APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 1978, p. 756-761
0099-2240/78/0035-0756/$02.00/0

Vol. 35, No. 4
Printed in U.S.A.

Evaluation of the Accuracy and Precision of Enumerating Aerobic Heterotrophs in Water Samples by the Spread Plate Method

JAMES B. KAPER, AARON L. MILLS,† AND RITA R. COLWELL*
Department of Microbiology, University of Maryland, College Park, Maryland 20742

Received for publication 12 October 1977

Factors associated with accuracy and precision in the enumeration of aquatic aerobic heterotrophs by the spread plate method were evaluated by using a nested analysis of variance experimental design. Variances associated with individual components of the spread plate procedure were isolated, and optimal replications of each step were allocated. A practical scheme for optimal allocation of resources is proposed, consisting of four subsamples and two plates per subsample and yielding a total variance decrease of 70% from a single-subsample, 10-plate series. Data transformation was, in general, unnecessary for intraexperiment or intrasample statistical analysis, whereas interexperiment or intersample comparisons may require transformation of data. Rapid changes in the numbers of organisms in stored water samples were observed that were not reproducible and did not follow detectable trends, with increases or decreases in counts occurring in samples regardless of whether they were stored at room temperature or refrigerated, or stored in plastic or glass containers. Rapid sample handling is strongly recommended to minimize variations in the microbial populations of samples for aquatic environments.

The most commonly used and widely accepted methods for enumerating aerobic heterotrophs from water and soils involve plate counts (1, 3, 7). Pour plates are not considered to be the method of choice for environmental studies because of the die-off of autochthonous organisms unable to survive exposure to temperatures of molten agar, i.e., 45 to 50°C (3). Thus, spread plating is the main alternative when growth of the organisms to be counted is required. Results of a number of studies have been published which describe media designed to provide a more complete assessment of heterotrophic bacterial populations in environmental samples (6). Until recently, however, relatively few studies have been reported that consider problems associated with the plate count method from the point of view of the precision and accuracy of the method in estimating the total number of aerobic heterotrophic microorganisms in samples collected from aquatic ecosystems. Palmer et al. (9) investigated the effect of patchiness on enumeration of heterotrophs in water samples and concluded that many samples must be taken at each station to minimize error. No consideration, however, was given to sources of error in the enumeration procedure, especially after the initial sampling was accomplished. The study reported here follows up that communication in dealing with handling and plating of samples after collection. The objectives of the study were to determine the major sources of error and to develop a procedure to minimize such errors.

MATERIALS AND METHODS

Water samples were collected during the summer of 1976 during several cruises in Chesapeake Bay aboard the R/V RIDGELY WARFIELD. Sampling stations included a midchannel location off Chesapeake Beach and a site near Parson's Island in Eastern Bay in the Upper Chesapeake Bay (see Fig. 1). Water samples were collected with a Niskin sterile bag sampler (8) at depths approximately 15 cm below the surface. Samples for plating or dilution were either withdrawn from the bag with a sterile pipette or transferred aseptically to a sterile glass dilution bottle. Samples that were refrigerated before plating were held at 4°C in a shipboard refrigerator between platings. Those samples left unrefrigerated were held at ambient temperature, which varied from cruise to cruise but was in the range of 25 ± 5°C.

Samples were plated by standard procedures. Serial decimal dilutions of the original samples were made, using sterile Upper Bay salts (NaCl, 5.0 g; $MgSO_4 \cdot 7H_2O$, 1.5 g; KCl, 1.6 g; distilled water, 1,000 ml, pH 7.2 to 7.4) as diluent. One-tenth milliliter of sample or appropriate dilution was spread uniformly over the surface of a plate containing 20 ml of Upper

† Present address: Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22903.

