JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 1993, p. 2841–2844
0095-1137/93/112841-04$02.00/0

Vol. 31, No. 11

GUEST COMMENTARY

Selection Criteria for an Antimicrobial Susceptibility Testing System

JAMES H. JORGENSEN

Department of Pathology, The University of Texas Health Science Center, San Antonio, Texas 78284-7750

The clinical microbiology laboratory plays an important role in antibiotic selection and use through performance of routine antimicrobial susceptibility testing on patients' clinical isolates. There are now a number of bacterial species which may exhibit acquired resistance to one or more classes of antimicrobial agents. For that reason, susceptibility testing is most important with members of the family *Enterobacteriaceae*, *Pseudomonas* spp., *Staphylococcus* spp., *Enterococcus* spp., *Haemophilus influenzae*, and *Streptococcus pneumoniae*, since clinical isolates may not be susceptible to empiric drug choices. Indeed, the most important function of a susceptibility test is the accurate detection of clinically relevant antimicrobial resistance in individual isolates. Failure of a test to predict antimicrobial resistance could place a patient in jeopardy of increased morbidity or mortality.

Clinical microbiology laboratories can choose from among several manual or instrument-based methods for performance of their routine susceptibility tests. These include the disk diffusion (or Bauer-Kirby) test, antibiotic gradient methods, agar dilution with mechanized inoculator, the broth microdilution method (with or without use of an instrument or a growth indicator), and short-incubation automated instrument methods. The broth microdilution susceptibility test with commercially prepared antibiotic panels has become the most popular of the current methods available to U.S. clinical microbiology laboratories (4). MICs can be determined with the aid of a simple viewing box, with a more elaborate viewer incorporating a computer-assisted data recording system, or with an automated photometer or fluorometer, which reads the growth endpoints in the trays. The reader devices may also be connected to a personal computer, which generates printed reports and can facilitate data storage and retrieval for periodic generation of cumulative susceptibility profiles for organisms which have been tested.

Automated instruments for susceptibility testing may provide test results following a conventional overnight incubation period, or they may allow results to be determined in a shortened analysis period of 3.5 to 10 h (11) by taking advantage of the greater sensitivity of the instruments' optical systems to detect subtle increases in bacterial growth. Most of the instruments may also be used to perform identifications of gram-negative or gram-positive bacteria and may be able to merge and print identification and antimicrobial susceptibility results into a single report (22). This guest commentary focuses on criteria which may be used for evaluation and selection of antimicrobial susceptibility testing systems.

ADVANTAGES AND DISADVANTAGES OF CURRENT SYSTEMS

The advantages of the microdilution susceptibility procedure include the convenience of performing a susceptibility test and an organism identification in the same tray, the generation of a quantitative result (i.e., the MIC), the fact that some organisms can be tested accurately only by an MIC procedure, and the appeal of the computerized reporting systems which are available from the commercial microdilution tray manufacturers. The computerized data management systems may be especially valuable to laboratories which lack a laboratory information system. However, it is also possible to use a computerized reporting system with disk diffusion testing (7). It is important to recognize that an MIC method should not be viewed from the perspective that MICs per se are more valuable to physicians. Indeed, physicians who are not infectious disease specialists are not likely to be able to interpret accurately the MICs of the multitude of contemporary antibiotics. In addition, there is no clear proof that the determination of MICs is superior to qualitative or susceptibility category results in patient management (6). A survey of members of the Infectious Diseases Society of America expressed the view that MIC results may be misinterpreted by some physicians, and that susceptible, intermediate, and resistant category results may be preferable (16).

A laboratory may choose to perform short-term incubation, automated antimicrobial susceptibility testing in order to generate test results more rapidly than can be accomplished by manual methods. Provision of important laboratory results 1 day sooner than by conventional methods is a logical advancement in patient care. However, there is little objective evidence that rapid susceptibility test results reduce mortality or morbidity (8) unless very aggressive measures are taken by the laboratory to make physicians aware of the data (26). This may be due in part to the fact that physicians may not be aware that such capabilities exist or that the rapid results are not available at the time that clinical decisions regarding antibiotic therapy must be made (13).

One of the previous shortcomings of rapid susceptibility testing methods was some sacrifice in the ability to detect certain inducible or otherwise subtle antimicrobial resistance mechanisms (11). The instruments most notable for such problems are no longer marketed, and the manufacturers of the two remaining instruments have made significant strides toward correcting earlier problems (11). However, it is important to emphasize that accuracy should not be sacrificed in an effort to generate a rapid susceptibility result.

