JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 1994, p. 732–739 Vol. 32, No. 3
0095-1137/94/$04.00+0
Copyright © 1994, American Society for Microbiology

Dilution Assay Statistics

LAWRENCE E. MYERS,[1]* LISA J. McQUAY,[1] AND F. BLAINE HOLLINGER[2]

Research Triangle Institute, Research Triangle Park, North Carolina 27709,[1] and Division of Molecular Virology, Baylor College of Medicine, Houston, Texas 77030[2]

Received 28 July 1993/Returned for modification 28 October 1993/Accepted 13 December 1993

A parametric method of statistical analysis for dilution assays is developed in detail from first principles of probability and statistics. The method is based on a simple product binomial model for the experiment and produces an estimate for the concentration of target entities, a confidence interval for this concentration, and an indicator of the quality of the assay called the *p* value for goodness of fit. The procedure is illustrated with data from a virologic quantitative micrococulture assay used to quantify free human immunodeficiency virus in clinical trials. The merits of the procedure versus those of nonparametric methods of estimating the dilution inducing a 50% response rate are discussed. Advantages of the proposed approach include plausibility of the underlying assumptions, ability to assess plausibility of specific experimental outcomes through their likelihood, and plausibility of confidence intervals.

A dilution assay is an experiment for estimating the concentration or frequency of target entities in a sample, in situations in which accurate counts of the organism are too difficult or costly to obtain. The original sample is divided into subsamples at lower concentrations by dilution. These subsamples may be further sampled to obtain replicate plates, tubes, or wells at each concentration level. Each replicate is then scored for the presence or absence of the target entity. This determination of positivity or negativity of each replicate may require an auxiliary test or procedure. For instance, the presence of bacteria may be deduced from the appearance of colonies after plates have been incubated for a time. In one example that motivated this work, it is desired to estimate the concentration of human immunodeficiency virus type 1 (HIV-1) p24 antigen-producing HIV-infected peripheral blood mononuclear cells (PBMCs) in a blood sample from a patient. Positivity is based on p24 antigen assay results obtained after 2 weeks of incubation. For our purposes, the key features of a dilution assay are that the sample is tested at certain dilutions and that the basic summary of each replicate is dichotomous or binary, with 1 representing a positive result and 0 a negative result.

The idea of a dilution assay is to choose a sufficiently broad range of dilutions that a transition from positive to negative results is virtually ensured as one proceeds through the dilution sequence. The dilutions at which the transition occurs contain information on the concentration of target entities in the original sample. We take estimation of this concentration as the primary purpose of the assay.

Dilution assays are widely used in microbiology, for instance, in the fields of public health (1), virology (9), and immunology (16). A common practical application is to estimate the density of coliform bacteria in water samples. We have encountered several different dilution assays in our work with the National Institute of Allergy and Infectious Diseases AIDS Virology Quality Control Program for virology laboratories serving the AIDS Clinical Trials Group. Specific cases include quantitative micrococulture, plasma viremia, and neutralization assays (9).

The purpose of this report is to describe and illustrate a parametric method of statistical analysis for dilution assays based on a simple probability model for experimental results. The method produces an estimate of the concentration of target entities, a confidence interval for this concentration, and an indicator of the quality of the assay called the *p* value for goodness of fit (PGOF). In describing the method, we briefly review some fundamentals of probability and statistics, including the likelihood function as a basis for inference. We also compare the proposed parametric approach with competing nonparametric methods, such as that of Spearman and Karber and that of Reed and Muench, and indicate why we prefer the proposed approach.

Our statistical approach is developed in Materials and Methods, the first section of which establishes some notation and terminology and describes a virologic quantitative micro-coculture example of dilution assays which is used to quantify free or blood-borne HIV in AIDS patients. The next sections review the notions of sample space, probability, and likelihood; develop the probability model used for inference; and give a detailed description of statistical methods in terms of a probability or likelihood matrix whose entries give the probabilities of experimental results as a function of the different possible values of the true concentration of target entities. All statistical calculations can be viewed as operations on this matrix. Results illustrates the calculations for a simple example and presents and discusses results for selected quantitative micrococulture outcomes.

The merits of our procedure versus those of nonparametric procedures for estimating the dilution inducing a 50% response rate are analyzed in the Discussion. The most credible nonparametric competitor appears from a literature review to be the method of Spearman and Karber. A disadvantage of our proposed parametric approach is its greater computational complexity. Advantages of the proposed approach involve plausibility of underlying assumptions, the ability to assess plausibility of specific experimental outcomes through their likelihood, and plausibility of confidence intervals. Specifically, for the virologic applications mentioned above, patient samples are commonly encountered for which the proportion of positive wells at each dilution assumes only the extreme value of 0 or 1. In such cases, confidence intervals associated with nonparametric procedures can be unreasonable, while confidence intervals produced by our approach are valid.

The exposition in Materials and Methods is detailed and traces the methodology to basic ideas of probability and

* Corresponding author. Mailing address: Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709. Fax: (919) 541-5966.

TABLE 1. Summary of notation for QMC dilution assays

Parameter or variable	Notation
Dilution ratio	$DR = 0.2$
No. of dilution levels	$D = 6$
Dilution level index	$d = 1, 2, \ldots, D$
No. of replicate wells tested at level d	$n_d = 2$
Replicate well index	$j = 1, 2, \ldots, n_d$
Indicator variable of +/− outcome, jth well, level d	$w_{dj} = 1$ for positive, 0 for negative
No. of positive wells at level d	$x_d = w_{d1} + w_{d2} = 0, 1,$ or 2
Generic outcome	$y = (x_1, x_2, \ldots, x_D)$
Proportion of positive wells at level d	$x_d/n_d = 0, 0.5,$ or 1.0
Expected no. of PBMC/well at level d	$u_d = 10^6 \cdot DR^{d-1}$
Concn of IUs/PBMC	C
Concn as IUs/10^6 PBMC	$\text{IUPM} = 10^6 \cdot C$
Expected no. of IUs/well at level d	$\text{IU}_d = C \cdot u_d$
Probability that a given well at level d is negative	$q_d \simeq \exp(-\text{IU}_d) = \exp(-C \cdot u_d)$
Probability that a given well at level d is positive	$p_d = 1 - q_d \simeq 1 - \exp(-C \cdot u_d)$

statistics. For the most part, this report is logically self-contained. The Appendix contains a glossary of terms.

