Contains Nonbinding Recommendations

Draft – Not for Implementation

Validation of Certain In Vitro Diagnostic Devices for Emerging Pathogens During a Section 564 Declared Emergency

---

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.

Document issued on January 7, 2024.

You should submit comments and suggestions regarding this draft document within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852-1740. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions about this document, contact IVDguidance@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Preface

Additional Copies

Additional copies are available from the Internet. You may also send an email request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please include the document number GUI00007020 complete title of the guidance in the request.

Table of Contents

Validation of Certain In Vitro Diagnostic Devices for Emerging Pathogens During a Section 564 Declared Emergency

---

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff or Office responsible for this guidance as listed on the title page.

I. Introduction

The Food and Drug Administration (FDA or Agency) plays a critical role in protecting the United States from threats such as emerging infectious diseases, potential public health emergencies, and public health emergencies. FDA is issuing this draft guidance to describe recommendations for validation of certain in vitro diagnostic devices (IVDs) for emerging pathogens when the Secretary of Health and Human Services has declared that the circumstances exist justifying emergency use authorizations (EUAs) for such IVDs under section 564 of the Federal Food, Drug, and Cosmetic Act (FD&C Act) (hereafter referred to as an "applicable 564 declaration"), based on an underlying determination under section 564 that there is a public health emergency or significant potential for a public health emergency.

For the current edition of the FDA-recognized consensus standards referenced in this document, see the FDA Recognized Consensus Standards Database. For more information regarding use of consensus standards in regulatory submissions, please refer to the FDA guidance entitled "Appropriate Use of Voluntary Consensus Standards in Premarket Submissions for Medical Devices."

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. Background

The Emergency Use Authorization (EUA) authority under section 564 of the FD&C Act allows FDA to help strengthen the nation's public health protections against chemical, biological, radiological, and nuclear (CBRN) threats by facilitating the availability and use of medical countermeasures (MCMs) needed during an actual or potential emergency or material threat. Under section 564 of the FD&C Act, when the Secretary of Health and Human Services (HHS) declares that the circumstances exist justifying the issuance of EUAs, FDA may authorize certain unapproved medical products or unapproved uses of approved medical products to diagnose, treat, or prevent serious or life-threatening diseases or conditions caused by CBRN agents when certain criteria are met, including when there are no adequate, approved, and available alternatives. FDA has used this authority to authorize emergency use of IVDs for eight infectious diseases that have emerged over the past years: H1N1 (2009), H7N9 (2013), MERS-CoV (2013), Ebola (2014), Enterovirus D68 (2015), Zika (2016), Coronavirus Disease 2019 (COVID-19) (2020), and mpox (formerly monkeypox) (2022).[1]

Accurate and reliable IVDs are critical to the diagnosis, tracking, treatment, and interruption of transmission of infectious diseases during outbreaks, as well as for diagnosing and treating diseases or conditions caused by CBRN threats. In the public health emergencies of COVID-19[2] and mpox[3], FDA issued guidances that included enforcement discretion policies for certain unauthorized tests to help rapidly increase national testing capacity early in the outbreaks. Certain tests were made available prior to or without an EUA as described in those policies. Regardless of whether a test is issued an EUA or offered as described in an enforcement discretion policy, it is critical that the test be appropriately validated. Therefore, FDA may take action, as appropriate, against violative tests, including those that lack appropriate validation. This guidance and associated templates are intended to help test manufacturers better prepare for future outbreaks by including FDA's recommendations for test validation during an applicable 564 declaration.

Also, this guidance and associated templates address the recommendations received from two independent assessments of FDA's response to COVID-19. Specifically, FDA selected Booz Allen Hamilton to do such an independent assessment, which culminated in an October 2021 report, "Emergency Use Authorization Assessment – Final Report," that recommended FDA "develop a framework for how to conduct validation of diagnostic tests for emerging pathogens in the setting of a declared PHE." Similarly, the HHS Office of the Inspector General's September 2022 report, "FDA Repeatedly Adapted Emergency Use Authorization Policies To Address the Need for COVID-19 Testing," recommended that FDA "develop a suite of EUA templates for future emergencies involving novel pathogens" and "expand and improve resources" on the EUA process, among other actions FDA has taken or is taking.

