ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Nov. 1990, p. 2246–2249
0066-4804/90/112246-04$02.00/0

Vol. 34, No. 11

Statistical Analysis of the Effects of Trial, Reader, and Replicates on MIC Determination for Cefoxitin

HANNAH M. WEXLER,[1,2]* PHILIP T. LAVIN,[3] ERIC MOLITORIS,[2] AND SYDNEY M. FINEGOLD[1,2,4,5]

Medical[4] and Research[2] Services, VA Wadsworth Medical Center, and Departments of Medicine[1] and Microbiology and Immunology,[5] University of California at Los Angeles School of Medicine, Los Angeles, California 90073, and Department of Surgery, Harvard Medical School, Boston, Massachusetts 02215[3]

Received 26 June 1989/Accepted 12 September 1990

A pilot study was designed to estimate the variance components in the determination of the MIC of cefoxitin for isolates of the *Bacteroides fragilis* group. Twenty different organisms were tested, and replicate, trial, and reader variabilities were examined. When the total-variance component was used, if the true MIC was 16 μg/ml, then the chance that the observed MIC was between 8 and 32 μg/ml, inclusive, was 95%. For all analyses, the isolate ($P = 0.0001$) and reader ($P < 0.03$) effects were significant. The probability of specific MIC observations for various true MICs (over the range of 16 to 32 μg/ml at 4-μg/ml increments) was calculated. For true MICs of 20, 24, and 28 μg/ml, the probabilities of observing an MIC of 16 or 32 μg/ml (inclusive) were 86, 75, and 62%, respectively. An upward bias was shown to exist in addition to sources of sizeable variation. The recommendation stemming from recognition of this inherent variability is that ranges of percent susceptibility at various concentrations be included in reports of in vitro susceptibility studies.

Bacteroides fragilis group strains are the anaerobes most frequently isolated from clinical infections and are among those most resistant to antimicrobial agents. In vitro susceptibility testing may provide valuable information about activities of certain agents and trends of anaerobic resistance patterns, but variation in testing methods among laboratories has led to a great deal of confusion regarding the reported resistance of *Bacteroides* species to various antimicrobial agents (1, 6, 7). The procedures themselves are considered accurate to within ±1 twofold dilution, i.e., if the MIC for the quality control strain is within 1 twofold dilution of the quality control value, the study run is acceptable. Clustering of MICs about the breakpoint for certain antimicrobial agents has been seen in our laboratory and described by our group (7). In studies completed in our laboratory, we found that the MICs for 50 to 60% of all anaerobes (70% of the *B. fragilis* group) were within 1 twofold dilution of the breakpoint for cefoperazone, cefotaxime, ceftizoxime, cefoxitin, and penicillin G while 38% (46% of the *B. fragilis* group) were within 1 dilution for clindamycin (5, 6).

We were concerned that the clustering of MICs combined with the inherent technical variability might result in significant nonmeaningful inconsistencies in reported results and decided to analyze the significance of a given reported MIC statistically. This pilot study was designed to estimate the variance components in the determination of the MIC of cefoxitin for isolates of the *B. fragilis* group.

MATERIALS AND METHODS

Twenty different cultures were analyzed. For cultures 1 to 5, there were three replicates, four different trial days, and two readers for each test. For cultures 6 to 10, there were two replicates, four different trial days, and two readers for each test. For cultures 11 to 20, there were four trial days and two readers for each test. All cefoxitin dilutions were from the same stock, and the same two readers evaluated all 20 cultures. MICs were measured in powers of 2, with outcomes ranging from 1 to 128.

Bacteria tested were recent clinical isolates from the Veterans Administration Wadsworth Medical Center, Los Angeles, Calif. MICs were determined by a twofold agar dilution technique described previously (4), with an inoculum of 10^5 CFU and brucella base laked-blood agar.

The data were analyzed on the basis of random-effects model (8). The isolates were regarded as fixed effects; all isolates were readily available. The readers, tests, and replicates were regarded as random effects. Nesting terms were included because of replication of test samples for given isolates (8). (Nesting is a measure of replication of the basic experimental unit being evaluated.) In this study, we use the general linear model (GLM) to identify component sources of variation (8). The order of terms in the GLM for MIC outcome were isolate, replicate nested in isolate, test, test times replicate nested in isolate, reader, and an error term. (The order corresponds to the experimental design structure.) The replicate nested in isolate term was included to address specimen division error, while the test times replicate nested in isolate term addressed test-to-test handling error. Results were analyzed separately for isolates 1 to 5, 6 to 10, and 11 to 20 because of the different numbers of replicates in the three subgroups. Separate analyses for each isolate set allowed for confirmation of results across each set. The analysis was repeated across all isolates by using the average over all replicates. The replicate nested in isolate and the test times replicate nested in isolate terms were dropped from this model. In the analysis, base 2 logs of the MICs were taken to stabilize the variance.