A significant problem with commercial microdilution or automated systems is the inflexibility of the standard antibi-

otic batteries or test panels. With the current availability of more than 50 antibiotics in the United States and the diversity which exists between antibiotic formularies in different health care facilities, it is often difficult for a laboratory to find a standard commercial test panel which completely satisfies its needs. With one exception (17), the inherent flexibility of the Bauer-Kirby disk diffusion test has not been matched by the microdilution or automated test systems. There may also be delays in the availability of new antibiotics in an instrument system or in making changes in breakpoints, quality control values, and interpretive categories following their clearance by the U.S. Food and Drug Administration or approval by the National Committee for Clinical Laboratory Standards.

A problem with automated susceptibility test instruments which use both conventional and short incubation periods has been limited instrument quality control procedures. This has occurred largely because the standard control strains often result in many offscale values, i.e., MICs less than or equal to the lowest concentration or greater than the highest concentration tested by the instrument (14). This problem is particularly acute with breakpoint testing panels, in which only one or two concentrations of an antibiotic are tested. The lack of on-scale control values means that the potencies of the antimicrobial agents and the functioning of the instrument may not be determined with the level of precision that clinical microbiologists have come to expect with the disk diffusion susceptibility test.

The concept that automated susceptibility instruments reduce labor requirements through increased efficiency has been overly optimistic. A review of current College of American Pathologists work load values for both manual and automated susceptibility test methods reveals that only minimal labor savings (i.e., less than 1 min of savings per test) are provided by current instrumentation (5). These modest labor savings may be most meaningful to large clinical microbiology laboratories which perform large numbers of tests daily and which may be able to take advantage of even a slight reduction in labor requirements.

OTHER NEW OR NOVEL TEST METHODS

There are two antibiotic agar gradient susceptibility methods which laboratories may consider for testing of fastidious or anaerobic bacteria (3, 10, 12). These methods may be especially useful for testing fastidious organisms such as *H. influenzae*, *S. pneumoniae*, and anaerobic bacteria because special enriched media or a special incubation atmosphere (e.g., anaerobic or increased CO_2) can be used.

There is considerable interest in the possibility of using molecular genetic methods for detection of antimicrobial resistance mechanisms. It is possible to detect *mec* (the gene responsible for the production of PBP 2a in *Staphylococcus* spp., resulting in methicillin resistance), *vanA* (the gene causing high-level vancomycin resistance in enterococci), and various β-lactamases and aminoglycoside-inactivating enzymes (23). Methods employing the use of probes or nucleic acid amplification techniques offer the promise of excellent sensitivity, specificity, and speed for the detection of resistance genes. However, none of these are commercially available, and other problems may limit their application in clinical microbiology, i.e., the presence of genes which may not be expressed or which may not result in clinically relevant levels of resistance (23).

WHAT IS ACCEPTABLE REPRODUCIBILITY AND ACCURACY FOR A SUSCEPTIBILITY TEST?

It is essential that an antimicrobial susceptibility test system provide reproducible results in day-to-day laboratory use and that the results generated by the system be comparable to the results determined by an acknowledged "gold standard" reference method. Definitions of acceptable reproducibility and accuracy should be established with an understanding of the capabilities and limitations of our current technology. Acceptable precision has been defined operationally for MIC tests by the National Committee for Clinical Laboratory Standards (20) in stating that "Generally, the acceptable reproducibility of the test is within one two-fold dilution of the actual endpoint." This is further quantified by stating that with repeat testing, more than 95% of MICs should fall within an expected range, usually 3 $\log_2$ dilutions. The contention that the acceptable precision of an MIC based on dilution testing is a range of three twofold dilutions (target value ± 1 dilution) now seems well established and supported by pioneering studies (1).

Defining the acceptable accuracy of a susceptibility test is a more challenging task when one considers the numerous and often subtle antimicrobial resistance mechanisms that have been elucidated in recent years. Such issues arose in the mid to late 1970s, when anaerobe susceptibility tests were under development and when the first automated antimicrobial susceptibility instruments were marketed. Metzler and DeHaan (18) proposed the error rate-bounded method for determining disk diffusion zone size interpretive criteria for anaerobe susceptibility tests. Error rate-bounded analysis allows zone size breakpoints to be chosen with the goal of minimizing interpretive errors compared with categories derived from MIC determinations for a group of test strains. Those authors proposed that errors of false-susceptible classifications by the disk test should not occur in more than 1% of all tests, and errors of false-resistant classifications should not occur in more than 5% of all tests (18).