III. Scope

[1] The year in each parentheses represents when the first EUA for an IVD was issued for each outbreak.
[2] See FDA Guidance document "Policy for Coronavirus Disease-2019 Tests During the Public Health Emergency (Revised)."
[3] See FDA Guidance document "Policy for Monkeypox Tests to Address the Public Health Emergency."

This guidance describes general recommendations for validation of certain IVDs for emerging pathogens during an applicable 564 declaration. The IVDs in the scope of this guidance are diagnostic tests[4] intended to detect a newly identified, previously unknown, or unusual pathogen(s) to aid in the diagnosis of a serious or life-threatening infectious disease or condition; or to detect a known pathogen(s) that aids in diagnosing a newly identified or unusual clinical presentation of such a disease or condition.

These recommendations apply to test data and information submitted in a pre-EUA[5], an EUA request, or to a test offered as described in an applicable enforcement discretion policy. This guidance does not address the EUA regulatory process; refer to "Guidance for Industry and Other Stakeholders on Emergency Use Authorization of Medical Products and Related Authorities" for additional information.

While the information and recommendations provided in this guidance are intended to be broadly applicable to potential future emerging pathogens, most examples throughout are based on SARS-CoV-2 and similar respiratory viral pathogens. Test manufacturers may also look to the mpox and COVID-19 EUA templates on FDA's website for additional examples.[6] FDA may provide more tailored recommendations for tests for a specific outbreak through separate guidance or pathogen-specific templates, as needed. In any outbreak, FDA continually monitors and assesses the testing landscape in the U.S. and will update its policies and recommendations as appropriate. FDA generally will work interactively with the manufacturer during the development and review of an EUA request to help ensure appropriate validation of a test, particularly given potential changes in recommendations due to the changing circumstances of any outbreak.

This guidance applies to all stages of an outbreak and includes discussion about when appropriate validation may depend on the stage of the outbreak. For example, FDA recognizes that use of a highly sensitive comparator may not be available in the early stages of an outbreak and discusses alternate options for such circumstances.

[4] These IVDs are in vitro diagnostic products as defined in 21 CFR 809.3 that are intended to aid in the diagnosis of disease (referred to herein as "diagnostic tests"), such as molecular or antigen tests. Screening tests, which are used for testing individuals without symptoms or other reasons to suspect illness, are a subset of diagnostic tests. In contrast, serology/antibody and other adaptive immune response tests generally are not used to diagnose a current acute infection and are outside the scope of this guidance. Diagnostic tests may be designed for use in various settings, such as in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory, at the point of care site covered by a laboratory's CLIA certificate, or at home.

[5] A pre-EUA can be submitted *prior to* or *during* an applicable 564 declaration before submitting an EUA request, to provide for early engagement between a manufacturer and FDA. A pre-EUA can only transition to an EUA request if there is a current applicable 564 declaration. The recommendations in this guidance may be helpful to manufacturers preparing for early engagement such as a pre-EUA, even prior to an applicable 564 declaration, as it could help facilitate the completeness of a potential future EUA request.

[6] See mpox templates at: https://www.fda.gov/medical-devices/emergency-use-authorizations-medical-devices/monkeypox-mpox-emergency-use-authorizations-medical-devices#templates and COVID-19 templates at: https://www.fda.gov/medical-devices/covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas#covid19ivdtemplates

Due to differences across tests, including technology and indications for use, as well as different circumstances across outbreaks, some sections of this guidance may not be applicable to all tests. Test manufacturers should consider which sections are applicable based on the stage of the outbreak/availability of validation materials and the design and proposed indication for use of their test.

Alternative approaches may be considered. Please consult with the FDA regarding the potential use of alternative validation approaches and materials via CDRH-IVD-EUA@fda.hhs.gov.

IV. Availability of Templates

FDA has found the use of templates to be beneficial during prior emergencies, for both manufacturers and FDA reviewers, to help facilitate the preparation and submission of pre-EUAs and EUA requests to FDA, and any resulting authorization. A generic template entitled "General IVD Emergency Use Authorization (EUA) Request/Pre-EUA Template" is made available through download from our website[7], and it reflects FDA's current thinking on validation study recommendations, and data and information that should be submitted in pre-EUAs and EUA requests. FDA may provide more tailored recommendations for tests for a specific outbreak through separate guidance or pathogen-specific templates, as needed. Additional templates may be added to our website. For example, FDA plans to provide updated templates as appropriate in the event of a specific outbreak. Templates should be viewed only as recommendations, and alternative approaches can be used unless specific regulatory or statutory requirements are cited.

