML estimate of the concentration parameter C. The desired confidence intervals are found by determining appropriate percentiles from the estimated sampling distribution of $\hat{C}$. For example, an approximate 90% confidence interval for C is estimated as the interval defined by the 5th and 95th percentiles of the observed bootstrap versions of $\hat{C}$. This method yields what are known as bootstrap percentile intervals.

Two problems arise in the above prescription. The first is that the product binomial model depends on an unknown parameter, namely, the concentration parameter which is precisely what we want to know. This problem can be overcome by substituting the ML estimate based on the observed data; this step constitutes the first step in the bootstrap paradigm - estimating the probability mechanism. The second problem is due to the well known fact that the ML estimator of C is positively biased. This manifests itself at two levels: the ML estimator $\hat{C}$ based on the observed data overestimates C, and the bootstrap replicates $\hat{C}^*$ overestimate the concentration "parameter" used to generate the bootstrap samples, namely $\hat{C}$. Applying the naive percentile method to the distribution of the $\hat{C}_b^*$ can therefore lead to intervals providing inaccurate coverage.

A partial solution to the second problem is provided by an improved version of the

naive percentile method, the bias-corrected and accelerated (BC_a) percentile method (Efron, 1987; Efron and Tibshirani, 1993). Following Efron and Tibshirani (1993), let $\hat{\theta}^{*(\alpha)}$ denote the 100·αth percentile of B bootstrap replicates $\hat{C}_b^*$. The percentile method purporting to have a confidence level of 1 - 2α estimates the interval as

$$\hat{\theta}^{*(\alpha)} \text{ to } \hat{\theta}^{*(1-\alpha)}. \tag{2}$$

The BC_a intervals are also based on percentiles of the bootstrap replicates, but, to correct for certain problems of the naive percentile method, the percentiles are not necessarily those given by equation (2); they are defined as

$$\hat{\theta}^{*(\alpha_1)} \text{ to } \hat{\theta}^{*(\alpha_2)}, \tag{3}$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right) \tag{4}$$

and

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right). \tag{5}$$

The numbers $\hat{z}_0$ and $\hat{a}$ are the bias-correction and acceleration constants, respectively; $\Phi(\cdot)$ is the standard normal distribution function; and $z^{(\alpha)}$ is the 100·αth percentile point of a standard normal distribution function. The number $\hat{z}_0$ roughly measures the discrepancy between the median of the $\hat{C}_b^*$ values and $\hat{C}$:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{C}_b^* < \hat{C}\}}{B}\right), \tag{6}$$

and is seen to be zero if half of the values of the $\hat{C}_b^*$s are less than $\hat{C}$. The bias in estimating C by $\hat{C}$ is effectively being estimated by the bias in estimating $\hat{C}$ by $\hat{C}_b^*$, and the BC_a is thus seen to account for the bias in the ML estimate of C. The acceleration constant $\hat{a}$ measures the rate of change of the standard error in $\hat{C}$ with respect to C and can be estimated in various ways. Since our main interest, however, is in correcting for the bias in the ML estimate of C, we refer to Efron and Tibshirani (1993) for estimating formulas for $\hat{a}$. For the purposes of this application of the BC_a method we set $\hat{a}$ to zero.

4 Small Sample Simulation Studies

This section reports on simulation experiments designed to investigate actual coverage of various intervals based on sample sizes that are commonly found in the application below.

Simulation Methods

We investigate actual coverage and mean length of four types of confidence intervals: LR, bootstrap percentile, bootstrap BC intervals, and standard normal based intervals. Each simulation experiment involved the generation of K sets of positivity outcomes following the product binomial model for fixed configurations of true con-

centration, number of dilutions, number of replicates per dilution, confidence interval procedure, and confidence level. Actual coverage was calculated as the number of intervals including the true concentration divided by K. The number of iterations per simulation run was K=500 and the number of bootstrap replicates was B=1000.

The simulation runs can be categorized into four groups. The first group follows a design discussed by Myers et al. (1994) and is a six fivefold dilution experiment in duplicate. The second group of runs also involves six dilutions but with more replicates per dilution. The third group involves relatively fewer dilutions with different dilution factors and a moderate number of replicates. The last group follows the design presented in the application below. Each simulation experiment investigates the true concentration parameter at twelve levels: $C = 10, 25(25)250, 500$.