The SAS Institute (Cary, N.C.) guide was used to perform data analysis on a COMPAQ 386 microcomputer (3). A PROC VARCOMP routine was used to estimate the error term of the model and the standard deviation of the MIC. The factor variance component refers to the contribution of the particular factors (isolate, replicate, or reader) to the total variance. Probability statements for the true mean MIC were computed by using a Z statistic (9). The *P* values for significance of individual sources of variation were obtained

* Corresponding author.

TABLE 1. MIC data for cultures, replicates, trials, and readers

Culture no.	Replicate	MIC (μg/ml)							
		Trial 1		Trial 2		Trial 3		Trial 4	
		Reader E	Reader D	Reader E	Reader D	Reader E	Reader D	Reader E	Reader D
1	A	2	2	4	4	4	4	8	2
	B			4	4	8	4	8	4
	C	2	2	4	4	8	4	4	4
2	A	32	32	64	64	32	32	32	32
	B			64	32	64	64	64	64
	C	32	32	64	64	64	64	64	32
3	A	4	4	16	16	16	8	8	8
	B			8	8	8	8	8	8
	C	4	4	16	8	16	8	16	8
4	A	32	32	64	64	32	32	32	32
	B			32	32	32	32	32	32
	C	32	32	32	32	64	32	32	32
5	A	32	32	64	64	64	32	64	32
	B			64	32	128	32	64	32
	C	32	32	64	32	64	32	64	32
6	C	2	2	2	2	4	4	2	2
	B			4	4	8	4	2	2
7	C	8	8	16	16	16	8	16	8
	B			8	8	8	8	16	16
8	C	8	8	16	16	32	8	16	8
	B			8	8	16	8	16	8
9	C	32	32	64	64	64	32	32	32
	B			64	32	64	64	32	32
10	C	32	32	32	32	32	32	32	32
	B			32	32	32	32	32	32
11		16	16	32	32	32	32	32	32
12		32	32	64	32	64	32	64	32
13		32	32	32	32	32	32	32	32
14		32	32	64	64	64	64	32	32
15		64	64	64	64	64	64	64	32
16		32	32	64	64	64	64	32	32
17		32	32	64	64	64	32	32	32
18		1	1	1	1	8	1	16	2
19		16	16	32	32	32	32	16	16
20		16	16	16	16	32	16	16	16

by using the test option on the random statement from the type II sum of squares within SAS PROC GLM (3).

RESULTS

The MICs are displayed in Table 1. Some data from trial 1 were not reported and were excluded from analysis (because of technical problems in medium preparation); all other data were used in the analysis.

The results of separate analyses are presented in Table 2. Table 3 summarizes the results for all 20 isolates obtained by using the mean value over the replicates. For all four analyses, the isolate effects were significant ($P = 0.0001$). There was also a significant reader effect ($P < 0.03$ for all four analyses); reader E always read the MIC at levels equal to or higher than those of reader D.

For isolates 1 to 5, the test ($P = 0.0001$) and reader ($P = 0.0001$) effects were highly significant, while the replicate effect ($P = 0.80$), the replicate nested in isolate effect ($P = 0.08$), and the test times replicate nested in isolate effect ($P = 0.10$) were not significant. In other words, the replicate effect, specimen division error, and test-to-test handling error were not significant. The total of the variance components was 0.3847, with an error component of 0.1572 (Table 2). Thus, when the total-variance component was used if the true MIC was 16 μg/ml, then the probability that the observed MIC was 32 μg/ml was 44.6%, the probability that the observed MIC was 16 μg/ml was 44.6%, and the probability that the observed MIC was 8 μg/ml was 5.3%.