V. Validation Study Recommendations

Validation should objectively demonstrate that a finished device can consistently fulfill defined user needs and its intended use. We recommend that validation testing is performed under defined operating conditions on the final design of the device. In the case of distributed test kits, validation testing should be performed on initial production units, lots, or batches, or their equivalents.

Accordingly, validation studies should be conducted with the final design of the test system that will be used clinically. Such a test system should include the instrument, reagents, and any other components needed to perform the test, including test materials that are required but not provided. Validation studies should also include necessary software (see Section V.B(13)), such as a software algorithm to apply a threshold/cut-off for result interpretation, and the final labeling including instructions for specimen collection. If the validation studies are conducted with an earlier iteration of the test system, the performance of the final design of the test system can sometimes, depending on the specific change(s) made to the system, be addressed through an equivalency study rather than repeating all the validation studies.

FDA generally recommends that test manufacturers conduct the validation studies outlined in this section that are applicable to the type of test systems for an emerging pathogen. See **Table 1** for more details. Generally, for rapid response to an emergency, FDA recommends developing

[7] Available at: https://www.fda.gov/media/184828/download

test systems that include existing instruments that are lawfully marketed[8] for clinical use. In such cases, FDA review of additional validation data for the components that are already lawfully marketed, such as the instrument and software, may not be needed, including where a right of reference has been granted. For innovative technologies, FDA may request technology-specific studies to assess the known and potential benefits and risks associated with the test.

[8] A "lawfully marketed" device means a device that is in compliance with FDA requirements, which may include premarket authorization.

Table 1. Validation Study Recommendations Based on Test Type

Test Type	Clinical Performance Evaluation	Limit of Detection (LoD)	Inclusivity	Cross-Reactivity and Microbial Interference	Endogenous/Exogenous Interference	High-Dose Hook Effect	Carry-Over/Cross-Contamination	Specimen Stability	Reagent Stability	Fresh/Frozen Specimens	Flex Studies	Usability and User Comprehension	Analytical Equivalency	Software Validation	Basic Safety and Essential Performance	Electromagnetic Compatibility (EMC)	Predetermined Change Control Plan (PCCP)
Lab-based	X	X	X	X	X	A	O	X	X	O	N	N	O	O	O	O	O
Home Collection	X	X	X	X	X	A	O	X	X	O	N	X	O	O	O	O	O
POC	X	X	X	X	X	A	O	X	X	O	X	X[9]	O	O	O	O	O
Home Use	X	X	X	X	X	A	O	X	X	N	X	X	O	O	O	O	O

X = Recommended validation studies
O = Validation studies recommended in certain situations, as described in this guidance
A = Applicable to antigen tests only
N = Generally not applicable

A. Clinical Performance Evaluation

A clinical performance evaluation with at least 30 positive and 30 negative specimens of the appropriate specimen[10] type should demonstrate the performance of the test compared to a highly sensitive comparator method, when available. In situations where an appropriate comparator is not available, such as early in an outbreak, initial test validation could be limited to contrived sample evaluation as discussed in subsection A(1) below.

A highly sensitive comparator method is typically a molecular method (e.g., RT-PCR) that utilizes a nucleic acid isolation method (e.g., silica bead extraction) and multiple target regions for the detection of the analyte with a high sensitivity based on clinical performance from testing natural clinical specimens of the appropriate specimen type. FDA generally considers PPA ≥

[9] Assessment of usability and user comprehension is typically incorporated into the clinical performance evaluation for POC tests, which include representative operators under intended use settings.

[10] For additional context, FDA issued EUAs for tests in public health emergencies prior to COVID-19 based on 50 contrived positive and 50 contrived negative clinical specimens. For similar products outside a declared emergency, FDA generally expects an "all-comers" study of natural clinical specimens until at least 50 positives are obtained. The 30 positive and 30 negative described represents what FDA generally considers to be the minimum number of specimens needed to provide appropriate assurance of performance in an outbreak. Evaluation of fewer specimens may not accurately characterize the true performance of the test. For example, FDA received an EUA request for a molecular test for COVID-19 that included validation with only 12 positive samples, showing perfect performance among this limited sample set. FDA requested evaluation of additional specimens to confirm. When an additional 12 samples were evaluated, the cumulative performance dropped to an unacceptable positive percent agreement (PPA) of 71%, and the EUA request was withdrawn.

