INTERNATIONAL STANDARD

ISO 20776-2

Second edition
2021-12

Clinical laboratory testing and in vitro diagnostic test systems — Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices —

Part 2: Evaluation of performance of antimicrobial susceptibility test devices against reference broth microdilution

Systèmes d'essais en laboratoire et de diagnostic in vitro — Sensibilité in vitro des agents infectieux et évaluation des performances des dispositifs pour antibiogrammes —

Partie 2: Évaluation des performances des dispositifs pour antibiogrammes par rapport à une méthode de référence de microdilution en bouillon

Reference number
ISO 20776-2:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 212, *Clinical laboratory testing and in vitro diagnostic test systems,* in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 140, *In vitro diagnostic medical devices,* in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

This second edition cancels and replaces the first edition (ISO 20776-2:2007), which has been technically revised.

The main changes are as follows:

— Revision in the title of this document to better align with the intended information.

— Addition of an Introduction (not present in the first edition).

— Revised Clause 3 as follows:

  — Removed definitions for category agreement, susceptible, intermediate, resistant, non-susceptible, major discrepancy, minor discrepancy, very major discrepancy, breakpoint test and zone diameter;

  — Added definition for contemporary isolate (3.11.1), and removed definitions for fresh isolate, recent isolate;

  — Added definitions for reproducibility (3.9), bias of the test method (3.10.3), sensitivity analysis (3.10.4.1), specificity analysis (3.10.4.2), bacterial organism group (3.16);

  — Added definition for qualitative test (3.7) and removed definition for breakpoint test;

  — Revised definitions for minimum inhibitory concentration test (3.4), breakpoint (3.6), quality control (3.8), discrepancy (3.10.1).

— Reordered Clause 4 (Test methods);

— Moved general requirements for a performance evaluation as a separate section, to the overview (now renamed general section, subclause 4.1) under test methods);

— Revised quality control section, subclause 4.2, and referenced EUCAST and CLSI documents for quality control ranges;

— Revised subclause 4.2.1 (Reference method) to add variability;

— Revised subclause 4.2.2 (Strain selection) and incorporated new definition of contemporary isolates (3.11.1);

— Revised subclause 4.2.5 (Reproducibility testing);

— Updated subclause 4.2.8 (Discrepancy resolution testing);

— Combined data analysis and acceptance criteria subclauses (Clause 5);

— Revised subclause 5.1 (Accuracy of test device) to remove category agreement;

— Revised data analysis for MIC devices to remove category agreement. Added bias requirement;

— Removed acceptance for breakpoint AST devices;

— Added provisions on acceptance criteria for qualitative AST devices (5.1.3) and included sensitivity and specificity requirements;

— Revised subclauses on quality control of test device and reproducibility of test device (5.2 and 5.3);

— Revised Bibliography;

— Added Annex A — Evaluation the Performance of MIC Tests, Annex B — Rationale for Bias Analysis, and Annex C — Sensitivity and Specificity Analyses for Qualitative Tests.

A list of all parts in the ISO 20776 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

In vitro antimicrobial susceptibility tests are performed on bacteria suspected of causing disease, particularly if the isolate is thought to belong to a species that can exhibit resistance to frequently used antimicrobial agents. The tests are also important in resistance surveillance, epidemiological studies of susceptibility and in comparisons of new and existing agents.

Dilution procedures are used to determine the minimum inhibitory concentrations (MICs) of antimicrobial agents for antimicrobial susceptibility testing. MIC methods are used in resistance surveillance, defining and identifying wild type phenotypes, comparative testing of new agents, to establish the susceptibility of organisms that give equivocal results in routine tests, for tests on organisms where routine tests can be unreliable and when a quantitative result is required for clinical management. In dilution tests, bacterial strains are tested for their ability to produce visible growth in broth (broth dilution) containing serial dilutions of the antimicrobial agent or on a series of agar plates (agar dilution).

The lowest concentration of an antimicrobial agent (in mg/l) that, under defined in vitro conditions, prevents the appearance of visible growth of an isolated bacterial strain within a defined period of time, is known as the MIC. Careful control and standardization are required for intra- and interlaboratory reproducibility of broth MIC tests. The MICs of quality control (QC) strains generally span three doubling dilutions with a dominant central value, but can have a four-dilution range.