For isolates 6 to 10, the test effect ($P = 0.0008$), reader effect ($P = 0.0025$), the replicate nested in isolate effect ($P = 0.04$), and the test times replicate nested in isolate effect ($P = 0.008$) were all significant, while the replicate effect (P

TABLE 2. GLM analysis summary[a]

Model component[b]	GLM model results[c]					
	Isolates 1–5		Isolates 6–10		Isolates 11-20	
	P value[d]	Variance	*P* value	Variance	*P* value	Variance
Isolate	0.0001		0.0001		0.0001	
Replicate	0.7956	−0.0018	0.4861	−0.0113		
Replicate (isolate)	0.0761	0.0014	0.0358	0.0052		
Test	0.0001	0.0967	0.0008	0.0272	0.0031	0.0737
Test × replicate (isolates)	0.1031	0.0608	0.0080	0.1106		
Reader	0.0001	0.0700	0.0025	0.0370	0.0285	0.0360
Error		0.1572		0.1344		0.3591

[a] The type of GLM computation refers to the order in which terms are brought in.
[b] The model component refers to the term brought in last.
[c] GLM model: log MIC = isolate + replicate(isolate) + test + test × replicate(isolate) + reader + error. The total variances for isolates 1 to 5, 6 to 10, and 11 to 20 were 0.3847, 0.3144, and 0.4688, respectively.
[d] *P* values were computed by using type II GLM estimates (3).

= 0.49) was not significant. The total of the variance components was 0.3144, with an error component of 0.1344 (Table 2). Thus, when the total-variance component was used, if the true MIC was 16 μg/ml, then the probability that the observed MIC was 32 μg/ml was 46.3%, for an observed MIC of 16 μg/ml it was 46.3%, and the probability that it would be 8 μg/ml was 3.7%.

For isolates 11 to 20, the test effect ($P = 0.003$) and the reader effect ($P = 0.03$) were significant. The total of the variance components was 0.4688, with an error component of 0.3591 (Table 2). Thus, when the total variance component was used, if the true MIC was 16 μg/ml, then the probability that the observed MIC would be 16 or 32 μg/ml was 42.8%, with the same upward bias.

The replicate effect was not significant for either of the two sets of isolates. Working with the mean for each set of replicates, we pooled the data for all 20 isolates. The test ($P = 0.0001$) and reader ($P = 0.0001$) effects were significant. The total of the variance components was 0.3666, with an error component of 0.2327 (Table 3). Thus, when the total-variance component was used, if the true MIC was 16 μg/ml then the chance that the observed MIC would be between 8 and 32 μg/ml, inclusive, was 95%. Table 4 displays the probability of specific MIC observations for true MICs over the range of 16 to 32 μg/ml at 4-μg/ml increments under the assumption of a standard deviation of 0.605 (as obtained from the total-variance component in Table 3). For a true MIC of 24 μg/ml, the probability of reading 8 μg/ml was 0.5%, that of reading 16 μg/ml was 15.6%, that of reading 32 μg/ml was 59.1%, that of reading 64 μg/ml was 23.8%, and that of reading 128 μg/ml was 1.0%. This illustrates that the chance of reading 16 or 32 μg/ml was 74.7%. For a true MIC of 28 μg/ml, this probability dropped to 62.0%, with a 35.4% probability of reading the MIC as 64 μg/ml. For a true MIC of 20 μg/ml, this probability rose to 85.5% and the probability of being more than one dilution away was 14.5%. Clearly, an upward bias exists, in addition to sources of sizeable variation.

DISCUSSION

The net effect of this analysis is that an MIC determination is subject to a rounding bias, as well as to significant sources of variation, such as test (day) and reader effects. Specimen handling and test-to-test variations introduce significant errors. The results of this study indicate significant potential for error in any MIC studies in which quotas are based on observed MICs as criteria for selection of an isolate for inclusion in the study. For example, if a study were required to have 100 isolates for which the MIC is at least 32 μg/ml for a given antibiotic, then selection of 100 organisms for which the observed MIC is 32 μg/ml would include 5 isolates for which the MICs are 16 μg/ml or lower.

The extent of the upward bias may be estimated in this type of study and corrected if desired. In this study, the correction factor (0.7, reciprocal of the square root of 2) was determined by simulating possible MIC distributions for various true MICs. This type of correction could be used to amend mean MICs; i.e., if the mean MIC were reduced by 30%, the bias would be corrected. Another approach would recognize that when one establishes a breakpoint of 32 μg/ml, for example, the actual cutoff point measured for a sample population is about 30% less, or ~22 μg/ml. (This fact might be of some use in trying to correlate clinical outcomes with susceptibility results, for example.) However, for MICs measured for clinical isolates, or even for percent susceptibility for sample populations, this correction factor would not be useful.