MATERIALS AND METHODS

Description and terminology for QMC dilution assays. The proposed method for statistical analysis of dilution assay data will be illustrated with a virologic quantitative micrococulture (QMC) assay used to quantify free or blood-borne HIV in patients participating in multicenter HIV clinical trials. Table 1 summarizes the main features of experimental design and notation for the QMC dilution assay. Table 1 indicates general symbols for dilution assays and the specific numeric values assumed for the QMC assay. A more detailed description of the QMC assay is in reference 9.

The QMC assay is performed, in duplicate, in a 24-well tissue culture plate by using six fivefold dilutions, beginning with 10^6 patient PBMC per well at the first dilution level, 2×10^5 PBMC per well at the second level, and so on. Viewing the 24-well plate as an array with four horizontal rows of six wells and six vertical columns of four wells, each plate can accommodate two patient samples, allocating the top two rows to one patient sample and the bottom two rows to the other patient sample. Undiluted specimen is placed in the first column on the left, 1:5 dilutions are in the second column, and so on, with 1:3125 dilutions in the last column on the right of the plate. Each sample of patient cells is cocultured with phytohemagglutinin-stimulated normal donor PBMC for 14 days. Supernatant from each well is then assayed for viral expression of HIV-1 p24 antigen with standard HIV p24 enzyme immunoassays.

A key concept is the infectious unit (IU), an aggregate of one or more infected cells whose presence in a well is necessary and sufficient to produce a positive HIV-1 p24 result. The target entity for QMC assays is the IU. T denotes the number of these harbored by the patient, C denotes the concentration of these per individual patient PBMC, and IUPM denotes the concentration of these as IUs per million patient PBMC ($\text{IUPM} = 10^6 \cdot C$). The purpose of the QMC assay is to estimate the concentration (C) of IUs per PBMC or, equivalently, the concentration (IUPM) of IUs per 10^6 patient cells. C and IUPM are examples of parameters, i.e., unknown constants which affect the probability distribution of experimental results. The critical assumptions which underlie the mathematical model used below are that infectious units are distributed purely at random within each sample, that detectable growth will occur in each well containing one or more IUs, and that dilution errors are negligible.

Fundamental notions of probability and statistics. (i) The sample space. The set of all possible outcomes for an experiment is called the sample space. A subset of the sample space is called an event. The sample space for a dilution assay is a finite set which may be constructed as follows. For the QMC assay, we may define 12 binary indicators to summarize experimental outcomes for the 12 wells. If $d = 1$ to 6 indexes the six dilution levels and $j = 1$ or 2 indexes the duplicate wells, then define w_{dj} to be 1 or 0, according as the jth well at the dth level is positive or negative for the p24 antigen. To connect this with the geography of a 24-well plate, the w_{dj} variables associated with the patient specimen allocated to the top two rows of the plate would be configured as follows:

	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
$j = 1$	w_{11}	w_{21}	w_{31}	w_{41}	w_{51}	w_{61}
$j = 2$	w_{12}	w_{22}	w_{32}	w_{42}	w_{52}	w_{62}

Each of the 12 wells can conceivably be either positive or negative, so there are 2^{12} (=4,096) different possible experimental outcomes. The size of the sample space can be reduced by observing that the experiment is symmetric with respect to the duplicate wells at any given dilution level. The only relevant information at a given dilution level is the number of positive wells, regardless of where they occur. The number (x_d) of positive wells at the dth dilution level is the sum of the binary indicators at that level, e.g., $x_1 = w_{11} + w_{12}, x_2 = w_{21} + w_{22}$, and so on. In this way, the QMC sample space can be viewed as the set of all ternary sextuples $(x_1, x_2, x_3, x_4, x_5, x_6)$, where x_d is the number of positive wells among the duplicates at the dth dilution level ($x_d = 0, 1,$ or 2). The reduced sample space for the QMC assay contains 3^6 (=729) possible outcomes. Specific numeric outcomes will be written without parentheses and commas. For instance, 210000 indicates that both wells are positive at the first level, one of two wells is positive at the second level, and all wells are negative at lower levels.

(ii) Probability. Three properties are sufficient to define a probability function (p) on a finite sample space: P1, the probability [$p(A)$] of any event (A) is between 0 and 1; P2, the probability of the whole sample space is 1; P3, if A and B are events with no outcomes in common, then the probability that either event occurs is the sum of their individual probabilities [$p(A \text{ or } B) = p(A) + p(B)$]. Events A and B with no outcomes

in common are called mutually exclusive; the occurrence of either event precludes the occurrence of the other event.

P3 is the addition rule for probability and extends in an obvious way to a finite number of mutually exclusive events, such as distinct individual outcomes.

The multiplication rule of probability is related to the notion of statistical independence. Events A and B are said to be independent if

$$p\,(A \text{ and } B) = p\,(A) \cdot p\,(B). \quad (1)$$

(iii) Likelihood. To apply probability models to experimental data, it is useful to extend the probability notation to explicitly display dependence on a parameter such as the concentration of target entities. Let $p(A|c)$ [or $p(y|c)$] denote the probability of event A [or outcome (y)], assuming that c is the true concentration of target entities. (Strictly speaking, we should write $p(\{y\}|c)$ instead of $p(y|c)$, since probability is a function of events and events are sets.} A mathematical model for $p(y|c)$ is derived from equation 1 and P3 in the section on the probability model below.