Broth micro-dilution denotes the performance of the broth dilution test in micro-dilution trays. Broth micro-dilution is now one of the most common methods used globally to perform antimicrobial susceptibility tests.

This document is the second edition of ISO 20776-2. It is designed for the evaluation of antimicrobial test devices against the standard broth micro-dilution reference method (ISO 20776-1) using pure cultures of aerobic bacteria that are easily grown by overnight incubation on agar and grow well in standardized micro-dilution trays containing standardized Mueller-Hinton broth (volume of ≤200 µl), which can need to be modified depending on the antimicrobial agent being tested.

Quantitative MIC and qualitative evaluations detailed in this revised document measure the accuracy, reproducibility and QC of tests performed with antimicrobial test devices that generate MIC values against the standard broth micro-dilution reference method. Antimicrobial agar disc diffusion tests are not included in this document.

This document has been revised using the premise that the MIC test is an in vitro assay, subject to intra- and interlaboratory assay variation. When making the comparison between any derivative test and that of the reference method, it is appropriate to apply measures of assay performance only and not result interpretation. For this reason, and because interpretive categories were removed from the second edition of ISO 20776-1, categorical agreement (CA) and its associated terminology, as described by the U.S. Food and Drug Administration (FDA), the Clinical and Laboratory Standards Institute (CLSI) M23 document, and other international documents, has not been applied. Avoiding an assessment of CA also assists in reducing the requirement to reassess assay performance automatically when the only change has been a breakpoint change (which is external to the assay itself).

This document applies to new performance evaluations initiated after the publication date; studies conducted prior to the acceptance date of this document should not need to be re-designed and/or re-analysed using these criteria. Studies conducted prior to these standards or acceptance of this document follow standard practice or guidance at the time of the study.

For derivative tests with more than three two-fold dilutions, assay performance is assessed with tools designed to measure accuracy using essential agreement (EA) and bias, and precision using EA only. For derivative tests with 1 to 3 concentrations, assay performance is assessed using standard sensitivity and specificity measures.

Clinical laboratory testing and in vitro diagnostic test systems — Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices —

Part 2: Evaluation of performance of antimicrobial susceptibility test devices against reference broth micro-dilution

1 Scope

This document establishes acceptable performance criteria for antimicrobial susceptibility test (AST) devices that are used to determine minimum inhibitory concentrations (MIC) of bacteria to antimicrobial agents in medical laboratories.

This document specifies requirements for AST devices and procedures for assessing performance of such devices. It defines how a performance evaluation of an AST device is to be conducted.

This document has been developed to guide manufacturers in the conduct of performance evaluation studies.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 20776-1, *Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices — Part 1: Broth micro-dilution reference method for testing the in vitro activity of antimicrobial agents against rapidly growing aerobic bacteria involved in infectious diseases*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
antimicrobial susceptibility test device
AST device
device, including all specified components used to obtain test results that allow MIC determination of bacteria with specific antimicrobial agents

Note 1 to entry: Specific components of the device include inoculators, disposables and reagents, media used to perform the test, and readers or analysers. Non-specific components, such as swabs, pipettes and tubes, are not part of the device.

3.2
reference method
method of analysis recognized by experts or used as a reference by agreement between parties, which gives, or is supposed to give the accepted reference value of the measurand

Note 1 to entry: For the purpose of this document, the reference method described in ISO 20776-1 is employed. This reference method describes dilution procedures to determine the *minimum inhibitory concentration* (3.3) of antimicrobial agents.

[SOURCE: ISO/TS 22176:2020, 3.1.20, modified — Note 1 to entry added.]

3.3
minimum inhibitory concentration
MIC
lowest concentration that, under defined in vitro conditions, prevents visible growth of bacteria within a defined period of time

Note 1 to entry: The MIC is expressed in mg/l.