95% with clinical specimens to be reflective of high sensitivity. For multianalyte tests, we recommend using an FDA-cleared/approved/authorized molecular test with prospective clinical study data from the past 5 years as the comparator test for assessing clinical performance of the non-emergency analytes on your device. FDA may include further information on what constitutes a highly sensitive comparator method on our website, as applicable.

(1) Initial Stages of the Outbreak - Alternative Specimen Types

Natural clinical specimens are the preferred sample type for validation of a diagnostic test. However, at the early stages of an emerging disease outbreak, disease prevalence may be low and natural clinical specimens may not be readily available.

In such cases, use of contrived (e.g., spiked) specimens could be acceptable. Contrived specimens are specimens that are constructed in the laboratory by placing known concentrations of a microorganism or analyte into individual (not pooled)[11] human specimens known to be negative for that microorganism or analyte (i.e., negative clinical matrix). A minimum of 30 contrived positive samples should be tested including a minimum of 20 samples within 2-fold of the test Limit of Detection (LoD), and the rest spanning the assay testing range.[12]

Additionally, the use of archived samples[13] consisting of positive and negative clinical specimens could be a reasonable alternative, if readily available. Ideally, archived specimens should be accompanied by information to determine sample adequacy, such as the specimen collection date, and date of onset of symptoms, as applicable.

In situations where pathogen stocks are not available, such as at the early stages of an outbreak, use of synthetic material could be considered.[14] When synthetic material is used, it should closely mimic natural materials. For example, if the pathogen is an RNA virus, then synthetic RNA, rather than synthetic DNA, should be used in most cases.

Due to limitations of validation with contrived samples, including those prepared using synthetic or natural materials, emergency use authorization of such tests will typically include a Condition of Authorization (CoA) requiring a clinical performance evaluation with natural patient specimens when it becomes feasible to do, as it is necessary to protect public health.

---

[11] Generally, FDA recommends individual negative matrix for studies such as confirmatory LoD and for constructing contrived specimens as described in Section V.A(1), to represent a range of mucus, particulate matter, etc. which may be present in samples. For different specimen types, or direct swab methods, other approaches could be acceptable.

[12] If too much viral RNA is used, the evaluation might not assess how well the test performs on specimens near the cutoff used to distinguish positive and negative results. This can result in a poorly performing test appearing to perform well. See Section V.B for discussion of analytical validation, including LoD.

[13] For purposes of this guidance, archived samples are defined as specimens collected from a human subject that are known to harbor the analyte of interest (i.e., positive) or not harbor the analyte (i.e., negative). Archived samples should be selected to minimize bias; for instance, samples should not be selected for archiving based on the candidate test. Archived samples are sometimes referred to as retrospective specimens or banked specimens. The appropriateness for use of archived samples, such as length of time in storage or other factors, will vary based on the individual emerging pathogen during an outbreak.

[14] For example, FDA authorized certain COVID-19 tests that were validated with synthetic material through April 2020.

(2) Study Design

Ideally, clinical performance should be established through a prospective, all-comers clinical study in the intended use environment, by the intended user(s), and with natural clinical specimens from the intended use patient population(s). FDA may provide more tailored recommendations for tests for a specific outbreak through separate guidance or pathogen-specific templates, as needed.

Generally, the study size should be determined by the disease prevalence and the number of consecutive patients needed to achieve a minimum of 30 positive and 30 negative individuals representing the intended use population.

In some cases, it might be appropriate for the clinical performance evaluation to evaluate only the most challenging clinical matrix type included in the intended use of the device (e.g., nasopharyngeal (NP) swabs for common upper respiratory types, sputum for common lower respiratory types). For atypical specimen types (e.g., saliva, oral fluid, and buccal swabs for respiratory viruses), the clinical performance evaluation should evaluate each specimen type included in the intended use of the device. For example, for validation of COVID-19 tests for use with sputum and any other typical respiratory specimen, we recommended testing either 30 sputum specimens or a combination of upper respiratory specimens and sputum specimens, such as 15 NP and 15 sputum specimens, or 15 combined upper respiratory specimens and 15 sputum specimens.