Two distinct terms, probability and likelihood, can be applied to the function $p(y|c)$, depending on the situation and perspective. In common parlance, these terms are synonymous, but in the domains of probability and statistics, an important distinction exists between them. If the state of nature or true concentration (c) is fixed and the outcome (y) is viewed as varying over the sample space, then $p(y|c)$ is called probability. For the statistical perspective, at hand is a particular experimental outcome (y), and we want to draw inferences about the true concentration of target entities. Following R. A. Fisher (6), attention is directed to the likelihood function, which is the same function, $p(y|c)$, but with outcome y fixed and parameter c varying.

Probability model. The purpose of this section is to develop a simple product binomial model for the likelihood, summarized in equations 3 and 4 below, used to calculate the probability of any dilution assay outcome for a given design and a given value of C (or IUPM). The development is an elaboration of Cochran (2) and uses only elementary probability.

(i) Coin tossing. Suppose we have a possibly biased coin with probability p of falling heads and probability $q = 1 - p$ of falling tails. For any positive integer n, let $n!$ equal the n-fold product of all integers from n down to 1 [$n! = n \cdot (n - 1) \cdot (n - 2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$] and define 0! as 1. Then the probability of exactly k heads in n independent tosses of the coin is the binomial probability

$$p\,(k \text{ heads in } n \text{ tosses}) = \{n!\,/\,[k!\,(n - k)!]\} \cdot p^k \cdot q^{n-k} \quad (2)$$

The derivation is straightforward if it is assumed that the tosses are independent and equation 1 and P3 are used. It follows from equation 1 that any particular sequence of n tosses with exactly k heads and $n - k$ tails has probability $p^k \cdot q^{n-k}$, and the factorial term in braces is simply the number of such sequences.

(ii) Probability model for a single dilution. Consider a dilution assay with n replicates at a single dilution. Let f denote the fraction of the patient's total PBMC population which is tested in each replicate. If the patient harbors T target entities, then the probability that a given target entity is not in a given replicate is $1 - f$ and the probability that none of the T target entities is in a given replicate is $q = (1 - \text{f})^T$, according to equation 1. The probability of a negative replicate is q.

The following approximation is critical and implies that we do not need to know f. By the first-order Taylor approximation $\exp(-f) \simeq 1 - f$, for f near zero, q is closely approximated by $\exp(-f \cdot T)$, which can, in turn, be rewritten as $\exp(-C \cdot u)$, where C is the patient's concentration of target entities per PBMC and u is the number of PBMC per replicate.

By analogy with coin tossing, associating heads with positive and tails with negative, the probability that exactly k of the n replicates are positive and $n - k$ are negative is given by equation 2 with $q = 1 - p = \exp(-C \cdot u)$.

(iii) Probability model for a general dilution assay. By the preceding argument, the probability that no target entities end up in a given well at the dth dilution level is approximately

$$q_d = \exp(-C \cdot u_d) \quad (3)$$

Thus, q_d is the probability that a given well at the dth dilution level is negative, i.e., devoid of IUs and p24 antigen. Assuming further that all wells are independent, equation 1 implies that the overall likelihood for the experiment is the product (Π) over the six dilutions of binomial probabilities:

$$L = \prod_{d=1}^{D} \{n_d\,!\,/\,[x_d\,!\,(n_d - x_d)\,!\,]\} \cdot p_d^{x_d} \cdot q_d^{n_d - x_d} \quad (4)$$

$q_d = 1 - p_d$ is given in equation 3. The probability model specified by equations 3 and 4 will be called the simple product binomial model. An example calculation of the likelihood is contained in Results.

Statistics. (i) Reduction to finite parameter space. The possible values for IUPM range over a virtual continuum from 0 to 10^6. To simplify the discussion of statistical calculations, we will treat the parameter space as finite and consisting of 0 and the 2,779 values obtained by taking integral powers of 1.01 from $-1{,}389$ to 1,389. This finite parameter set of 2,780 values includes 0, 0.000000995, . . . , 1,005,514. The positive parameter values range approximately from 10^{-6} to 10^6, with each succeeding term 1.01 times larger than its predecessor. The loss of accuracy entailed by this reduction of the parameter space is negligible.

The reason for reformulating the problem in terms of a finite parameter space is that this permits a conceptually simple and direct approach in terms of certain explicit vectors and matrices. These vector and matrix elements of the problem, and relationships among them, are indicated in Table 2.

For any outcome $y = (x_1, x_2, x_3, x_4, x_5, x_6)$, we want (i) an estimate of the IUPM in the original sample, (ii) a 95% confidence interval for IUPM, and (iii) an indicator of the quality of the data $(x_1, x_2, x_3, x_4, x_5, x_6)$. We will approach these problems in the order of their complexity: i, iii, and ii.

(ii) Probability or likelihood matrix. The probability (or likelihood) matrix (P) is central to the statistical computation. In our formulation of the problem, both the sample space and the parameter space are finite sets. The QMC sample space contains 729 outcomes, and the reduced parameter space has 2,780 values. The probability matrix has 729 rows corresponding to the outcomes and 2,780 columns corresponding to the parameter values. It has 2,026,620 entries.

The two correspondences, between rows and outcomes and between columns and parameter values, are depicted in Table 2. The parameter values, in increasing order, appear as column labels. To obtain an ordering for the 729 outcomes, each sextuple can be viewed as a real number, beginning with 000000, 000001, . . . , and ending with 222221, 222222. With these associations, the outcome corresponding to row i will be denoted y_i and the parameter value corresponding to column j will be designated c_j. Hence, for the QMC example with six