3.4
minimum inhibitory concentration test
MIC test
test that is capable of determining a *MIC* (3.3) covering a range of at least four consecutive doubling dilutions, and for which *essential agreement (EA)* (3.10.2) can be determined

3.5
on-scale MIC test result
on-scale minimum inhibitory concentration (MIC) test result
result from a *minimum inhibitory concentration (MIC) test* (3.4) when there is growth in at least one dilution below the MIC endpoint and no growth in at least one dilution above

3.6
breakpoint
specific values of parameters, such as *MICs* (3.3), on the basis of which bacteria can be assigned to clinical categories such as "susceptible" (S) or "resistant" (R)

Note 1 to entry: For current interpretive breakpoints and interpretive categories, reference should be made to the latest publications of organizations employing this *reference method* (3.2) (e.g. CLSI[2] and EUCAST[3]).

3.7
qualitative test
test for which the principal objective is to provide a qualitative result

EXAMPLE Using a *breakpoint* (3.6) or screening concentration.

Note 1 to entry: Such tests have a limited range of 1 to 3 doubling dilutions.

3.8
quality control
QC
use of carefully selected bacterial strains with given expected *minimum inhibitory concentration* (*MIC)* (3.3) results

Note 1 to entry: *MICs* of antimicrobial agents for control organisms should be within the ranges given in the latest editions of the CLSI M100 document[2] or the EUCAST Quality Control document.[4] It is not possible to provide a single Quality Control Table.

3.9
reproducibility
extent to which consistent results such as MICs are obtained when the test is repeated

3.10 Terms relating to the evaluation of test results

3.10.1
discrepancy
difference in a result between the test method [either a *minimum inhibitory concentration* (*MIC) test* (3.4)] or a *qualitative test* (3.7)) and the result of the *reference method* (3.2) outside the region of *essential agreement (EA)* (3.10.2) (*MIC test*), or outside the region of sensitivity and specificity (*qualitative test*)

3.10.2
essential agreement
EA
minimum inhibitory concentration (*MIC)* (3.3) result obtained with the *antimicrobial susceptibility test (AST) device* (3.1) that is within plus or minus one two-fold dilution step from the *MIC* value established with the *reference method* (3.2)

Note 1 to entry: Used for *MIC* devices.

Note 2 to entry: Another representation of the concept is:

$$\frac{N_{EA}}{N} * 100$$

where:

N_{EA} is the number of bacterial isolates with an EA;

N is the total number of bacterial isolates tested.

Note 3 to entry: The overall EA is expressed as a percentage.

3.10.3
bias of the test method
evaluation of test device results to determine whether the results that differ from the *reference method* (3.2) are significantly skewed or predominantly in one direction

Note 1 to entry: Used for *minimum inhibitory concentration (MIC) tests* (3.4).

3.10.4 Terms relating to sensitivity analysis

3.10.4.1
sensitivity analysis
<screening or *breakpoint* (3.6) test> measure of agreement between test device results and *reference method* (3.2) results that are positive or above a published *breakpoint* (3.6)

Note 1 to entry: This can also be considered as positive percent agreement when reference results are interpreted as positive.

Note 2 to entry: Used for *qualitative tests* (3.7). See Table 1.

Table 1 — Sensitivity analysis for a qualitative (screening or breakpoint) test

		Reference method		Total
		(−) or no growth	(+) or growth	
Test method	(−) or no growth	*a*	*b*	*a+b*
	(+) or growth	*c*	*d*	*c+d*
Total		*a+c*	*b+d*	Sum of (*a,b,c,d*)
Sensitivity = $100*[d \div (b+d)]$				

3.10.4.2
sensitivity analysis
<three-dilution *qualitative test* (3.7)> measure of agreement between test device results and *reference method* (3.2) results that have the *MICs* (3.3) at the high end of the scale

Note 1 to entry: Used for *qualitative tests* (3.7). See Table 2.