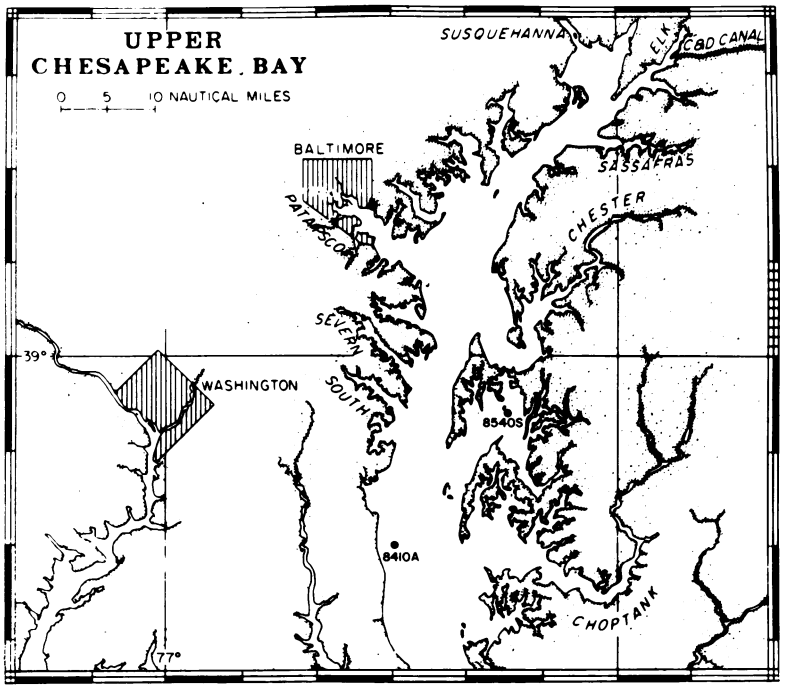

FIG. 1. *Map of Chesapeake Bay showing sampling stations. Chesapeake Beach is station 8410A and Eastern Bay is station 8540S.*

Bay yeast extract agar (UBYE) medium (yeast extract, 1.0 g; peptone, 1.0 g; agar, 20 g; Upper Bay salts, 1,000 ml, pH 7.2 to 7.4). The plates were incubated for a period of time, depending upon the experiment involved. At the end of the incubation period, colonies were counted with a Spencer plate counter, the number of colony-forming units (CFU) per milliliter was calculated, and the data were recorded.

Statistical analyses were conducted by using programs available on the University of Maryland UNIVAC 1108 computer.

An initial experiment was done to determine the stability of the number of heterotrophs recovered from a water sample over a 24-h period. A 100-ml portion of the sample was removed from the Niskin bag after collection and placed in a glass bottle. This subsample was held at ambient temperature on the bench top on board the ship between platings.

Because of the significant change noted in numbers of organisms recovered in the first 2.5 h, a similar experiment was conducted to determine changes over a much shorter period of time. In addition, several other treatments were tested. Two samples were taken simultaneously from a given cast. One of the samples was divided into two aliquots, with one aliquot immediately plated and the other held in the bag at ambient temperature. The second sample was refrigerated in the sampling bag between platings. Also, from each of the Niskin bag samples, two 100-ml subsamples were taken and placed in sterile glass bottles. One bottle from each of the two bags was held in the refrigerator between platings, while the other was held at ambient temperature on the bench top aboard ship. The samples were plated out very quickly, with less than 1 min between the time of removal of the sampler from the hydro wire and the actual spreading of the water samples on the agar plates. The samples plated out immediately upon retrieval of the sampler bag from the hydro wire constituted the t_0 sample in each experiment.

To determine the variance associated with plating of natural water samples, 20 replicate plates on UBYE agar for each of two dilutions were prepared. A 2-ml pipette was used for each dilution, and all 20 plates within a dilution were spotted from the same pipette. Thus, each plate represented the number of organisms in each of the 20 0.1-ml portions in the pipette, excluding changes due to currents within the pipette and differential adhesion of cells to the inner walls of the pipette.

To isolate factors contributing the most error in the plating procedure, five components were identified as potential contributors to total variance for a single sample: technician, subsampling (i.e., patchiness within the sample), dilution (i.e., patchiness within the subsample), pipette (for lack of a better term, the variance within a single dilution "tube"), and individual plate error. Figure 2 presents the different components as viewed in this study.