In addition, specimens from the same anatomical site but different in collection or transport methods, such as with and without liquid transportation medium, are considered as two distinct types of specimens and should be validated separately. For validation of multiple workflows and/or optional components refer to Section V.B(12).

Further, when a clinical performance evaluation is not a prospective, all-comers clinical study, FDA recommends that manufacturers ensure that their evaluation include samples that appropriately represent the range of pathogen levels expected in clinical specimens. For example, for COVID-19, FDA generally expected evaluation of approximately 20% low positive samples (approximately 25% was recommended for molecular tests and 10-20% was recommended for antigen tests). For these evaluations, FDA generally considered low positives to have a Ct (cycle threshold) value within 3 Ct of the mean Ct at the Limit of Detection (LoD) of the comparator test.

If the test is intended for use with asymptomatic individuals, individuals enrolled in the clinical performance evaluation should be documented as free of any symptoms of the target infection prior to enrollment and sample collection. The study protocol and report should document how individuals were screened and confirm that all enrolled individuals were asymptomatic. Sufficient subjects should be prospectively enrolled to achieve an appropriate number of positives and negatives (both specimen positivity and negativity defined by a comparator test). The total number of subjects needed depends on the prevalence of the pathogen in the intended U.S. population. For example, for COVID-19, FDA generally expected 20 positives and 100 negatives to validate an intended use in asymptomatic individuals following a successful validation of a symptomatic intended use for an EUA. In such a case, since the test would have

already been validated for use on symptomatic individuals, such as with the 30 positive/30 negative study design discussed earlier, validation for use on asymptomatic individuals could be performed with fewer positive samples than the original validation on symptomatic individuals. Obtaining even 20 positive samples from asymptomatic individuals can be challenging given potentially lower analyte prevalence. Therefore, when 20 positives cannot be obtained, enrichment strategies could be considered if prevalence in asymptomatic individuals is low. For example, conducting an additional prospective study in an asymptomatic screening population that is under quarantine due to possible exposure may increase the chances of obtaining more positive specimens. You should consult FDA prior to implementing enrichment approaches in your clinical performance evaluation.

(3) Clinical Data Analysis

FDA generally expects all samples meeting the pre-defined inclusion criteria for the clinical performance evaluation to be included in the analysis. When a sample is excluded from the data analysis, justification should be documented and included in any EUA request.

FDA generally recommends that clinical data analysis include the calculation of positive percent agreement (PPA) and negative percent agreement (NPA) with a highly sensitive comparator method, if available at the time. As stated in Section V.A above, if comparator tests are not available, evaluation of contrived samples could be acceptable for initial test validation. The level of PPA and NPA that helps ensure adequate performance of a diagnostic test depends on the test type and indications for use as well as a benefit/risk assessment in the context of the emerging outbreak. For example, FDA generally expected ≥ 95% PPA and NPA for EUA-authorization of molecular tests during the COVID-19 outbreak. With certain mitigations, lower PPA was generally considered acceptable for certain types of tests. For example, for COVID-19 antigen tests, FDA generally expected a PPA of ≥ 80% and NPA of ≥ 95% for EUA authorization. In some cases, such as for Point-of-Care (POC) or at-home tests, an even lower PPA was generally considered acceptable for authorization, with certain mitigations. For all tests with a PPA lower than 95%, FDA generally expected certain mitigations, such as reporting of negative test results as "presumptive" and recommendations for serial testing. In contrast, in certain cases, such as for screening tests for asymptomatic individuals, a higher NPA (≥ 98%) was expected.

If the test is intended for symptomatic individuals, the data should include time from symptom onset to test for each enrolled subject and the data analysis should include consideration of performance shifts in relation to time from symptom onset.