The reported results do not include confidence intervals of infinite length. Specifically, we ignore those cases where the number of positive outcomes equals the sample size $\sum n_d$ of the design, thereby leading to one-sided infinite intervals. In practice, such a situation would typically be followed by higher dilutions until a transition phase in positivity was found. At the other extreme, because the true concentration is bounded below by zero, it is possible and natural to construct finite, one-sided intervals when all outcomes are negative. More discussion on this issue is given by Strijbosch et al. (1990) and Loyer and Hamilton (1984) with the main focus being

on the former extreme case. It suffices to say that the mean lengths of the intervals are highly sensitive to how one treats a result of all positive outcomes. In short, we have chosen to proceed in accordance with how the interval procedure is actually implemented in practice.

Simulation Results

Table 1 shows simulation results from a six fivefold dilution assay running six dilutions with dilution factors $u_d^{-1} = 1$, 5, 25, 125, 625, and 3125. Four types of intervals are constructed at the nominal level: bootstrap percentile (BS-P); bootstrap bias-correction (BS-BC); likelihood ratio (LR); and standard normal intervals where the SE is based on the parameteric bootstrap. When the nominal level is 95% and there are two replicates per dilution, Table 1a, the two bootstrap intervals tend to undercover while the LR and normal intervals overcover relative to the nominal level. However, compared to the BS-BC intervals, the LR and BS-P coverages are at the cost of considerably longer intervals. The normal intervals were constructed as previously discussed and then truncated at zero to reflect the natural parameter space as is commonly done in practice. It is the coverage and length of the truncated intervals denoted by "T-Length" that are actually reported in Table 1a and elsewhere. The operating characteristics of this interval are similar to those of the LR, relatively long intervals exceeding the nominal level. In addition, they tend to be shorter than the

LR intervals.

A graphical analysis of the relationship between mean length and concentration shows that length is approximately proportional to concentration; equivalently, standardized length defined by the ratio of mean length to C is approximately constant. As indicated in Table 1, however, the constant of proportionality need not be the same for each interval method. This fact allows for a comparison of the different methods with respect to a joint analysis of length and coverage with varying concentration parameter values. The numerical results in Table 1a are thus graphically summarized in Figure 1a where the observed coverage is plotted against standardized length; the horizontal line indicates the nominal level. From the figure one can easily see the distinct operating characteristics of the different methods. In particular, note that none of the procedures provide coverage at the nominal level.

Table 1b (Figure 2a) is similar to Table 1a (Figure 1a) except for the nominal level being reduced by 5%. The LR coverage in this case is in closer agreement with the nominal level than previously with no systematic overcoverage. Similarly, both bootstrap methods seemingly perform better in this situation. That this is the case, however, may be an artifact of the experimental design. Since there are two replicates per dilution, only a very limited number of (probable) ML estimators of C are obtained for a fixed true concentration. Therefore, over many bootstrap replications,

each ML estimator will be repeated many times resulting in a highly discrete bootstrap distribution with few distinct percentiles points. Comparing BS-BC coverage of Table 1a with that of Table 1b, we see that for the most part, actual coverage levels are almost the same. This is reflecting the highly discrete bootstrap distribution for this sample size. For larger sample sizes, and consequently more distinct values of $\hat{C}^*$ that are observed, this problem is lessened. The point is illustrated by comparing the BS-BC points in Figure 1b to those in Figure 2b where the results are based on a design with three replicates per dilution. Similarly, compare Figure 1c to 2c showing results based on four replicates per dilution level. In each case now, unlike that with two replicates per dilution, the BS-BC intervals appropriately reflect the nominal level while maintaining considerable gains in length.

The LR intervals are not reported in Figures 1b,c and 2b,c because the implementation as prescribed by Myers et al. (1994), and used for Table 1, was computationally impractical to apply in these cases. Nevertheless, the cited LR procedure does not suffer from the small sample problem discussed above since the method effectively considers a large continuous range of C values in constructing the intervals. In other words, Myers et al. (1994) invert an exact test to obtain the LR intervals.