Thornsberry and colleagues (24) in 1980 used the categories very major, major, and minor to describe errors of false susceptibility or false resistance or a response involving an intermediate result (respectively) when they evaluated the performance of a new automated system and compared them with the results generated by broth microdilution and disk diffusion. Notably, the authors acknowledged that with certain selected "challenge organisms, disparate results were obtained from the two standard reference methods in one or more laboratories." For the purposes of their study, a reference antibiogram was established for each of their isolates on the basis of knowledge of the organisms' resistance mechanisms. This represents one of the first studies to use a challenge collection of organisms with known resistance properties for evaluation of a new susceptibility testing method. It was also an important acknowledgment that there were imperfections in the conventional methods used to evaluate the results of the new instrument method.

Thornsberry and Gavan (25) suggested in 1980 that "an arbitrary rule of thumb" for evaluating a new instrument test method would be that complete category agreement should be over 90% and that the total of the very major and major errors should be less than 5%. In 1981, Sherris and Ryan (21) offered several tentative suggestions for the acceptable performance of new susceptibility testing methods. They proposed that "very major errors *attributable to the new procedure* should be less than 1.5% for all individual species to be tested." They also suggested that "the overall percent-

age of errors *attributable to the new procedure* should not exceed 5% in tests on random clinical isolates." They contended in their article that errors can be attributed to a procedure if similar results are obtained following repeat testing of both the new and the old procedures. They also pointed out that if the reproducibility of a reference procedure is 95%, "a perfectly reproducible and accurate automated procedure could obviously not exceed 95% correspondence. If both are 95% reproducible, correspondence will be 90.25%, with 0.25% of results being incorrect by both procedures." Thus, a different conclusion could have been reached from their statements, i.e., that a new susceptibility testing procedure would have acceptable accuracy if agreement with a reference method was at least 90%. It is important to recognize that they did not specify the characteristics of the population of isolates on which the requirement for fewer than 1.5% very major errors was based or if all strains in the population would be used in the denominator of the calculation.

To illustrate these points, if a group of 1,000 consecutive clinical isolates were examined in many clinical microbiology laboratories today, few (e.g., <5%) strains resistant to the latest quinolone or carbapenem antibiotics would likely be encountered. If very major testing errors occurred with one-fifth of the resistant strains, the error calculation would be <1.5% very major errors if the entire organism population was used as the denominator. However, if a challenge set of 100 organisms which included 50% resistant strains were tested, and if very major errors occurred with one-fifth of those strains, the very major error calculation would be 10%, if calculated in the same manner. These calculations would result even though test performance was the same with both collections of organisms. Thus, in the same manner that disease prevalence affects the calculation of predictive values (9), the characteristics of the test isolate population affect the susceptibility test error rates when calculated as described above.

In evaluating a new automated susceptibility testing instrument in 1987, Murray and colleagues (19) correctly asserted that calculations of very major errors should be based only upon the number of resistant strains tested, since the susceptible strains cannot contribute false-susceptible results. Likewise, major errors should be calculated only on the basis of the susceptible strains in the population since they cannot contribute errors of false susceptibility. To distinguish this method of error calculation from that proposed earlier (18), Bradford and Sanders (2) later coined the term "risk corrected error rate" to describe the statistically correct means for calculating interpretive category errors. Application of this unassailable logic to the hypothetical examples cited above reveals that in both organism populations, the very major error rates were actually 20%.

In rigorous evaluations of new susceptibility testing methods, it is important to examine a collection of challenge strains which contains an adequate number of resistant strains to verify the ability of the new test to detect resistance. The next step should then be a determination of the new test's performance by examining a number of susceptible strains as might be expected in a large population of unselected isolates to determine the extent of major and minor errors that might be anticipated in a typical clinical laboratory setting. Testing only one type of culture collection could conceal important aspects of any new test that should be known by its potential users. Having made that assumption, it is important to define the level of acceptable accuracy on the basis of this proposed method of evaluation. Referring to the earlier statements regarding the reproducibility of conventional MIC tests as approximately 95%, and assuming that any new test should have equal reproducibility, a new test method should provide >90% agreement with MICs determined by the reference method. I propose that very major interpretive category errors determined on a large sample ($n = \geq 35$) of known resistant isolates should be $\leq 3\%$ (i.e., not more than 1 in 33 isolates should repeatedly test falsely susceptible) and that the combination of major and minor errors attributable to the new test should be $\leq 7\%$ when determined on a large known susceptible population or a large unselected sample of clinical isolates. For antibiotics without an intermediate interpretive category, slight modification of these criteria is required, since it might be possible to spuriously categorize a 1-dilution MIC error as either very major or major if the reference MIC occurred at the single breakpoint for such a drug. Thus, the legitimate measure should be whether the within-1-dilution agreement between the two procedures is >90%. If strains resistant to a new antibiotic have not been recognized, the accuracy of a susceptibility test method for that drug cannot be predicted accurately.