Table 2 — Sensitivity analysis for a three-dilution qualitative test

		Reference method			Total
		≤ Low MIC	Middle MIC	≥ High MIC	
Test method	≤ Low MIC	*a*	*b*	*c*	*a+b+c*
	Middle MIC	*d*	*e*	*f*	*d+e+f*
	≥ High MIC	*g*	*h*	*i*	*g+h+i*
Total		*a+d+g*	*b+e+h*	*c+f+i*	Sum of (*a* to *i*)
Sensitivity = $100*[i \div (c+f+i)]$					

3.11 Terms relating to bacterial isolates

3.11.1
contemporary isolate
isolate recovered from a clinical sample within the previous six months that has been minimally sub-cultured

Note 1 to entry: Ideally, they are consecutive and prospectively collected. These isolates can have been frozen prior to use.

3.11.2
stock isolate
isolate recovered from a clinical sample that has been retained, stored or obtained from a culture collection

Note 1 to entry: Stock isolates are usually included because they have known or rare resistance mechanisms or are of a genus or species for which the antimicrobial agent is indicated but are not commonly isolated. Such organisms are unlikely to be available in *contemporary isolates* (3.11.1) used in the evaluation. There is no requirement for how long ago the isolate was obtained.

3.12
coordinator
person empowered by the manufacturer or *investigator* (3.13) with responsibility for the entire performance evaluation

3.13
investigator
person responsible for the execution of the performance evaluation at a certain location

3.14
evaluation plan
description of a planned performance evaluation

3.15
evaluation report
description of and conclusions from a performance evaluation

3.16
bacterial organism group
group of related bacterial genera and species that share similar characteristics

4 Test methods

4.1 General

The manufacturer or investigator takes the responsibility for the initiation and the conduct of a performance evaluation according to the evaluation plan. The manufacturer shall define the responsibility and the interrelation of all personnel who manage and conduct a performance evaluation.

The manufacturer or investigator shall appoint a coordinator with overall responsibility for the performance evaluation and the evaluation report. This may include a coordinator, who shall assess and document criteria used and indicate which performance claims are met.

An evaluation conducted by a manufacturer shall consist of accuracy using contemporary and stock strains, reproducibility and quality control (QC) testing performed in at least three different laboratories, of which a maximum of one can be the manufacturer's laboratory. Alternatively, these studies may be conducted at a single site that mimics three sites (e.g. multiple users/instruments, geographically diverse source of organisms). This single site can be within the manufacturer's laboratory. The complete testing protocol should focus on the most commonly used manufacturer's instructions for use or primary methods.

Alternatively, studies incorporating variations of the manufacturer's instructions for use (e.g. inoculation procedures or manual reads of the device) or other secondary methods should consist of the reproducibility and QC sections described below. These studies can be conducted at a single site that mimics three sites (e.g. multiple users/instruments, geographically diverse source of organisms). This single site can be within the manufacturer's laboratory.

Changes made to the test device to correspond to published breakpoint changes for one group of organisms (e.g. Gram-negative fermentative bacilli) or changes made to QC range changes should not require re-test or re-analysis of all bacterial isolates if those changes involve previously validated concentrations in the test device. Initial evaluation of test devices can be performed with longer dilution sequences than used on final products, to provide a comprehensive evaluation while giving flexibility in dilutions provided on product for routine use.

4.2 Methods

4.2.1 Reference method

The reference method shall be as described by ISO 20776-1.

The reference method procedure can be performed either simultaneously with the test device at all testing sites, or at a single site for all isolates tested in the study, which can be that of the manufacturer. If the reference method and test device are tested at the same site, the reference method and the test device shall be set up on the same day from the same inoculum source.

The variability of the reference method can be determined prior to or during AST device studies. A single reference method test point can be used. If reference variability is noted, the results should indicate either the need to perform reference testing at a single site to eliminate variability, or to perform replicate testing of the reference method, in which case the mode or median value of the replicate tests is used.

NOTE In some cases, results from other widely accepted methods can be used in addition to the reference MIC result to arbitrate results. For example, tests that detect the presence of a specific resistance gene, such as the *mecA* gene (encoding oxacillin resistance) or the gene product (PBP 2a), are widely employed and are considered reference methods for detecting oxacillin resistance in staphylococci.

4.2.2 Strain selection

An evaluation protocol should incorporate at least 300 clinical isolates overall (100/site) relevant to an antimicrobial agent. Only one isolate per species per patient shall be included. The collection should

include isolates from as many genera and species as feasible within the intended use of the device. It should include as many unrelated strains representing different degrees of susceptibility to the antimicrobial agents as possible.