Designs involving duplicates, say at each of six fivefold dilutions as in Table 1, are suitable for roughly estimating C. Their main purpose, however, is exploratory,

the results of which are then used to collect more data at and nearby the dilutions indicating a transition from a positive majority to a negative majority. This generally translates into subsequent experiments with relatively fewer dilutions and a higher degree of replication; thus providing more information in a "known" local neighborhood of the true parameter value. The designs underlying the results in Figures 1 and 2, panels d and e, are such examples. Three dilutions (u_d^{-1} = 10, 100, 1000) with five replicates per dilution are considered in Figures 1d and 2d. As in the previous cases, the LR interval coverage tends to hover above the nominal level while the bootstrap coverages hover below; and the BS-BC lengths are distinctly less than the others. Figures 1e and 2e show results based on four dilutions (u_d^{-1} = 10, 50, 250, 1250) each with four replicates. The BC intervals provide good coverage and have relatively shortest length. Although the discrepancy in length between BC and the other intervals in this case is not as substantial as noted in Figures 1d and 2d, it is still distinct and appreciable.

Finally, we report on a simulation experiment, Figures 1f and 2f, following the actual design used in an AIDS application (see below). There are five dilutions (u_d^{-1} = 1, 10, 50, 250, 1250) and the design is unbalanced with replicate numbers 23333 at the respective dilution levels. Knowing that the ML estimate, $\hat{C}$=220, is positively biased, the range of C investigated is likely to include the true concentration associated with the data. The results previously discussed apply here as well. On

balance, the bootstrap BC intervals are optimal in that they yield, as before, the shortest intervals and good coverage.

5 Application: AIDS

The domestic cat provides an important model for AIDS. The feline immunodeficiency virus (FIV) is an immunosuppressive lentivirus in the same family and with marked clinical and biological similarities to the human immunodeficiency virus (HIV). Proviral DNA and other HIV nucleic acids have been shown to increase with the progression of the disease and decrease in response to treatment (Aoki et al., 1990; Clark et al., 1992; Oka et al., 1990; Oka et al., 1991; Pang et al.,1990), thus a test which can reliably quantitate these nucleic acids might be useful to assess therapeutic efficacy or herald disease progression. In this case, a PCR assay which is sensitive to a single copy of DNA is employed with dilutional analysis to quantitate FIV provirus in cat blood (Read et al., 1995). In this section we report on an application of the bootstrap, LR, and standard normal interval methods to data arising from such an assay.

A blood sample from a cat with AIDS was assayed according to a dilution experiment having five dilutions with dilution factors 1 (original concentration), 10, 50, 250, and 1250, and 23333 replicate samples (5μl), respectively. The number of positive results were 23320, respectively. The ML estimate of C is 220.3 molecules per 5μl

or 44,060 molecules per ml of blood. In units of molecules per 5μl, the bootstrap percentile interval is (82.3, 750.4), the bootstrap bias-corrected interval (62.9, 479.7), the truncated normal interval (0, 668.73), and the LR interval (55.7, 762.7). The bootstrap BC interval is based on percentile points α_1=0.0041 and α_2=0.8989 from equations (4) and (5) as opposed to the usual percentile points of 0.025 and 0.975. Note that the bias-corrected interval is considerably shorter and more closely centered on the ML estimate. The former result was expected based on the simulation experiments.

6 Summary

Various methods have been proposed to find confidence intervals for concentration parameters from dilution assays. The most common intervals are based on inverting a hypothesis test and are readily available for relatively small, balanced (same number of replicates per dilution) experiments. In practice it is difficult to adhere to special designs for which intervals are available. For example, loss of data points or sample limitations frequently result in unbalanced data. In these situations, it is necessary to determine intervals specific to the realized design. This may entail a nontrivial implementation of the interval method; perhaps demanding a large amount of memory space or memory space manipulation. We depart from this tradition by proposing

a flexible, easily implemented bootstrap procedure for finding (approximate) confidence intervals for C. Results reported here show that in general the bootstrap bias-corrected intervals provide good coverage with relatively short intervals.

Unlike the traditional intervals, the bootstrap BC intervals explicitly account for the fact that the ML estimate of C is biased. Adjusting for the bias results in shorter intervals. Some care, however, must be taken when dealing with small sample sizes as in Table 1. In this case the resulting bootstrap distribution of the MLE will be highly discrete possibly leading to unreliable results. Thus with small sample sizes it may be worth the effort to implement an alternative method.