TABLE 2. Key vectors and matrices for likelihood-based statistics with finite sample and parameter spaces

sample space outcomes y_i	PROBABILITY (LIKELIHOOD) MATRIX $P = [P[i,j]] = [p(y_i \mid c_j)]$ parameter values c_1	c_2	. . .	c_n	MLE $\hat{c}_i$	MAXIMIZED LIKELIHOOD LMAX[i]= $p(y_i \mid \hat{c}_i)$
y_1	$p(y_1 \mid c_1)$	$p(y_1 \mid c_2)$	. . .	$p(y_1 \mid c_n)$	$\hat{c}_1$	$p(y_1 \mid \hat{c}_1)$
y_2	$p(y_2 \mid c_1)$	$p(y_2 \mid c_2)$	. . .	$p(y_2 \mid c_n)$	$\hat{c}_2$	$p(y_2 \mid \hat{c}_2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_m	$p(y_m \mid c_1)$	$p(y_m \mid c_2)$	. . .	$p(y_m \mid c_n)$	$\hat{c}_m$	$p(y_m \mid \hat{c}_m)$

sample space outcomes	RELATIVE LIKELIHOOD MATRIX $R = [R[i,j]] = [p(y_i \mid c_j)/p(y_i \mid \hat{c}_i)]$			
y_1	$\frac{p(y_1 \mid c_1)}{p(y_1 \mid \hat{c}_1)}$	$\frac{p(y_1 \mid c_2)}{p(y_1 \mid \hat{c}_1)}$	. . .	$\frac{p(y_1 \mid c_n)}{p(y_1 \mid \hat{c}_1)}$
y_2	$\frac{p(y_2 \mid c_1)}{p(y_2 \mid \hat{c}_2)}$	$\frac{p(y_2 \mid c_2)}{p(y_2 \mid \hat{c}_2)}$	. . .	$\frac{p(y_2 \mid c_n)}{p(y_2 \mid \hat{c}_2)}$
⋮	⋮	⋮	⋮	⋮
y_m	$\frac{p(y_m \mid c_1)}{p(y_m \mid \hat{c}_m)}$	$\frac{p(y_m \mid c_2)}{p(y_m \mid \hat{c}_m)}$	. . .	$\frac{p(y_m \mid c_n)}{p(y_m \mid \hat{c}_m)}$

fivefold dilutions in duplicate, $m = 729$, $n = 2{,}780$, $y_1 = 000000$, $y_2 = 000001, \ldots, y_{728} = 222221, y_{729} = 222222$, $c_1 = 0$, $c_2 = 0.000000995, \ldots, c_{2,780} = 1{,}005{,}514$. For a specified row i (outcome y_i) and column j (parameter c_j), the entry $P[i, j]$ is the probability $[p(y_i|c_j)]$ of outcome y_i, assuming $C = c_j$, computed with equations 3 and 4. Any particular column specifies a distribution of probability over the sample space. Therefore, any column of the matrix adds to unity.

(iii) Estimation of C and IUPM by the method of maximum likelihood. The dominant method of statistical estimation is the method of maximum likelihood (11). For a given experimental outcome, the maximum-likelihood estimate (MLE) is the parameter value which maximizes the probability of obtaining the outcome. Operationally, relative to the probability matrix of Table 2, the outcome y_i uniquely identifies row i of the matrix. The ith row is searched to identify the column where the maximum likelihood occurs. The index of this column is designated $j(i)$. The corresponding parameter value $c_{j(i)}$ is the MLE of C and is designated $\hat{c}_i$. The value $p(y_i|\hat{c}_i)$ of the likelihood at this maximum is called LMAX[i]. The MLE of IUPM is $10^6 \cdot \hat{c}_i$.

(iv) P value for goodness of fit (PGOF). A large majority of the 729 outcomes are in fact very unlikely to occur, regardless of the value of C. To quantify this and to obtain an indicator of the data quality for a particular assay, a quantity called PGOF is used. The intent of PGOF is to identify rare or implausible experimental outcomes which may indicate procedural problems with the assay. Examples of such patterns include 111111 (222000 would be much more likely) and large "skips," such as 220002.

The circumstance most favorable to the occurrence of y_i is that $\hat{c}_i$ is the true concentration. PGOF is the probability of an experimental result as rare as or rarer than that obtained, assuming that the model is correct and the parameter C is equal to its MLE. Given an outcome y_i, attention is again directed to column $j(i)$ of the probability matrix of Table 2, for which this outcome has maximum likelihood LMAX[i]. The probabilities in this column sum to unity. By P3, PGOF[i] is the sum of all probabilities in the column which do not exceed LMAX[i]. PGOF[i] is usually less than 1 and equals 1 only if y_i is the most probable outcome for $\hat{c}_i$. Low values of PGOF, e.g., PGOF < 0.01, indicate rare or implausible experimental results. If possible, any sample with a very low PGOF should be retested.

(v) Confidence intervals for C and IUPM. There is a duality between hypothesis testing and confidence intervals, whereby a point c_j is included in a 95% confidence interval for C just in case the null hypothesis H, that $C = c_j$, is not rejected at the 5% significance level. We use this duality to get a 95% confidence interval for C. Just as the method of maximum likelihood is the dominant method of statistical estimation, the preferred method of testing hypotheses uses likelihood ratio tests (LRTs) (reference 10, p. 48). The recipe described below amounts to construction of confidence intervals for C via exact LRTs. The appended glossary reviews the necessary hypothesis testing concepts.

Construction of confidence intervals for C involves relative likelihood matrix R (bottom of Table 2), as well as likelihood matrix P. The relative likelihood is the likelihood normalized relative to its maximum achievable value (11). R is obtained from probability matrix P and the column vector LMAX of maximized likelihoods via $R[i, j] = P[i, j]/\text{LMAX}[i]$, i.e., each row of the probability matrix is divided by the row maximum. Note that $R[i, j] \leq 1$ and $R[i, j(i)] = 1$.

Given an outcome of y_i, to test H ($C = c_j$) at the 0.05 level with an LRT, we assume that H is true and calculate the LRT p value, which is the probability of an outcome at least as implausible as y_i, as determined by the relative likelihood function. To put it another way, this LRT p value is the probability of an outcome which is relatively not more likely than that obtained. The p value calculation employs the jth columns, $P[\,.\,, j]$ and $R[\,.\,, j]$, of both likelihood matrix P and relative likelihood matrix R. The column of the relative likelihood matrix indicates which entries of $P[\,.\,, j]$ are summed. Specifically, by P3, the LRT p value is the sum of $P[k, j]$ over all rows k (outcomes y_k) for which $R[k, j]$ does not exceed $R[i, j]$. If this p value exceeds 0.05, then H is accepted at the 5% level and c_j is included in the 95% confidence interval.