(4) Human Subject Protection

Studies involving clinical specimens (human specimens) are subject to applicable requirements for Institutional Review Board (IRB) review and approval and informed consent (*see* 21 CFR parts 50, 56, and 812). In December 2023, FDA published a final rule that permits an IRB to waive or alter informed consent requirements for certain minimal risk clinical investigations that meet the conditions in 21 CFR 50.22. FDA anticipates that this new provision may be applicable to certain IVD studies involving clinical specimens (88 FR 88241). In addition, the FDA guidance "Guidance on Informed Consent for In Vitro Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable," describes a policy about informed

consent requirements for certain IVD studies that use leftover, de-identified specimens. Please note that additional requirements in the Investigational Device Exemption (IDE) regulations (21 CFR part 812) may be applicable to certain IVD clinical studies.

B. Analytical Validation Testing

(1) Limit of Detection (LoD) (Analytical Sensitivity)

The LoD provides a measure of the analytical sensitivity of a test for a particular target analyte, and is defined as the lowest concentration of target analyte that is consistently detected by the test in 95% of the specimen replicates.[15] The LoD results guide additional validation studies, including the clinical performance evaluation, as described throughout this document.

LoD should be determined using the entire test system from specimen preparation and extraction through detection and the result interpretation algorithm. For example, tests intended for use with collection swabs placed in Viral Transport Media (VTM) should be evaluated by spiking collection swabs with the target analyte prior to immersing them into VTM and running on the test system. Tests intended for use with dry swabs (i.e., not eluted in liquid specimen transport media) should be evaluated by applying the contrived specimen (e.g., virus spiked into real negative clinical matrix) directly to the swab prior to testing. Tests intended for swab collected specimens with either VTM or dry processing should be evaluated separately and LoD established for both liquid transport media and dry conditions.

In some cases, it may be appropriate to determine LoD only for the most challenging negative clinical matrix type included in the intended use of the device (e.g., NP swabs for common upper respiratory types, sputum for common lower respiratory types). For atypical specimen types (e.g., saliva, oral fluid, and buccal swabs), the LoD should be determined with each specimen type included in the intended use of the device.

In situations where neither live nor inactivated stocks, nor a known positive clinical specimen is available, such as very early in an outbreak, use of synthetic material[16] might be considered for use in the LoD evaluation in real clinical matrix. When synthetic material is used, it should closely mimic the natural target analyte.[17] Simulated or artificial specimen matrix (e.g., clean liquid transport media spiked with mucin, human DNA, and HeLa cells) or recombinant antigen (e.g., for an antigen test) should not be used in an LoD study as this material does not accurately mimic actual patient samples and, therefore, testing with this material may not accurately reflect performance of the device. Developers should discuss potential use of alternative matrices for unique circumstances with FDA. As more specimens become available, FDA generally recommends that LoD be evaluated by spiking individual or pooled natural negative clinical

[15] *See* definition in CLSI EP17-A2 *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures.*

[16] For example, FDA authorized COVID-19 tests that were validated with synthetic full length or long strand RNA through April 2020.

[17] For example, for COVID-19 tests, FDA generally expected synthetic material to consist of full length or long strand RNA, as SARS-CoV-2 is an RNA virus for EUA authorization. In cases where tests were instead validated with synthetic DNA or short fragments of RNA, FDA requested revalidation. Results demonstrated that use of such materials, which did not closely approximate SARS-CoV-2 RNA, over-estimated test performance and masked some unacceptably poorly performing tests.

matrix[18] with well characterized, quantified stocks of the target analyte (live or inactivated), for each clinical specimen type included in the intended use of the device. For example, in lieu of quantified live or inactivated virus (e.g., heat treated, chemically modified, or irradiated virus), a quantified known positive clinical specimen as determined by an FDA-cleared/approved/authorized test could be used to create dilutions in real clinical matrix for the LoD study.

The preliminary LoD should be determined by testing a 2-3-fold dilution series of three replicates per concentration. The lowest concentration at which all tested replicates are positive is considered the preliminary LoD. The preliminary LoD study should include at least one concentration that does not yield 100% positive results.

The LoD should be confirmed by testing a minimum of 20 individual replicates of the concentration determined to be the preliminary LoD. The final LoD is the lowest concentration resulting in positive detection of at least 95% of the replicates (e.g., at least 19 out of 20 replicates). In the case where the final LoD study achieves a positivity of 100%, a lower concentration (using a 3-fold dilution) should be tested (with 20 replicates) until < 95% positivity is obtained.