Counts were compiled and the need for data transformation was assessed, since the analysis of variance (ANOVA) technique assumes homogeneity of variance of the data for all groups; i.e., if homogeneity of variance is not found, data transformation is necessary in

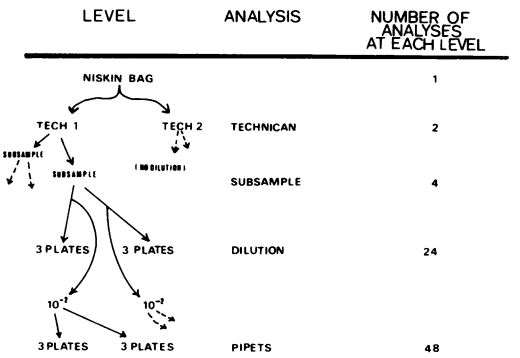

FIG. 2. *Experimental design for analysis of plating errors. Dilution factors were for the dilution as plated, since 0.1 ml was used for each spread plate. The structure of the experiments varied somewhat, but the design principle depicted was used throughout the study.*

order to use the ANOVA technique. A number of procedures exist to test for homogeneity, including Bartlett's test and the *F-max* test (12). The latter was used in this study. Transformation of the data, when performed, was by the "log e" transformation. The data were analyzed by using a nested ANOVA design (12), allowing partitioning of the total variance into individual components (technician, subsample, dilution, pipette, and plate).

RESULTS AND DISCUSSION

Results of the platings of the subsample held at ambient air temperature aboard ship at 0, 2.5, 5.0, 20.25, and 24.25 h were (CFU per milliliter): 5.4×10^3, 1.4×10^4, 8.4×10^4, 1.8×10^6, and 1.6×10^6, respectively. These counts represent the average of four replicate platings. Counts recovered from the various treatments are shown in Table 1. The number of organisms recovered from each Niskin bag at t_0 was identical, 6.2×10^3 cells/ml. After storing the sample for 180 min, counts of the water in the sampling bags held at ambient temperature and at refrigerated temperature increased to 1.3×10^4 and 1.0×10^4/ml, respectively. When subsamples were transferred from the Niskin bags to glass bottles and immediately sampled, counts in all of the bottles were 52 to 63% of the counts obtained from samples withdrawn from the Niskin bag with a pipette. Counts of water samples stored in glass containers held at ambient temperature increased only slightly or, in some cases, decreased, whereas counts obtained from water samples stored under refrigeration in glass containers increased in approximately the same proportion as samples held in the Niskin bags.

Results of the 20 replicate plate series are given in Table 2. Although the mean values obtained for the two dilutions compared favorably (4.52×10^3 versus 4.58×10^3), replicates within each plating varied significantly. Standard deviations and confidence intervals showed that, even with 20 replicates, precision was not good. It is interesting to note that the distribution of organisms within each pipette followed no discernible pattern.

Results of the statistical analysis are shown in Table 3. From the *F* ratios, significant differences can be seen in three of the components of error in the spread plate procedure. No significant differences were found between technicians in any of the experiments carried out. On one occasion, significant differences were observed among subsamples taken from the same sample,

TABLE 1. *Changes in total viable count recovered from samples stored in the sampling bag or in glass bottles, either refrigerated (R) or held at ambient (A) temperature*

Sample	Temp	CFU/ml × 10^{-3}		
		0[a]	35	180
Niskin[b]	A	6.2	7.3	13
Bottle[b]	A	3.6	3.6	3.9
Bottle[b]	R	3.9	3.1	7.2
Niskin[c]	R	6.2	9.3	10
Bottle[c]	A	3.6	2.0	2.2
Bottle[c]	R	3.2	4.7	5.4

[a] Minutes after sample retrieval.
[b, c] Indicates source of samples was Niskin bag having similar superscript notation.

TABLE 2. *Total viable counts obtained from 20 replicate plates from each of two decimal dilutions*

Plate no.	Colonies counted	
	10^{-2}	10^{-3}
1	TNTC[a]	0
2	24	Confluent
3	34	0
4	35	3
5	34	4
6	53	4
7	56	9
8	74	11
9	95	3
10	78	9
11	24	1
12	17	0
13	25	0
14	25	8
15	50	2
16	24	3
17	34	7
18	62	7
19	45	4
20	69	12
Mean	45.16	4.58
Standard deviation	21.6	3.77
95% confidence interval	$35.4 < x < 54.8$	$0 < x < 13.5$

[a] Too numerous to count.