Historical notes. Sir Ronald Aylmer Fisher influenced the practice of statistics more than any other individual. Fisher's 1922 treatise (6) was a pathbreaking work which defined and investigated several important concepts, including likelihood, MLEs, and sufficiency. The paper also includes an interesting dilution assay example using the simple product binomial model. Essentially the same model was used by Greenwood and Yule (7), Eisenhart and Wilson (4), Finney (5), and Strijbosch et al. (16). To a great extent, Fisher's 1922 paper (and the modern practice of parametric statistics) can be summarized with the statement that likelihood is the basis of inference. The text by Kalbfleisch (11) is in this spirit, has been acclaimed for the clarity of its presentation (14), uses many biologic examples, and is highly recommended to interested readers.

RESULTS

Example with singleton wells undiluted and diluted fivefold. We illustrate the calculation of statistical quantities with a small artificial example involving singleton wells, undiluted and diluted fivefold. The possible outcomes are 00 (both wells negative), 01 (undiluted well negative, diluted well positive), 10 (undiluted well positive, diluted well negative), and 11 (both wells positive). The parameter space will be taken to be {0, 0.04, 0.2, 1, 5, 25}, i.e., we assume that it is somehow known in advance that the true concentration must be one of these six values. This situation is not intended to correspond to any real experiment but is chosen for simplicity of computation. The relevant matrices are in Table 3.

TABLE 3. Likelihood matrix, relative likelihood matrix, p values for LRTs of H ($C = c_j$), and binary indicators of 95% confidence sets for dilution assay example

Outcome	Likelihood matrix with parameter space of:						MLE	LMAX	PGOF	Relative likelihood matrix with parameter space of:						p values for LRT of H ($C = c_j$) with parameter space of:						95% Confidence sets
	0	0.04	0.2	1	5	25				0	0.04	0.2	1	5	25	0	0.04	0.2	1	5	25	
00	1.000	0.953	0.787	0.301	0.002	0.000	0.0	1.000	1.000	1.000	0.953	0.787	0.301	0.002	0.000	1.000	1.000	1.000	0.416	0.002	0.000	{0, 0.04, 0.2, 1}
01	0.000	0.008	0.032	0.067	0.004	0.000	1.0	0.067	0.067	0.000	0.115	0.481	1.000	0.064	0.000	0.000	0.047	0.213	1.000	0.007	0.000	{0.2, 1}
10	0.000	0.039	0.174	0.518	0.365	0.007	1.0	0.518	1.000	0.000	0.075	0.337	1.000	0.706	0.013	0.000	0.039	0.181	1.000	1.000	0.007	{0.2, 1, 5}
11	0.000	0.000	0.007	0.115	0.628	0.993	25	1.000	1.000	0.000	0.000	0.007	0.116	0.632	1.000	0.000	0.000	0.007	0.115	0.635	1.000	{1, 5, 25}

The calculation of likelihood will be illustrated by verifying that the entry in the second row and third column of the likelihood matrix in Table 3 is 0.032, i.e., that the probability of outcome 01 is 0.032 when IUPM is 0.2. We have $n_1 = n_2 = 1$ (singleton wells), $x_1 = 0, x_2 = 1$, so the likelihood reduces to $q_1 \cdot p_2$. Also, $q_1 = \exp(-C \cdot u_1) = \exp(-0.2 \cdot 1) = 0.81873$ and $p_2 = 1 - q_2 = 1 - \exp(-0.2 \cdot 0.2) = 1 - \exp(-0.04) = 0.03921$, so that $q_1 \cdot p_2 = 0.032$. The other entries of the likelihood matrix are calculated similarly.

For each of the four outcomes, the MLE of IUPM is determined by the column which maximizes the likelihood over the row corresponding to the outcome. For instance, for outcome 01 (second row), the maximum likelihood of 0.067 occurs in the fourth column, so that the MLE is IUPM = 1 and LMAX = 0.067. PGOF is also 0.067, because there are no other outcomes which are less likely.

The relative likelihood is calculated separately for each row as the likelihood relative to (divided by) its maximum achievable value. For instance, the relative likelihood of outcome 01 is obtained by dividing each likelihood value in the second row by 0.067.

Next we verify that the LRT p value is 0.213 for testing hypothesis H (IUPM = 0.2) when the experimental result is 01. Assume that H is true, so IUPM equals 0.2. The relative likelihood (R) equals 0.481 when the IUPM is 0.2 and the outcome is 01. Two other outcomes are relatively less likely: 10 ($R = 0.337$) and 11 ($R = 0.007$). Therefore, the LRT p value, i.e., the probability (P) of outcomes which are not relatively more likely than 01, is $0.032 + 0.174 + 0.007 = 0.213$. The parameter value $C = 0.2$ is therefore included in the confidence set for the outcome 01, since the p value 0.213 exceeds 0.05. The confidence sets for each outcome are indicated at the bottom right of Table 3. (The term confidence set is used here rather than the term confidence interval because of the artificial nature of the example, whereby we have assumed that it was known in advance that the true IUPM must be one of the six values {0, 0.04, 0.2, 1, 5, 25}.)

Example with six fivefold dilutions in duplicate. Table 4 contains selected results for the QMC dilution assay design with six fivefold dilutions in duplicate which is used to quantify HIV burden in patients participating in National Institute of Allergy and Infectious Diseases HIV clinical trials. The selected results are from proficiency testing conducted within the AIDS Virology Quality Control Program. The table contains results for all 22 outcomes with PGOFs of at least 0.2, as well as all outcomes which have occurred with PGOFs below 0.001. Confidence intervals are not given for the outcomes with PGOFs below 0.001, because these should be retested.