References

Aoki, S., Yarchoan, R., Thomas, R.V., Pluda, J.M., Marczyk, K., Broder, S. and Mitsuya, H. (1990). Quantitative analysis of HIV-1 proviral DNA in peripheral blood mononuclear cells from patients with AIDS or ARC: Decrease in proviral DNA content following treatment with 2',3'-dideoxyinosine (ddI). *AIDS Research and Human Retroviruses*, 6:1331-1339.

APHA (American Public Health Association, American Waterworks Association and Federation of Sewage ANd Industrial Wastes Associations) (1976). *Standard Methods for the Examination of Water and Wastewater*, 14th ed. Washington: American Public Health Association.

Blyth, C. R. and Still, H.A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association*, 78:108-116.

Brinchmann, J.E., Albert, J. and Vartdal, F. (1991). Few infected CD4+ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. *Journal of Virology*, 65:2019-23.

Clark, A.G.B., Holodniy, M., Schwartz, D.H., Kazenstein, D.A. and Merigan, T.C. (1992). Decrease in HIV provirus in peripheral blood mononuclear cells during zidovudine and human rIL-2 administration. *Journal of Acquired Immune*

Deficiency Syndromes, 5:52-59.

Cochran, W.G. (1950). Estimation of bacterial densities by means of the 'most probable number.' *Biometrics*, 6:105-116.

deMan, J.C. (1975). The probability of most probable numbers. *European Journal of Applied Microbiology*, 4:307-316.

Does, R.J.M.M., Strijbosch, 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. *The Annals of Statistics*, 7:1-26.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82:171-185.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* New York, Chapman Hall.

Finney, D.J. (1978). *Statistical Method in Biological Assay*, 3rd ed. New York, Macmillan.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Royal Society of the London Philosophical Transactions (Series A)*, 222:309-368.

Kalbfleisch, J.G. (1985). *Probability and Statistical Inference, Volume 2: Statistical Inference.* Springer-Verlag, New York.

Loyer, M. W. and Hamilton, M. A. (1984). Interval estimation of the density of organisms using a serial-dilution experiment. *Biometrics*, 40:907-916.

Marth, E.H. (ed.) (1978). *Standard Methods for the Examination of Dairy Products*, 14th ed. Washington: American Public Health Association.

McGrady, M.H. (1915). The numerical interpretation of fermentation tube results. *Journal of Infectious Diseases*, 17:183-212.

Myers, L.E., McQuay, L.J., and Hollinger, F.B. (1994). Dilution assay statistics. *Journal of Clinical Microbiology*, 32:732-739.

Oka, S., Urayama, K., Hirabayashi, Y., Ohnishi, K., Goto, H., Mitamura, K., Kimura, S. and Shimada, K. (1990). Quantitative analysis of human immunodeficiency virus type-1 DNA in asymptomatic carriers using the polymerase chain reaction. *Biochemical and Biophysical Research Communications*, 167:1-8.

Oka, S., Urayama, K., Hirabayashi, Y., Ohnishi, K., Goto, H., Mitamura, K., Kimura, S. and Shimada, K. (1991). Quantitative estimation of human immunodeficiency virus type-1 provirus in CD4+ T lymphocytes using the polymerase chain reaction. *Molecular and Cellular Probes*, 5:137-142.

Pang, S., Koyanagi, Y., Miles, S., Wiley, C., Vinters, H.V., and Chen, I.S.Y. (1990). High levels of unintegrated HIV-1 DNA in bran tissue of AIDS dementia patients. *Nature*, 343:85-89.

Read, R., Nie, L., Richardson, J., Lue, Y., Basu, S., and Guerra, R. (1995). Dilutional analysis coupled with nested PCR single copy detection for quantitation of retroviral nucleic acids. Manuscript.

Simon, F., Matheron, S., Tamalet, C., Loussert-Ajaka, I., Bartczak, S., Pepin, J.M., Dhiver, C., Gamba, E., Elbim, C., Gastaut, J.A., et al. (1993). Cellular and plasma viral load in patients infected with HIV-2. *AIDS*, 7:1411-7.