TABLE 3. *Summary of nested ANOVA design*

Determination	Sample				
	1	2	3	4	5
F ratios					
Technician	0.063 (NS)[a]	1.35 (NS)	1.854 (NS)	0.521 (NS)	2.973 (NS)
Subsample	2.416 (NS)	0.966 (NS)	0.750 (NS)	12.67[b]	3.208 (NS)
Dilution tubes			2.591[c]		0.798 (NS)
Pipette	3.904[c]	0.930 (NS)	1.143 (NS)	2.28 (NS)	1.55 (NS)
Variance component[d]					
S_{tech}^2	0 (2)[e]	1 (2)	376 (2)	0 (2)	26,561 (2)
S_{ss}^2	55 (2)	0 (2)	0 (4)	102,909 (4)	18,529 (2)
S_{dil}^2			2,884 (2)		0 (2)
S_{pi}^2	58 (2)	0 (2)	453 (2)	9,902 (2)	14,450 (2)
S_{pl}^2	59 (3)	40 (3)	6,347 (2)	15,450 (2)	81,382 (3)

[a] NS, Not significant.
[b] Significant at the 0.05 level.
[c] Significant at the 0.001 level.
[d] $\times 10^3$.
[e] Numbers in parentheses indicate number of replicates.

indicating patchiness within a single 1.5-liter sample. In two experiments a 10^{-2} dilution was counted, thus necessitating the use of replicate dilution tubes. Results were conflicting, with one experiment showing significant difference between the replicate dilution tubes and the other indicating no significant difference in the dilution steps. The procedure of inserting a pipette into a dilution tube or subsample and withdrawing an aliquot to spot onto a plate showed that there was a significant difference associated with the pipettes in only one of the experiments. In the other four experiments, no significant differences among pipettes were detected.

The nested ANOVA design allows separation of variance components for each step in the procedure. The individual variance components derived from untransformed data are listed in Table 3. In nearly every case, the level with the largest variance was the plating step, reflecting variability of numbers occurring from one plate to another. In several cases, the calculated variance component was a negative number. Since there cannot be a negative variance, these are reported as zero. Isolated instances of large variance revealed that there was patchiness in the samples, such as the large subsample variance (S_{ss}^2) in experiment 4. Of course, it is understood that variance components should not be compared across experiments, since different populations are represented. Therefore, intraexperiment trends are emphasized; i.e., plate variance (S_{pl}^2) was usually greatest, and technician variance (S_T^2) was lowest.

When subjecting bacterial count data to statistical analysis, the primary consideration is the need for data transformation. Bacteria in the environment are usually not normally distributed, but neither are they always found in Poisson or other common distribution (5). A variety of transformations have been proposed to treat viable count data (2, 6), but no single transformation will apply in every case. Frequently, transformation of data is essential so that the assumptions of the statistical procedure being performed are not violated, thereby resulting in loss of sensitivity or validity. Thus, each set of data for a particular analysis should be examined individually by a procedure such as Bartlett's test or the *F-max* test (12) to determine the need for transformation. Interestingly, the data examined here did not require transformation to insure preservation of the assumptions of the ANOVA procedure, at least for intraexperiment conditions. Since we were interested in the spread plate technique itself and in the various components of the technique that were potential sources of error, interexperiment statistical comparisons were not made, and therefore transformation of the data was not required.

To examine the effects of transformation on the data, both log e and square root transformations were performed, and the results were compared with those of the untransformed analysis. As expected, the mean squares and variance components changed considerably in magnitude, but the *F* ratios and the resulting probabilities changed only slightly. In no case did statistically significant differences become nonsignificant or vice versa. It is concluded that intraexperiment or intrasample data often require no transformation for statistical analysis, such as the nested ANOVA, but that interexperiment or intersample comparisons more likely will require transformation of data to insure validity of the statistical analysis.

The nested ANOVA design is extremely useful for examining variance associated with individual steps or components of the spread plate procedure. In one experiment, significant differences were associated with the pipette used, indicating that duplicate pipettes used in spotting samples onto triplicate plates yielded significantly different results. The difference is concluded to arise from patchiness in the given subsample involved, which contained 10 to 15 ml of water sample. Another experiment showed that there were significant differences among duplicate dilution tubes, each of which had been inoculated with 1 ml of subsample. Again, patchiness within the subsample was indicated. In other experiments, however, significant differences within subsamples were not observed, but significant differences among subsamples were noted, arising most likely from patchiness in the sample.