The outcome 222222, with all wells positive, is censored, in that if the dilution series were extended further, negative wells would eventually occur. To obtain a finite concentration estimate for such a censored case, a common approach is to postulate duplicate negative wells at the hypothetical next dilution level, i.e., 2222220. The corresponding IUPM estimate is 5,608. If such an approach is taken, it is important to distinguish (flag) such censored cases in a data base. Ideally, if resources permit, the sample should be retested with a broader design, so that results can be based on real rather than hypothesized data. If censored outcomes occur too frequently with a fixed design for routine testing, then the design should be changed.

The last outcome listed (101222) is virtually impossible for all values of IUPM. It turned out that this outcome resulted-from reversing the order of dilutions on the tray, so the laboratory software interpreted the highest dilution as the

TABLE 4. Likely (PGOF > 0.2) and unlikely (PGOF < 0.001) outcomes from National Institute of Allergy and Infectious Diseases AIDS Virology Quality Control QMC Proficiency Testing Program[a]

Outcome	MLE of IUPM	PGOF	LO_{95}[b]	UP_{95}[c]
000000	0.000	1.000000	0.000	1.220
100000	0.508	1.000000	0.026	2.732
200000	1.612	1.000000	0.248	7.031
210000	3.235	1.000000	0.639	13.831
220000	8.081	1.000000	1.138	35.241
221000	16.218	1.000000	2.759	70.021
222000	40.509	1.000000	7.101	180.201
222100	82.105	1.000000	13.969	365.237
222200	205.086	1.000000	35.594	1,059.16
222210	419.830	1.000000	70.721	1,640.98
222220	1,124.32	1.000000	182.003	4,711.59
222221	2,492.30	1.000000	368.889	11,422.7
222222[d]	∞[d]	1.000000	1,017.83	∞[d]
211000	5.648	0.362795	1.041	17.737
221100	28.313	0.360267	6.493	88.908
110000	1.105	0.360197	0.190	3.538
222110	143.339	0.347318	32.545	463.754
222211	747.676	0.277522	164.765	2,125.48
201000	2.815	0.254773	0.596	9.016
220100	14.109	0.253046	2.678	45.195
222010	70.721	0.245868	13.558	231.095
222201	361.621	0.206533	68.641	1,371.88
222122	423.141	0.000964		
020100	1.312	0.000932		
012000	1.274	0.000882		
011100	1.265	0.000692		
120001	2.511	0.000681		
112100	3.084	0.000571		
001010	0.807	0.000274		
221002	40.416	0.000262		
022000	1.780	0.000248		
202110	8.843	0.000177		
111110	3.043	0.000095		
100101	1.607	0.000087		
021010	1.764	0.000040		
202101	8.830	0.000030		
111101	3.042	0.000022		
021001	1.764	0.000011		
022100	2.232	0.000010		
122200	4.956	0.000007		
221122	95.294	0.000002		
101222	4.770	0.000000		

[a] QMC assay design involves duplicate wells at each of six fivefold dilutions.
[b] LO_{95}, lower 95% confidence limit.
[c] UP_{95}, upper 95% confidence limit.
[d] Censored outcome. Both the MLE and upper confidence limit are infinite. See text.

lowest, the second highest as the second lowest, and so on. That is, the actual experimental outcome was 222101.

This illustrates how an explicit probability model can be used to advantage to identify unlikely outcomes for closer inspection and possible retesting.

In addition to PGOF, another natural indicator of assay quality might involve precision as expressed by the confidence interval width or the ratio of upper to lower confidence limits. In fact, the probability distribution of the MLE $\hat{c}$ is known to be skewed, while the distribution of the logarithm of $\hat{c}$ is much more symmetric and closer to normal (2). This suggests that the difference between the logarithms of the confidence interval endpoints, or, equivalently, the ratio of the endpoints, is a reasonable summary of precision.

These two aspects of quality, PGOF and the ratio of confidence interval endpoints, are actually quite different. Outcomes which are rarer in PGOF terms can be more informative in confidence interval terms. This is illustrated by comparison of the results in Table 4 for outcomes 210000 and 201000, which have respective PGOFs of 1.0 and 0.255 and respective confidence interval ratios of 21.6 and 15.1. This should not be interpreted as a commentary on the validity of the PGOF as an indicator of assay quality, and laboratories should not aspire to produce outcomes with low PGOFs.

DISCUSSION

The methodology presented here is applicable to any dilution assay producing binary (+/− or 1/0) data at the replicate level. We have illustrated the method in terms of a virologic quantitative micrococulture assay whose purpose is to quantify HIV in infected patients as IUPM. Application to other assays is straightforward, as long as they qualify as dilution assays with dichotomous outcomes. For instance, the plasma viremia assay (9) is similar to the QMC assay, employing duplicate wells at each of six fivefold dilutions, except that a fixed volume of patient plasma is input to the replicates at each level (0.4 ml at the first dilution level, 0.08 ml at the second level, 0.016 ml at the third level, and so forth.) In this case, u_d is defined as the volume of plasma at the dth level and the concentration (C) of target entities is defined relative to a unit volume of plasma equal to the volume input to a well at the first dilution level, i.e., 0.4 ml. The estimated concentration (C) will then be the number of IUs per 0.4 ml, which is multiplied by 2.5 to obtain the concentration of IUs per milliliter. Neutralization assays are treated similarly.

Competitors to the parametric method of analysis based on the simple product binomial model include nonparametric procedures going under the names of Spearman and Karber, Reed and Muench, Dragstedt and Behrens, Litchfield and Wilcoxon and moving average methods (5, 8). These methods might be applied to dilution assays to estimate the ED_{50}, i.e., the dilution at which 50% of the replicates would be expected to be positive. Regarding the first three methods, in his classic work, Finney (reference 5, p. 394) opines the following. "The time has come for a change: under the most favorable conditions, neither [the Reed-Muench nor the Dragstedt-Behrens] method is as precise as the Spearman-Karber, and often Spearman-Karber is markedly superior. . . . Except as part of statistical history, both methods should be forgotten."