Simmonds, P., Balfe, P., Peutherer, J.F., Ludlam, C.A., Bishop, J.O., and Brown, A.J.L. (1990). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of Virology*, 64:864-872.

Sterne, T.E. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, 41:275-278.

Strijbosch, L.W.G., Does, R.J.M.M., and Albers, W. (1990). Multiple-dose design and bias-reducing methods for limiting dilution assays. *Statistica Nerlandica*, 44:241-261.

Woodward, R.L. (1957). How probable is the most probable number? *Journal of*

the American Water Works Association, 49:1060-1068.

Table 1: Coverage and expected widths of confidence intervals based on design with dilution factors $u_d^{-1} = 1, 5, 25, 125, 625, 3125$, and replicates 222222. Interval methods are bootstrap percentile (BS-P), bootstrap bias-correction (BS-BC), likelihood ratio (LR), and standard normal truncated on left at zero. The true concentration is denoted by C.

(a) 95% nominal level

C	BS-P		BS-BC		LR		Normal	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	T-Length
10	91.8	67.09	88.0	33.81	97.2	54.04	98.6	51.59
25	85.8	178.43	89.6	85.25	97.0	140.95	99.6	137.00
50	89.8	367.55	88.2	178.87	96.6	300.79	98.8	277.91
75	95.4	536.62	94.4	252.95	98.4	436.55	97.2	408.32
100	89.8	763.57	84.2	365.84	96.8	603.59	98.8	568.82
125	85.8	935.87	88.8	441.81	96.2	742.88	99.6	697.14
150	95.6	1143.59	91.6	563.92	95.6	898.50	97.2	848.22
175	93.8	1300.73	92.8	620.07	96.0	1032.24	98.2	953.79
200	93.4	1430.84	92.0	658.49	97.4	1146.05	98.4	1037.82
225	92.2	1631.50	85.2	794.13	97.8	1329.94	98.4	1205.43
250	91.2	1714.81	88.0	820.40	97.0	1422.79	98.0	1241.30
500	91.6	3008.38	86.2	1574.85	98.2	2831.90	99.0	2369.30

(b) 90% nominal level

C	BS-P		BS-BC		LR		Normal	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	T-Length
10	90.8	52.92	87.6	28.45	91.6	42.24	95.6	45.56
25	83.2	136.89	89.4	73.05	97.0	106.34	97.0	120.93
50	86.2	289.88	87.6	151.34	89.2	218.47	96.6	245.17
75	80.2	413.15	92.6	214.94	93.0	320.46	97.2	359.94
100	86.8	583.06	84.2	294.27	90.2	454.27	97.4	501.57
125	83.6	722.01	88.8	358.73	96.2	574.01	97.4	614.77
150	81.2	911.40	70.4	450.15	90.2	703.15	95.8	748.60
175	80.6	1005.62	74.8	506.20	90.2	819.05	98.0	842.09
200	75.4	1090.81	77.4	524.49	91.0	924.96	98.4	916.32
225	91.0	1255.05	85.2	626.70	91.0	1078.36	98.2	1066.03
250	88.2	1373.86	88.0	661.04	90.2	1165.46	96.2	1096.79
500	89.6	2571.98	85.6	1245.06	90.8	2337.07	97.2	2105.87

Figure Legends:

Figure 1: Coverage and mean length operating characteristics of various confidence interval procedures at 95% nominal level. Method labels: 1=bootstrap percentile; 2=bootstrap bias-corrected percentile; 3=likelihood ratio; 4=truncated normal. The sample size of each dilution assay design is indicated by the number of replicates per dilution factor. Results based on 500 replications per simulation configuration; true parameter values (C) as shown in Table 1; dilution factors given in text.

Figure 2: Coverage and mean length operating characteristics of various confidence interval procedures at 90% nominal level. Method labels: 1=bootstrap percentile; 2=bootstrap bias-corrected percentile; 3=likelihood ratio; 4=truncated normal. The sample size of each dilution assay design is indicated by the number of replicates per dilution factor. Results based on 500 replications per simulation configuration; true parameter values (C) as shown in Table 1; dilution factors given in text.