No consistency in significant differences detected in all the experiments was noted, even though the experiments were conducted over a period of 2 months, an observation in agreement with that of Palmer et al. (9), who examined patchiness over a much larger scale. Examination of the individual components of variance revealed that there was a similar inconsistency. For example, one level or component contributed almost no variance in one experiment, but essentially all of the total variance in another experiment. Two general trends that should be noted, however, were that differences among technicians, in general, contributed the least to total variance, whereas individual plates prepared during the count procedures usually contributed the most variance.

That the plating step usually yielded the largest variance component suggests that increasing the number of replicate plates from two or three to five or ten replicates should decrease the variance. Some investigators routinely include a large number of replicate plates in their procedure, which decreases the variance of the plating step (5, 10). However, significant variance can remain in the upper levels, i.e., subsamples, dilutions, and pipettes, which are unaffected by a decrease in plating variance. It is the overall or total variance (S_T^2) that is pivotal in providing the best estimate of the sample population density. The obvious answer to this problem is to increase replication at all levels in order to decrease the variance at each level, with a resulting decrease in total variance. However, this clearly would lead to great increases in cost and time and could quickly become unreasonable. Fortunately, with variance components in hand, the optimum allocation of replications for decreasing the total variance can be determined. In a system consisting of subsamples (ss), pipettes (pi), and plates (pl), the total variance is equal to:

$$S_T^2 = \frac{S_{ss}^2}{n_{ss}} + \frac{S_{pi}^2}{(n_{ss})(n_{pi})} + \frac{S_{pl}^2}{(n_{ss})(n_{pi})(n_{pl})}$$

where S^2 represents the variance for the respective components and n equals the number of replications at each step.

By increasing replication at different steps, the total variance can be calculated for evaluating the effectiveness of different replication schemes. Table 4 shows several examples of such manipulation. By increasing the number of replicate plates from two to three, a 7% decrease in total variance was realized; an increase to ten plates decreased S_T^2 by 17%. However, by increasing n_{ss} from one to two, with two plates for each subsample, a 50% decrease in variance resulted. Comparing the $n_{ss} = 2$, $n_{pl} = 2$ system with the $n_{ss} = 1$, $n_{pl} = 10$ system, a threefold decrease in variance resulted with the former method, with four instead of ten plates. Thus, increasing replication at the upper levels is the key to decreasing total variance. This idea, although not new to statisticians, is rarely seen in bacteriological reports, where the lower levels become increasingly replicated in hopes that dramatically lower variance will result. This principle holds true no matter what variance components are used for calculation.

Selection of replication need not be based solely on variance. For many investigators, cost may be a more important factor. The most efficient experimental design with respect to cost and variance will result when the product of variance and cost is minimized. The following equation represents the concept, where $S_{\bar{x}}^2$ = overall variance of the population mean and C_t = total cost (11):

$$C_t S_{\bar{x}}^2 = \left(\frac{S_{ss}^2}{n_{ss}} + \frac{S_{pi}^2}{(n_{ss})(n_{pi})} + \frac{S_{pl}^2}{(n_{ss})(n_{pi})(n_{pl})}\right) \cdot (C_{ss}n_{ss} + C_{pi}n_{ss}n_{pi} + C_{pl}n_{ss}n_{pi}n_{pl})$$

The product is minimized when the optimal ratio of n_{pl}/n_{pi} equals

$$\sqrt{C_{pi}S_{pl}^2/C_{pl}S_{pi}^2}$$

and the ratio of n_{pl}/n_{ss} equals

$$\sqrt{C_{ss}S_{pl}^2/C_{pl}S_{ss}^2}.$$

When devising a replication scheme for precision in the spread plate procedure, other sources of error that can be minimized should not be overlooked. For example, diluent and dilution tubes should be autoclaved separately and aseptically combined very precisely. Sub-

TABLE 4. *Decrease in total variance with increasing replication*[a]

n_{ss}	n_{pi}	n_{pl}	Total n_{pl}	S_T^{2b}	% Decrease[c] in S_T^2
1	1	2	2	142,500	
1	1	3	3	133,000	7
1	1	10	10	118,900	17
2	1	2	4	71,250	50
4	1	2	8	35,625	75

[a] n_{ss}, Number of replicate subsample; n_{pi}, number of replicate pipettes; n_{pl}, number of replicate plates; S_T^2, total variance.