This opinion is echoed by Hamilton (reference 8, p. 75): "Both large sample and small sample evaluations of these estimators have clearly established the superiority of the Spearman-Karber procedure. Apparently the continued use of the Reed-Muench and/or Dragstedt-Behrens methods in some areas . . . is motivated only by tradition." Hamilton (reference 8, p. 70) also indicates that the Litchfield-Wilcoxon method is outdated because the investigator fits regression lines "by eye."

Finney (reference 5, p. 435) also asserts that the underlying assumptions are more realistic for the simple product binomial model than for the nonparametric approaches: "The reader is warned against attempting to use any analogue of the Dragstedt-Behrens, Reed-Muench or moving average method for dilution assays. The exponential formula for the probability of a sterile plate is not symmetric about any point, and these methods are even less appropriate than with normal or logistic sigmoid response curves."

It thus appears that the only serious competitor among these nonparametric methods of estimating the ED_{50} is the Spearman-Karber estimate. If the equation (3) for the probability of a positive response is accepted as a reasonable approximation, then the ED_{50} is that number of PBMC per well such that exp

$(-C \cdot ED_{50}) = 0.5$, i.e., such that $C \cdot ED_{50} = 0.69$. This relationship can be solved for either parameter in terms of the other. If $\hat{c}$ is the MLE of C, then $0.69/\hat{c}$ is the MLE of the ED_{50} and conversely. Hence, the two parameters, C and ED_{50}, are equivalent in that either can be obtained from the other, if equation 4 is accepted. Although the ED_{50} and IUPM are, in a sense, interchangeable, the IUPM seems inherently more clinically relevant than the dilution at which 50% of the replicates wells are expected to be negative.

We prefer the parametric approach based on the simple product binomial model, essentially because it supports the calculation of auxiliary quantities such as the PGOF and the confidence interval for IUPM. The use of the probability model to identify implausible experimental outcomes as cases with low PGOFs is an important benefit of the parametric approach. This assists laboratories in identifying individual problem assays or procedural problems.

Three key assumptions underlying the simple product binomial model were identified at the end of the QMC dilution assay section in Materials and Methods. Like all mathematical models, the simple product binomial model is an idealization which does not exactly correspond to any real experiment. However, the assumptions that mixing is homogeneous and dilution errors are negligible are undeniably worthwhile goals for laboratorians. We believe it is useful to subject dilution assays to the PGOF criterion to detect assays that deviate severely from these ideals.

In preparing the dilution series and replicate wells, it is important to mix each sample adequately before subsampling from it. Inadequate mixing prior to pipetting can allow settling of cells to the bottom. This can induce a bias in subsampling, depending on where the pipette is inserted in the sample, and might produce a result such as 002000. This outcome has a PGOF of 0.001791. The sample should be retested if possible. For a laboratory using proper technique, 5% of the assays would be expected to have PGOFs below 0.05. A laboratory with an appreciably higher percentage of PGOFs below 0.05 should critically examine its procedures. A reasonable approach would be to monitor the percentage of assays with PGOFs below 0.05 and periodically test the "in control" hypothesis that PGOF is 0.05 or less. If the hypothesis is rejected and the empirical proportion of PGOFs below 0.05 is 0.1 or larger, then search for an assignable cause.

In many cases, a confidence interval for the key parameter is needed. This need provides an important justification for the use of our parametric procedures in preference to the Spearman-Karber method. The Spearman-Karber standard error of the ED_{50} is estimated to be 0 for any dilution assay outcome with all empirical proportions equal to 0 or 1 (reference 8, p. 75). For instance, for the QMC example used in this report, with duplicate wells at each of six fivefold dilutions, the following outcomes would have 0 standard error: 200000, 220000, 222000, 222200, 222220, 000222, 020202. The textbook Spearman-Karber confidence interval in these cases would have a width of 0 for any level of confidence, which is absurd. ("Confidence intervals associated with the nonparametric procedures are probably not reliable for such coarse data" [reference 8, p. 84].) The confidence intervals produced by our parametric approach are reasonable in all cases.

The method of summarizing dilution assays in terms of the endpoint, i.e., the lowest level at which a positive result is obtained, is imprecise and discards information. For instance, it does not distinguish between QMC outcomes 221000 (IUPM = 16.25, PGOF = 1.0) and 001000 (IUPM = 0.403, PGOF = 0.045).

It should be noted that the MLE of the concentration of target entities is rather biased, systematically overestimating the true concentration by a multiplicative factor as large as 1.5. Roughly speaking, this arises because the logarithm of the MLE of the concentration estimate has a fairly symmetric and unbiased distribution for the true log concentration, so that exponentiation induces asymmetry and bias. Does et al. (3) have examined this bias issue in detail and recommend a jackknife estimate of C, whereby the replicates are deleted one at a time to generate n pseudosamples of size $n - 1$, where n is the total number of experimental units in the original assay. The MLE of C is calculated for each pseudosample, and these n estimates are averaged to form the jackknifed estimate of C. Computer programs for this calculation are available from R. J. M. M. Does.

Loyer and Hamilton (12) compared several techniques of confidence interval construction and recommended a technique of Sterne (15). They did not consider our method, which is based on inversion of the LRT. We have conducted a study (13) of several techniques of confidence interval construction for dilution assays, including the method described in this report, the method of Sterne (15), a Bayesian technique, and some robust procedures. The methods were compared on the basis of expected confidence interval width, assuming a distribution for true patient concentration values which was based on experience from HIV clinical trials. The LRT inversion method described in this report was found to be the best method among those considered, although the Bayesian procedure was nearly as good. The result is not surprising, since LRTs are regarded by many statisticians as the general method of choice (reference 10, p. 48).