These proposed criteria should be viewed as representing the most desirable performance for a new test method. It is possible that a test with a lower degree of correspondence with traditional methods might in fact be useful for recognition of organisms with certain very difficult to detect resistance mechanisms. An example can be drawn from a study which evaluated a new gradient diffusion susceptibility testing method with four beta-lactam antibiotics on a special collection of stably derepressed type I β-lactamase-producing gram-negative bacilli (15). The results of the study revealed from 0 to 10% very major errors, 0 to 10% major errors, and 0 to 16.7% minor errors compared with the MICs determined by a conventional method. The authors concluded that such performance might be deemed acceptable because of the difficulties of detecting the resistance mechanisms found in this group of organisms by any method.

Thus, the criteria that I propose for the acceptable accuracies of new susceptibility testing methods should not be interpreted rigidly. Instead, it is important to consider the degree of difficulty involved in detecting resistance in some organisms. This is not to suggest that certain methods should be judged more leniently than others or that there are intrinsic reasons why some antibiotics cannot be tested accurately. Merely, it is important to recognize that some fastidious or slow-growing organisms or some subtle resistance mechanisms may not be recognized by any method with the same level of reproducibility and accuracy that can be expected with most other organism-antibiotic combinations.

In summary, contemporary clinical microbiology laboratories have many methods to choose from in selecting a method for their routine antimicrobial susceptibility testing. In an era of great emphasis on cost-containment, it is worth reemphasizing that the most economical susceptibility test method currently available is the Bauer-Kirby disk diffusion test. Its advantages include the fact that it is simple to perform and very reproducible, it does not require any special equipment, it provides category results readily interpreted by clinicians, and it offers great flexibility in selection of drugs for routine test batteries. Despite the attributes of the disk diffusion test, many laboratories will choose instead one of the commercial microdilution or rapid automated instrument methods. These latter methods offer some real or perceived benefits to the laboratory, including the provision

of more rapid or more quantitative results, the possibility of automating certain tasks, and improved data presentation and storage capabilities through use of the personal computer systems available with some instruments.

Few laboratories can afford the resources required to perform a rigorous evaluation of a susceptibility testing system which might be under consideration for acquisition. Microbiologists should make themselves aware of the performance evaluations which have been published in well-respected, peer-reviewed journals regarding the systems under consideration. In addition, a laboratory should perform a limited on-site evaluation using selected control and clinical isolates in an effort to validate the claims made in the literature and by the manufacturer regarding the convenience and reproducibility of a system. It is also important to assess the reputation of the instrument manufacturer with respect to routine service needs and the likely response to unanticipated mechanical or biological problems which might develop. Each laboratory must consider all of these factors in order to make an informed decision as to which methodology or system best suits its needs.

ACKNOWLEDGMENT

I thank C. Alex McMahan for review of the statistical assertions made here.