Figure 1. Coverage Results from 500 simulations for the Nominal Level of 95%

1 = BS-P; 2 = BS-BC; 3 = LR; 4 = Trunc-Normal

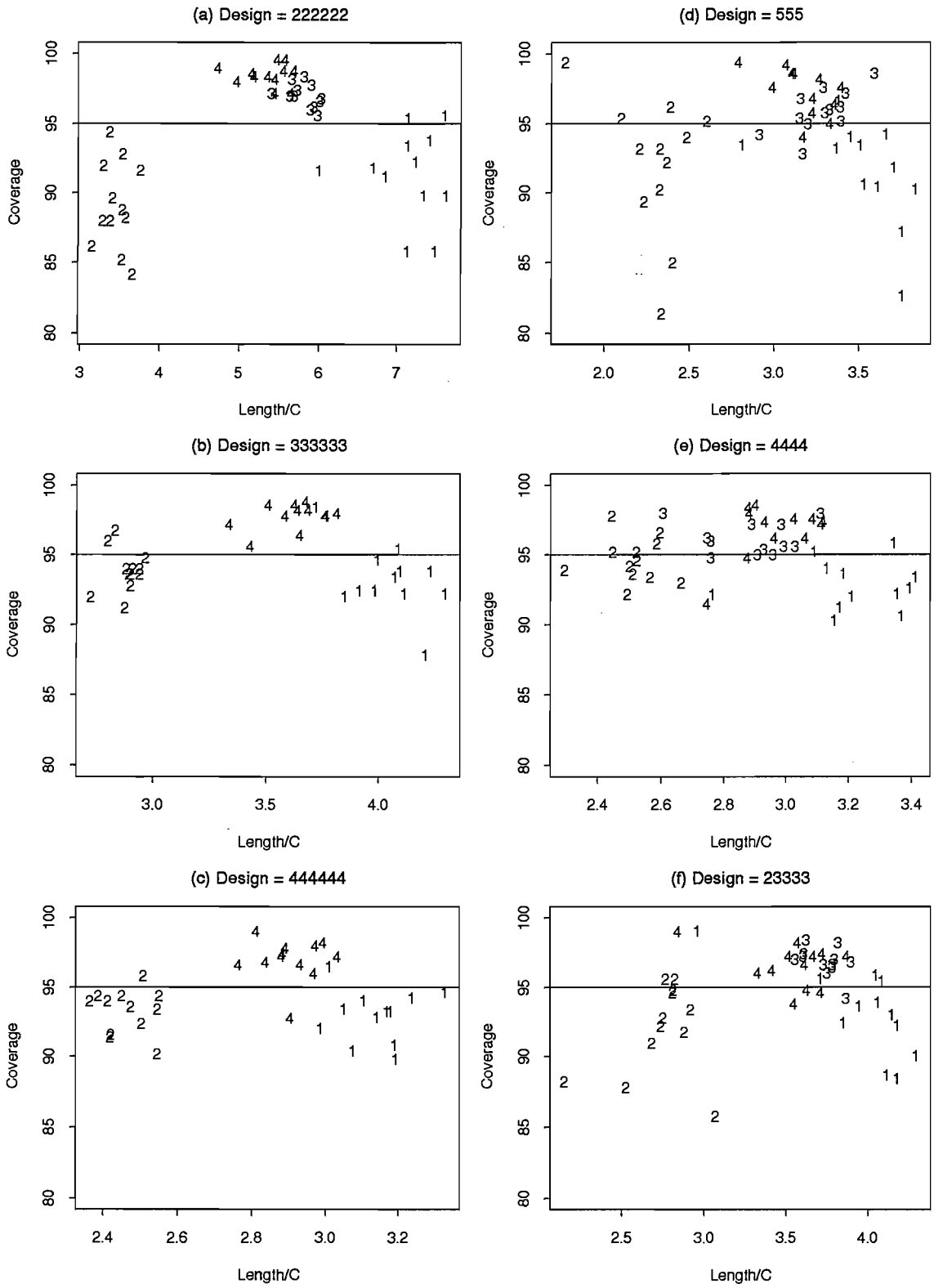

Figure 2. Coverage Results from 500 simulations for the Nominal Level of 90%

1 = BS-P; 2 = BS-BC; 3 = LR; 4 = Trunc-Normal