When their use can be justified, parametric procedures are more powerful and informative than nonparametric alternatives. MLE applied to the simple product binomial model has been widely used to estimate the concentration of target entities in a dilution assay. We have built on this tradition by using likelihood methods to obtain a confidence interval for this concentration and an indicator of assay quality called the PGOF. The statistical procedures are based on the likelihood function, using MLE and LRTs. These methods are in the mainstream of modern statistical practice.

APPENDIX

Glossary of probability and statistical terms.

binomial model: the probability model for the total number of heads obtained in a certain number of repeated independent tosses of the same coin

event: a subset of the sample space; a collection of outcomes (In this report, all subsets of the finite sample space are regarded as events.)

function: a mapping from one set to another which uniquely associates a member of the second set with each member of the first set

hypothesis test: a rule for deciding on the basis of data whether to accept or reject a given hypothesis (Usually the rule is based on a test statistic, with extreme values of the statistic leading to rejection. An incorrect rejection is called a false-positive error. The determination of extreme values depends on what false-positive error rate is tolerable in making the decision.)

independent events: a collection of events such that the probability that any subcollection of them will jointly occur is the product of the individual probabilities of events in the subcollection

likelihood: probability viewed as a function of the parameters for a fixed outcome

LRT p value for testing the simple hypothesis H that $C = c_0$, given data y_0: the probability of a relative likelihood less than or equal to that observed (y_0), assuming that $C = c_0$.

MLE of a parameter: the parameter estimate which maximizes the likelihood that the given results will be obtained.

mutually exclusive events: a collection of events with the property that no two of the events have any outcomes in common; i.e., the occurrence of either event precludes the occurrence of the other event

95% confidence interval for a parameter: an interval produced by using a statistical recipe, such that the recipe has a probability of at least 0.95 of producing an interval containing the true parameter value (Any specific interval of numbers either contains or does not contain the true value, so it would not make sense to assert that the interval has a probability of 0.95 of containing the true value.)

(null) hypothesis: a tentatively entertained statement about a parameter

parameter: a number or vector which affects the distribution of probability, e.g., the probability of heads in coin tossing

parameter estimate: a statistic whose purpose is to estimate a parameter

PGOF: the probability of results as rare as or rarer than those obtained, assuming that the probability model is correct and the true parameter value is equal to its MLE

probability: assignment of nonnegative numbers to events with the properties that the probability of the whole sample space is unity and the probability of a union of two mutually exclusive events is the sum of their individual probabilities

p value for testing the simple hypothesis H that $C = c_0$, given data y_0 and test statistic t: the probability of an outcome as extreme as or more extreme than that obtained, as determined by $t(y)$, assuming that H is true, i.e., the probability that $t(y)$ is less than or equal to $t(y_0)$, assuming that H is true

relative likelihood: the likelihood normalized relative to its maximum achievable value, i.e., the likelihood divided by the maximized likelihood

sample space: the set of all possible outcomes of an experiment (For this report, finite sample spaces were assumed.)

simple hypothesis: a hypothesis asserting that a parameter has a particular value, such as the hypothesis that $C = 3$

statistic: a number which can be computed from observational data

test statistic: a function of $t(y)$ of outcome y such that small values of $t(y)$ are regarded as surprising or extreme if the null hypothesis is true but are not surprising or extreme if the null hypothesis is false.

ACKNOWLEDGMENTS

This research was supported by NIH/NIAID/DAIDS contract NO1-AI-82517.

We gratefully acknowledge improvements suggested by James Bremer, Robert Coombs, Victor DeGruttola, Scott Hammer, William Meyer, Patricia Reichelderfer, and a very dedicated referee.

REFERENCES

1. **American Public Health Association.** 1976. Standard methods for the examination of water and wastewater, 14th ed. American Public Health Association, Washington, D.C.
2. **Cochran, W. G.** 1950. Estimation of bacterial densities by means of the "most probable number." Biometrics **6:**105–116.
3. **Does, R. J. M. M., L. W. G. Strijbosch, and W. Albers.** 1988. Using jackknife methods for estimating the parameter in a dilution series. Biometrics **44:**1093–1102.
4. **Eisenhart, C., and P. W. Wilson.** 1943. Statistical methods and control in bacteriology. Bacteriol. Rev. **7:**57–137.
5. **Finney, D. G.** 1978. Statistical method in biological assay, 3rd ed. Macmillan, New York.
6. **Fisher, R. A.** 1922. On the mathematical foundations of theoretical statistics. Trans. R. Soc. **222:**309–368.
7. **Greenwood, M., and G. U. Yule.** 1917. On the statistical interpretation of some bacteriological methods employed in water analysis. J. Hyg. **16:**36–54.
8. **Hamilton, M. A.** 1991. Estimation of typical lethal dose in acute toxicity studies, p. 61–88. *In* D. Krewski and C. Franklin (ed.), Statistics and toxicology. Gordon & Breach, New York.
9. **Hollinger, F. B. (ed.).** 1993. AIDS Clinical Trials Group virology manual for HIV laboratories, p. MIC-1–MIC-5. Division of AIDS, Rockville, Md.
10. **Kalbfleisch, J. D., and R. L. Prentice.** 1980. The statistical analysis of failure time data. John Wiley & Sons, Inc., New York.
11. **Kalbfleisch, J. G.** 1985. Probability and statistical inference, vol. 2: statistical inference. Springer-Verlag, New York.
12. **Loyer, M. W., and M. A. Hamilton.** 1984. Interval estimation of the density of organisms using a serial dilution experiment. Biometrics **40:**907–916.
13. **Myers, L. E.** Unpublished data.
14. **Roberts, R. A.** 1989. Review of probability and statistical inference. J. Am. Stat. Assoc. **84:**842–843.
15. **Sterne, T. E.** 1954. Some remarks on confidence or fiducial limits. Biometrika **41:**275–278.
16. **Strijbosch, L. W. G., W. A. Buurman, R. J. M. M. Does, P. H. Zinken, and G. Groenewegen.** 1987. Limiting dilution assays. J. Immunol. Methods **97:**133–140.