REFERENCES

1. **Balows, A.** 1972. Current techniques for antibiotic susceptibility testing. Charles C Thomas, Springfield, Ill.
2. **Bradford, P., and C. C. Sanders.** 1992. Use of a predictor panel for development of a new disk for diffusion tests with cefoperazone-sulbactam. Antimicrob. Agents Chemother. **36:**394–400.
3. **Citron, D. M., M. I. Ostovari, A. Karlsson, and E. J. C. Goldstein.** 1991. Evaluation of the E test for susceptibility testing of anaerobic bacteria. J. Clin. Microbiol. **29:**2197–2203.
4. **College of American Pathologists.** 1991. Proficiency Survey Program. CAP surveys final critique for Set D-A, 1991. College of American Pathologists, Northfield, Ill.
5. **College of American Pathologists.** 1992. Workload recording method and personnel management manual. College of American Pathologists, Northfield, Ill.
6. **Craig, W. A.** 1993. Qualitative susceptibility tests versus quantitative MIC tests. Diagn. Microbiol. Infect. Dis. **16:**231–236.
7. **D'Amato, R. F., L. Hochstein, J. R. Vernaleo, D. J. Cleri, A. A. Wallman, M. S. Gradus, and C. Thornsberry.** 1985. Evaluation of the BIOMIC antimicrobial test system. J. Clin. Microbiol. **22:**793–798.
8. **Doern, G. V., D. R. Scott, and A. L. Rashad.** 1982. Clinical impact of rapid antimicrobial susceptibility testing of blood culture isolates. Antimicrob. Agents Chemother. **21:**1023–1024.
9. **Galen, R. S., and S. R. Gambino.** 1975. Beyond normalcy—the predictive value and efficiency of medical diagnosis. John Wiley & Sons, Inc., New York.
10. **Hill, G. B., and S. Schalkowsky.** 1990. Development and evaluation of the spiral gradient endpoint method for susceptibility testing of anaerobic gram-negative bacilli. Rev. Infect. Dis. **12**(Suppl. 2):S200–S209.
11. **Jorgensen, J. H.** 1990. Antibacterial tests: automated or instrument-based methods, p. 1166–1172. *In* A. Balows, W. J. Hausler, Jr., K. L. Herrmann, H. D. Isenberg, and H. J. Shadomy (ed.), Manual of clinical microbiology, 5th ed. American Society for Microbiology, Washington, D.C.
12. **Jorgensen, J. H., A. W. Howell, and L. A. Maher.** 1991. Quantitative antimicrobial susceptibility testing of *Haemophilus influenzae* and *Streptococcus pneumoniae* by using the E-test. J. Clin. Microbiol. **29:**109–114.
13. **Jorgensen, J. H., and J. M. Matsen.** 1987. Physician acceptance and application of rapid microbiology instrument test results, p. 209–212. *In* J. H. Jorgensen (ed.), Automation in clinical microbiology. CRC Press, Inc., Boca Raton, Fla.
14. **Kellogg, J. A.** 1985. Inability to adequately control antimicrobial agents on AutoMicrobic System gram-positive and gram-negative cards. J. Clin. Microbiol. **21:**454–456.
15. **Knapp, C. C., and J. A. Washington.** 1992. Comparison of the E test and microdilution for detection of β-lactam-resistant mutants that are stably derepressed for type I β-lactamase. J. Clin. Microbiol. **30:**214–215.
16. **Kunin, C. M., and S. Chambers.** 1985. Responsibility of the infectious diseases community for optimal use of antibiotics: views of the membership of the Infectious Diseases Society of America. Rev. Infect. Dis. **7:**547–559.
17. **Lancaster, M. V., and R. Fields.** 1992. User-defined, colorimetric antimicrobial susceptibility assays. Am. Clin. Lab. **11:**14–16.
18. **Metzler, C. M., and R. M. DeHaan.** 1974. Susceptibility tests of anaerobic bacteria: statistical and clinical considerations. J. Infect. Dis. **130:**588–594.
19. **Murray, P. R., A. C. Niles, and R. L. Heeren.** 1987. Comparison of a highly automated 5-h susceptibility testing system, the Cobas-Bact, with two reference methods: Kirby-Bauer disk diffusion and broth microdilution. J. Clin. Microbiol. **25:**2372–2377.
20. **National Committee for Clinical Laboratory Standards.** 1990. Methods for dilution antimicrobial susceptibility testing for bacteria that grow aerobically. Approved standard M7-A2. National Committee for Clinical Laboratory Standards, Villanova, Pa.
21. **Sherris, J. C., and K. J. Ryan.** 1982. Evaluation of automated and rapid methods, p. 1–5. *In* R. C. Tilton (ed.), Rapid methods and automation in microbiology. American Society for Microbiology, Washington, D.C.
22. **Stager, C. E., and J. R. Davis.** 1992. Automated systems for identification of microorganisms. Clin. Microbiol. Rev. **5:**302–327.
23. **Tenover, F. C.** 1992. Bauer and Kirby meet Watson and Crick: antimicrobial susceptibility testing in the molecular era. ASM News **58:**669–672.
24. **Thornsberry, C., J. P. Anhalt, J. A. Washington II, L. R. McCarthy, F. D. Schoenknecht, J. C. Sherris, and H. J. Spencer.** 1980. Clinical laboratory evaluation of the Abbott MS-2 automated antimicrobial susceptibility testing system: report of a collaborative study. J. Clin. Microbiol. **12:**375–390.
25. **Thornsberry, C., and T. L. Gavan.** 1980. Automated procedures for antimicrobial susceptibility tests, p. 491–494. *In* E. H. Lennette, A. L. Balows, W. J. Hausler, Jr., and J. P. Truant (ed.), Manual of clinical microbiology, 3rd ed. American Society for Microbiology, Washington, D.C.
26. **Trenholme, G. M., R. L. Kaplan, P. H. Karahusis, T. Stine, J. Fuhrer, W. Landau, and S. Levin.** 1989. Clinical impact of rapid identification and susceptibility testing of bacterial blood culture isolates. J. Clin. Microbiol. **27:**1342–1345.