One common assumption was dispelled by this analysis. Generally, when we speak of a ±1 twofold dilution error, the

TABLE 3. GLM analysis summary for isolates 1 to 20

Model factor[a]	Factor *P* value[b]	Factor variance component[c]
Isolate	0.0001	
Test	0.0001	0.0919
Reader	0.0001	0.0420
Error		0.2327

[a] The GLM model was log MIC = isolate + test + reader + error.
[b] *P* values were computed by using type II GLM estimates.
[c] The total variance was 0.3666.

TABLE 4. Probabilities of specific MIC readings

Reading (μg/ml)	Probability that the true MIC (μg/ml) was:				
	16	20	24	28	32
4	0.001	0.000	0.000	0.000	0.000
8	0.048	0.015	0.005	0.002	0.001
16	0.451	0.283	0.156	0.090	0.048
32	0.451	0.572	0.591	0.530	0.451
64	0.048	0.127	0.238	0.354	0.451
128	0.001	0.003	0.010	0.024	0.048

presumption is that the determination would fall on the plus and minus sides equally; this study shows a clear upward bias to the determinations. Increasing the number of times that an isolate was tested did not substantially increase the accuracy of the measurement. The statistical probability of actually observing the true MIC ranged from 42.8 to 46.3% whether determinations were made once, in duplicate, or in triplicate. Thus, there appears to be no advantage in multiple measurements.

The importance of an accurate, generally agreed upon, and easily understandable definition of the endpoint is underscored by this study. In this study, both readers were experienced technicians who had worked in the area of susceptibility testing extensively; both knew that they were involved in a study designed to estimate sources of variability. Nevertheless, there was a consistent discrepancy in their endpoint readings. Other organisms (e.g., *Fusobacterium* spp.) may introduce even greater variability because of the difficulties in endpoint determination (2). Extensive training of technicians involved in susceptibility testing and frequent periodic surveillance of their technique are essential.

The trend (and, at times, editorial pressure) for reporting only MICs for 50 or 90% of the strains tested is disturbing in light of these considerations. The combination of the uncertain actual MICs as described in this study and the clustering effect of MICs about breakpoint concentrations that occurs with many beta-lactam agents, clindamycin, and chloramphenicol may result in a large but insignificant gap between the cumulative percent susceptibility at one concentration and the value at ±1 twofold dilution greater (or less). For example, in the current study, in trial 3, with a breakpoint of 32 μg/ml, isolate 2 would be described as either susceptible or resistant, depending on which replicate was referred to, even with a single reader. Thus, we recommend that workers report the percent susceptibility at a reasonable range of values about the breakpoint (e.g., a range of 3 twofold dilutions). Narrowing the increments tested around the critical breakpoint concentrations or using a system that can measure exact MICs (i.e., narrowing the error factor) would also alleviate some of the uncertainty inherent in this technique.

ACKNOWLEDGMENTS

This work was supported in part by Veterans Administration Medical Research funds and by Merck Sharp and Dohme, Rahway, N.J.

LITERATURE CITED

1. **Finegold, S. M., and the National Committee for Clinical Laboratory Standards Working Group on Anaerobic Susceptibility Testing.** 1988. Susceptibility testing of anaerobic bacteria. J. Clin. Microbiol. **26**:1253–1256.
2. **Johnson, C. C., H. M. Wexler, M. Garcia, S. Becker, and S. M. Finegold.** 1989. Cell-wall-defective variants of *Fusobacterium*. Antimicrob. Agents Chemother. **33**:369–372.
3. **SAS Institute.** 1985. SAS statistics guide for personal computers; version 6, p. 549–565. SAS Institute, Cary, N.C.
4. **Sutter, V. L., D. M. Citron, M. A. C. Edelstein, and S. M. Finegold.** 1985. Wadsworth anaerobic bacteriology manual, 4th ed. Star Publishing Co., Belmont, Calif.
5. **Wexler, H., and S. M. Finegold.** 1985. In vitro activity of imipenem against anaerobic bacteria. Rev. Infect. Dis. **7**(Suppl. 3):S417–S425.
6. **Wexler, H., and S. M. Finegold.** 1987. Antimicrobial resistance in *Bacteroides*. J. Antimicrob. Chemother. **19**:143–146.
7. **Wexler, H. M.** 1989. Susceptibility testing procedures, p. 715–729. *In* S. M. Finegold and W. L. George (ed.), Anaerobic infections in humans. Academic Press, Inc., Orlando, Fla.
8. **Winer, B. J.** 1971. Statistical principles in experimental design, 2nd ed., p. 309–430. McGraw Hill, New York.
9. **Zar, J. H.** 1984. Biostatistical analysis, 2nd ed., p. 83–86. Prentice Hall, Englewood Cliffs, N.J.