The collection should not be weighted with a single species or species complex, unless the device is intended only for that species or species complex. Stock isolates can be used to supplement the contemporary clinical isolates in order to provide strains with different resistance mechanisms. A reasonable approach is 25 contemporary isolates and 75 stock isolates per site. If a device is intended for testing a single genus or species, at least 100 isolates in total from all sites should be studied. When possible, at least 25 on-scale isolates should be tested overall, in order to obtain a meaningful estimate of bias.

A set of at least 10 strains shall be defined to assess intra- and interlaboratory reproducibility of the AST device. The QC strain collection shall, as a minimum, include strains defined in the AST device package insert and any other strain(s) needed to provide on-scale results.

4.2.3 Quality control (QC)

The QC strains shall be tested on the reference and test device every day that testing is performed on the reference or test device.

MICs of antimicrobial agents for control bacteria should be within the ranges given in the latest versions of the CLSI M100 document[2] or the EUCAST Quality Control document.[4] It is not possible to provide a single QC table. For current QC ranges, reference should be made to the latest publications of organizations employing this broth micro-dilution reference method (CLSI and EUCAST).

Some of the standard QC strains described by CLSI or EUCAST might not have on-scale results in the final test method configuration. The test method can have additional QC strains that are essential as part of the test procedure that are not described by CLSI or EUCAST. QC ranges employed for these strains should be as described by the manufacturer in the AST device package insert.

4.2.4 Quality control (QC) of the reference method

If QC results for any antimicrobial agent/bacterium combination is out of range on the reference method and the antimicrobial agent has only one on-scale QC organism, all testing for that day shall be repeated for that antimicrobial agent with the reference method. If the reference method and the test device are being tested simultaneously at the same site, all testing for that day shall be repeated.

For antimicrobial agents with two or more on-scale QC organisms, the following apply:

a) If QC results for one antimicrobial agent/bacterium combination is out of range on the reference method, whilst the other QC strain(s) is (are) within the expected range, the test results for that antimicrobial agent/bacterium combination for that day can be acceptable if the QC results are within the expected range on the next testing day.

b) If the QC result for any antimicrobial agent/bacterium combination is out of range on the reference method for two successive days, both days' results shall be repeated for that antimicrobial agent with both the reference method and test device.

c) If QC results for two or more on-scale QC strains are out of range on the reference method for any antimicrobial agent/bacterium combination on one day, all testing for that day shall be repeated for that antimicrobial agent with both the reference method and test device.

4.2.5 Reproducibility testing of test device

Triplicate testing of a minimum of ten strains (whenever possible including those with on-scale MIC test results for long dilutions of the antimicrobial agent being tested) shall be performed on at least three days at each site where the test device is under evaluation. The number of on-scale isolates should be indicated in the final report. It is likely that the final test configuration can only have a subset of the dilutions tested.

4.2.6 Isolate testing protocol

Isolate testing for the device, including QC and reproducibility, shall be according to the manufacturer's instructions for use. Comparison of the test device result is made to the MICs or qualitative results of the reference method.

4.2.7 Inoculum preparation

The standardization of the inoculum for the test device shall be according to the manufacturer's instructions for use.

4.2.8 Discrepancy resolution testing

Discrepancy resolution testing can be performed for organisms with MIC test results outside of essential agreement (EA) with the reference method. Additionally, discrepancy resolution can be used for qualitative tests. In this case, a discrepancy is defined as a result that does not agree with the reference. If there is reasonable evidence of a technical error (e.g. mixed culture was tested, wrong incubation conditions), both the reference and the test method shall be repeated and the repeat results generated shall replace the original results.

If there is no obvious indication of a technical error, discrepancies may be resolved by a one-time triplicate repeat of both the reference method and test method using separate bacterial inoculum suspensions. Alternatively, for discrepant results, a one-time duplicate repeat of both the reference method and test method using separate bacterial inoculum suspensions can be performed and combined with the original test result to generate three results used for analysis. When conducting additional testing to investigate discrepancies, an equal number of isolates having concordant results should also be tested. The concordant isolates should be selected based on similar organism types and/or similar test results, if available, with each isolate tested in triplicate by both methods.