[b] Total variance was compared using $S_{ss}^2 = 55{,}000$, $S_{pi}^2 = 58{,}000$, $S_{pl}^2 = 59{,}000$.

[c] Percent decrease in S_T^2 based on $n_{ss} = 1$, $n_{pi} = 1$, $n_{pl} = 2$.

sample tubes and dilution tubes should be mixed well on a Vortex mixer to achieve a homogeneous suspension. If simple points such as these are neglected, the value of optimum replication is negated.

Besides precision, microbiological analyses should also strive for accuracy. Often, microbiologists attempt to determine the number of organisms present in a sample collected from a given environment at a specific time. The large changes in numbers of organisms recovered from samples stored for long or short periods of time that can occur, even when samples are refrigerated, emphasizes the need for rapid handling of samples as soon as they are collected, since a delay of just a few minutes can cause alteration of the results, a fact noted in earlier work in marine microbiology by ZoBell (13). At times, of course, immediate plating of samples may be impractical or even impossible. In those instances, results from stored samples should be considered only as crude estimates of the original microbial populations. Whenever possible, studies should be conducted to determine the effect of storage on the microbial levels, i.e., do numbers increase? decrease? how much? and are the changes consistent?

Many investigators are now turning to direct counting methods such as epifluorescence microscopy to determine total numbers of microorganisms in water samples (4). Such methods should improve accuracy of the estimates of total numbers. However, these techniques are subject to the same sampling errors as are viable counts. Furthermore, direct counts cannot distinguish among physiological types, i.e., proteolytic, starch hydrolytic, and other ecologically important groups, so that plating of samples to enumerate CFU present in the samples will very likely continue to be used in quantitative microbiological analyses. Therefore, improvement in precision and accuracy in the plate counts should be sought whenever possible.

ACKNOWLEDGMENTS

This work was supported by Sea Grant 04-3-158-7, National Oceanic and Atmospheric Agency, and National Science Foundation grant GB-35261X. The computer time for this project was supported in full through the facilities of the Computer Science Center of the University of Maryland.

Larry W. Douglas provided extremely helpful advice and critical review for this project.

LITERATURE CITED

1. **American Public Health Association.** 1965. Standard methods for the examination of water and wastewater. American Public Health Association, Inc., New York.
2. **Ashby, R. E., and M. E. Rhodes-Roberts.** 1976. The use of analysis of variance to examine the variations between samples of marine bacterial populations. J. Appl. Bacteriol. **41**:439–451.
3. **Buck, J. D., and R. C. Cleverdon.** 1960. The spread plate as a method for the enumeration of marine bacteria. Limnol. Oceanogr. **5**:78–80.
4. **Daley, R. J., and J. E. Hobbie.** 1975. Direct counts of aquatic bacteria by a modified epi-fluorescent technique. Limnol. Oceanogr. **20**:875–882.
5. **Eisenhart, C., and P. W. Wilson.** 1943. Statistical methods and control in bacteriology. Bacteriol. Rev. **7**:57–137.
6. **Goulder, R.** 1976. Evaluation of media for counting viable bacteria in estuaries. J. Appl. Bacteriol. **41**:351–355.
7. **Jannasch, H. W., and G. E. Jones.** 1959. Bacterial populations in seawater as determined by different methods of enumeration. Limnol. Oceanogr. **4**:128–129.
8. **Niskin, S. T.** 1962. A water sampler for microbiological studies. Deep Sea Res. **9**:501–503.
9. **Palmer, F. E., R. D. Methot, Jr., and J. T. Staley.** 1976. Patchiness in the distribution of planktonic heterotrophic bacteria in lakes. Appl. Environ. Microbiol. **31**:1003–1005.
10. **Postgate, J. R.** 1969. Viable counts and viability, p. 611–628. *In* J. R. Norris and D. W. Ribbons (ed.), Methods in microbiology, vol. 1.
11. **Snedecor, G. W., and W. G. Cockran.** 1967. Statistical methods, 6th ed. State University Press, Ames, Iowa.
12. **Sokal, R. R., and F. J. Rohlf.** 1969. Biometry. W. H. Freeman and Co., San Francisco.
13. **ZoBell, C. E.** 1946. Marine microbiology. Chronica Botanica Co., Waltham, Mass.