Final interpretation of the test and reference results after discrepancy resolution testing will be based on comparing the mode or median from triplicate testing of both methods (see Table 3).

Table 3 — Interpretation of additional testing results

MIC test							
Isolate	Test method			Reference method			Essential agreement (EA) of modes/medians
	Initial	Additional	Mode/Median	Initial	Additional	Mode/Median	
A	1	1, 2, 2	2	4	2, 4, 4	4	Yes
B	1	1, 2, 4	2	4	2, 4, 4	4	Yes
C	8	2, 2, 4	2	1	1, 2, 4	2	Yes
D	1	1, 1, 2	1	4	2, 4, 4	4	No
Qualitative test							
E	+	+, +, +	+	-	+, +, +	+	Yes

4.2.9 System under evaluation

All components of the device (e.g. densitometer, reader, optics, interpretive algorithm) used for evaluation shall be equivalent to the commercial device configuration. The evaluation plan used for the study of the AST device shall not deviate from the standard procedure specified by the manufacturer.

5 Data analysis and acceptance criteria

5.1 Accuracy of test device

5.1.1 General

For MIC devices, overall EA and bias shall be calculated. For qualitative devices, sensitivity and specificity shall be calculated.

5.1.2 MIC devices

MIC AST devices should always have both an overall EA of ≥ 90 % when compared to the reference method result(s) and less than ±30 % bias. Overall EA should, at a minimum, be separated by Gram-positive, Gram-negative fermentative and Gram-negative non-fermentative organisms tested and should be calculated separately for organisms with different reference methods (e.g. *Streptococcus* spp.). Additionally, a separate analysis of bias will be performed and if bias of greater than ±30 % is observed, a comment can be added to labelling if not resolved with discrepancy resolution testing.

Analysis should be made of discrepancies to determine whether particular bacteria groups are affected and require limitations for use of the device with that bacterial species and particular antimicrobial agents (refer to Annex A and Annex B).

5.1.3 Qualitative AST devices

Qualitative AST devices should always have a sensitivity and specificity of ≥ 95 % when compared to the reference method result(s). Overall sensitivity and specificity should, at a minimum, be separated by Gram-positive, Gram-negative fermentative and Gram-negative non-fermentative organisms tested and should be calculated separately for organisms with different reference methods (e.g. *Streptococcus* spp.).

Analysis should be made of discrepancies to determine whether particular bacterial species or bacteria organism groups are affected and require limitations for use of the device with that bacterial species or groups and particular antimicrobial agents (see 4.2.8 and Annex C).

5.2 Quality control (QC) of test device

QC strains tested on the test device should always be in the expected range stated in the most current QC standard for either CLSI or EUCAST, for at least 95 % of the results from all sites combined during the device evaluation study period. If the test method has additional QC strains that are not described by CLSI or EUCAST, results for these strains should always be within the expected range described by the manufacturer for at least 95 % of the results obtained during the evaluation study period.

5.3 Reproducibility of test device

Reproducibility for the test device is conducted by comparing the test device to itself, and not to the reference broth micro-dilution panel. Reproducibility for an MIC device shall be within plus or minus one doubling dilution of the mode/median or span a maximum of three dilutions of that antimicrobial agent for ≥ 95 % of the results. Reproducibility for a qualitative or 3-dilution MIC device shall provide exact agreement for ≥ 95 % of the results.

5.4 Documents related to study

A product description and an evaluation plan shall be written prior to the start of an evaluation. A final report shall be produced, clearly indicating overall performance of the test device as compared to the reference method for each organism group and/or genus and each antimicrobial agent. The performance of the device shall be stated for each antimicrobial agent. The laboratories that participated in the evaluation study shall be listed.

Annex A
(informative)

Evaluating the performance of MIC tests

Performance of MIC tests is based on two metrics, essential agreement (EA) and bias. EA provides a measure of the agreement between the test method and the reference method. Bias provides a measure of the accuracy of the test results when compared to the reference method. Tests that are biased have results that are consistently above or below the reference method.

For the calculation of EA, the range of MICs to be compared should be the same for the test and reference methods. If the actual range of reference MIC values is wider than the MIC range of the test, the values less than the lowest MIC reported by the test should be combined with MICs that correspond to the low end of the test. Similarly, MICs greater than the highest MIC reported by the test should be combined with the MICs that correspond to the high end of the test.

As an example, assume that the test method provides the following MICs: ≤ 2, 4, 8, 16, 32, and > 32. During clinical testing, reference MICs ranging from ≤ 0,5 to > 128 were obtained. Tables A.1 and A.2 illustrate the adjustment to the distribution of reference results in order to calculate EA.

Table A.1 — Actual reference method MIC range and result frequency

≤ 0,5	1	2	4	8	16	32	64	128	> 128
44	85	92	48	13	3	8	3	1	3

Table A.2 — Reference method MIC range and result frequency for calculating essential agreement (EA)

≤ 2	4	8	16	32	> 32
221	48	13	3	8	7

EA is defined as MIC results obtained with the AST device, or test method, that are within plus or minus one doubling dilution of the MIC value from the reference method. Overall, an EA rate is calculated as 100 × (# isolates within EA / total # isolates). An acceptable overall EA rate is ≥ 90 %. Tables A.3 and A.4 summarize how EA is determined.

Table A.3 — MIC comparison

		Reference method						Total
		≤ 2	4	8	16	32	> 32	
Test method	≤ 2	154[a]	17	0	0	0	0	171
	4	66	30	8	1	1	0	106
	8	1	1	1	0	0	0	3
	16	0	0	3	0	2	0	5
	32	0	0	1	2	3	3	9
	> 32	0	0	0	0	2	4	6
Total		221	48	13	3	8	7	300

[a] The shaded cells represent those that are in EA.

Table A.4 — MIC doubling dilution difference distribution (Test MIC – Reference MIC)

							Essential agreement (EA)
≤ −3	−2	−1	0	+1	+2	≥ +3	
1	1	30[a]	192	74	2	0	296/300 98,7 %
[a] The shaded cells represent those that are in EA.							

To evaluate bias, the objective is to compare the percentage of test results greater than the reference and the percentage of test results less than the reference. An unbiased test would have relatively equal percentages. For the method defined in this document, all available data are used including results that occur at the ends of the MIC range.

To demonstrate the bias calculation, consider the distribution of reference results listed in Table A.2. It is only possible to have test results greater than the reference for the reference MICs of ≤ 2 to 32.

A. To determine the percentage of test results greater than the reference, complete the following:

a) Count the number of isolates having reference MICs of ≤ 2 to 32 that also have test MICs greater than the reference

b) Count the total number of isolates having reference MICs of ≤ 2 to 32

c) Divide the number from step 1 by the number from step 2 and multiply by 100

Using Table A.3:

— Number of isolates having reference MICs of ≤ 2 to 32 that also have test MICs greater than the reference: (66+1+1+3+1+2+2) = 76

— Number of isolates having reference MICs of ≤ 2 to 32: (221+48+13+3+8) = 293

— 100 × (76 / 293) ≈ 25,9 %

B. To determine the percentage of test results less than the reference, complete the following:

a) Count the number of isolates having reference MIC of 4 to > 32 that also have test MICs less than the reference

b) Count the total number of isolates having reference MICs of 4 to > 32

c) Divide the number from step 1 by the number from step 2 and multiply by 100

Using Table A.3:

— Number of isolates having reference MICs of 4 to > 32 that also have test MICs less than the reference: (17+8+1+2+1+3) = 32

— Number of isolates having reference MICs of 4 to > 32 (48+13+3+8+7) = 79

— Divide the number from step 1 by the number from Step 2 and multiply by 100 (100 × (32 / 79) ≈ 40,5 %

The last step in calculating bias is finding the difference between the percentage of results greater than the reference and the percentage of results less than the reference. In the example, this is 25,9 % – 40,5 %, or −14,6 %. Bias that is within the interval −30,0 % to 30 % is considered acceptable.

Bias is based on all isolates. When bias does not meet the criterion of being within ±30 %, the following can be considered. One, additional testing can be completed to examine whether test bias can be attributed to the test or random variation. Two, if bias cannot be resolved, labelling should be provided describing the source of bias, such as range of MICs on the scale or organism group.

It is important to note that the total number of on-scale isolates shall be 25 in order to calculate bias. This restriction minimizes the risk of obtaining an artificially high percentage when there are limited numbers of isolates across the MIC range. An example of such a case would be when the existence of isolates having resistant or elevated MICs is limited. When it is not possible to calculate bias, EA becomes the only performance metric.

Annex B
(informative)

Rationale for bias analysis

When comparing AST results between test and reference methods, it is possible to observe high EA while the test method provides results that are consistently above or below the reference. Such cases suggest that the test method results are biased, and thus, it becomes important to evaluate bias when examining AST method performance. Bias, in general, provides a measure of the accuracy of test results when compared to the reference method.

As motivation for the bias analysis described in this document, consider the following. A distribution of dilution differences between two AST methods was created to illustrate the need for evaluating bias. The distribution was based on review of clinical data from multiple antimicrobials when it was possible to achieve a high level of EA. The data included different classes of antimicrobials, different organism groups, and varying numbers of on-scale isolates. From these data the following distribution, provided in Table B.1, represents an average symmetric distribution of dilution differences between results from a test and reference method.

Table B.1 — Distribution of dilution differences between test and reference MIC results

≤ −2	−1	0	+1	≥ +2	Essential agreement (EA)
1,5 %	7,0 %	83,0 %	7,0 %	1,5 %	97,0 %

Table B.2 shows a theoretical approach to illustrate how test bias can affect results. To calculate a bias of +30 %, the following process was used:

— If bias is +30 %, 70 % of the results in a column of the distribution would be unaffected and 30 % would shift to the next higher dilution.

— For the column ≤ −2, 70 % of 1,5 % is 1,05 %, or 1,1 % rounding to the nearest 10th. Thus 1,1 % of the results remain in this column of the distribution.

— The column corresponding to −1 becomes 70 % of 7 % + 30 % of 1,5 % or, 4,9 % + 0,5 %, becoming approximately 5,4 %.

— This process continues from one column to the next.

The same process was applied to all columns and at different levels of bias in order to estimate the shift in the overall distribution of results.

Table B.2 — Effects of bias on the distribution of dilution differences between test and reference MIC results

Bias	≤ −2	−1	0	+1	≥ +2	Essential agreement (EA)
None	1,5 %	7,0 %	83,0 %	7,0 %	1,5 %	97,0 %
+30 %	1,1 %	5,4 %	60,2 %	29,8 %	3,5 %	95,4 %
+40 %	0,9 %	4,8 %	52,6 %	37,4 %	4,3 %	94,8 %
+50 %	0,7 %	4,3 %	45,0 %	45,0 %	5,0 %	94,3 %

The changes in the distributions after applying a bias demonstrate the need for an analysis of this type. The assumption in this illustration, however, is that every isolate can shift in either direction. Isolates that have MICs that are at the extremes of the range of results that can be reported are not capable of demonstrating bias in only one direction. Thus, the bias analysis was split into two groups: isolates

that have reference MICs less than the highest possible result and isolates that have reference MICs greater than the lowest possible result. From analysis of data for each group, it is possible to determine the percentage of test results greater than the reference and the percentage of results less than the reference. The difference between these two percentages becomes the estimate of test bias.

The acceptance criterion of bias being within the interval of −30 % and +30 % was determined by inspection of the effects of bias in Table B.2 and review of clinical data. A difference that falls outside of the given interval correlates with a definite skew in the distribution of MIC differences.

With respect to the selection of the minimum number of isolates available for analysis in each group for calculating bias, the use of at least 25 on-scale isolates as the cut-off provides reasonable probability of detecting a high level of bias, at least 50 %, without creating a risk of a false indication of unacceptable bias when the true percentage is relatively small, 10 % or less.