While the LoD for the entire test system, from specimen preparation and extraction through detection and the result interpretation algorithm, is most critical for test validation, FDA may also request the LoD for individual targets for multi-target tests to help the Agency evaluate the performance of the device.[19]

CLSI EP17-A2 *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures* is recognized by FDA and should be considered where applicable.

(2) Inclusivity (Analytical Reactivity)

An inclusivity study shows reactivity of the test with additional related (e.g., taxonomic, immunological, and genetic composition) target species or isolates. For molecular-based tests, FDA generally recommends 100% nucleotide sequence identity, meaning that the test reagents have no mismatches with known published sequences and therefore can likely detect all known species or isolates. If a test has less than 100% nucleotide sequence identity to a significant number of published sequences, FDA recommends performing a risk assessment on how such mismatches may impact the performance of the test.

Inclusivity of molecular-based tests should be evaluated through *in silico* analysis of the test primer and probes with all known sequence variants (past[20] and present) of the pathogen for each pathogen included in the intended use. The methods and results of this analysis should be

[18] For purposes of this guidance, clinical matrix is defined as a specimen taken from a human subject. Negative clinical matrix is a clinical specimen taken from a human subject which does not harbor the analyte of interest. Liquid transport media without clinical matrix and specimen transport medium included in a collection kit that was not used to collect a clinical specimen are not considered real clinical matrix.

[19] Multi-target tests detect multiple sites for the same analyte. For example, a multi-target SARS-CoV-2 test may detect N-gene, S-gene, and E-gene targets of the SARS-CoV-2 virus.

[20] Certain past variants may contain mutations that can reappear in the future.

documented and should show the extent to which variation in the target genome may affect sensitivity of test performance. *In silico* data should be supplemented with wet testing of currently circulating variants by testing clinical isolates and/or inactivated materials spiked into clinical matrix at or near the test LoD. This may only be possible when materials are widely available.

Inclusivity of antigen-based tests should be evaluated through wet testing of currently circulating variants, such as by testing clinical isolates and/or inactivated materials spiked into clinical matrix at or near the test LoD. This may only be possible when materials are widely available.

FDA recommends evaluation of inclusivity monthly. Test manufacturers should monitor new and emerging and/or clinically significant mutations and variants for their potential to affect test performance. This could include, for example, assessing the prevalence of mutations in well-established publicly available sequence databases (e.g., NCBI) and monitoring for credible reports that a given variant (which may have one or more mutations) has the potential to increase pathogenicity, increase transmission, or otherwise increase the risk to public health. FDA also conducts its own monitoring and may request additional testing, as applicable.

In this example, for any mutations and variants that are identified as prevalent and/or clinically significant, molecular test manufacturers should assess whether the mutations are in nucleic acid regions targeted by the test's primers/probes and antigen test manufacturers should assess whether the resulting predicted amino acid change(s) in the proteins caused by the mutations are critical to the test design. Mutations critical to the test design should be evaluated using clinical specimens to assess the impact of the mutation or variant on test performance. Testing should include both clinical performance evaluation and LoD studies using wet testing with a clinical specimen with the mutation, if available.

The aggregate impact of the mutations should be evaluated and should not result in the clinical performance point estimates for the test dropping below the clinical performance recommendations described in Section V.A.

If a greater than 3-fold reduction in analytical sensitivity is observed when comparing the pathogen harboring the mutation, and not harboring the mutation, you should conduct a risk analysis for the observed decrease in performance, consider further risk mitigations, and assess whether the known and potential benefits of the test continue to outweigh the known and potential risks.

(3) Cross-Reactivity (Analytical Specificity) and Microbial Interference

The purpose of the cross-reactivity evaluation is to establish that the test does not react with related non-target microorganisms, high prevalence disease causing agents, and commensal or pathogenic flora that are likely to be in the clinical specimen. The purpose of the microbial interference study is to establish test performance when the target analyte is present in a clinical specimen with other relevant non-target microorganisms. Cross-reactivity wet testing should be done using samples that *do not* contain the analyte included in the intended use and microbial interference wet testing should be done using samples that *do* contain the analyte included in the

intended use at low concentrations (e.g., ≤ 3-fold of the LoD). Ideally, the study design should incorporate the cross-reactivity and microbial interference validation so that analyte positive and negative specimens can be tested in a randomized and blinded manner.

For molecular tests, cross-reactivity and microbial interference could initially be assessed with an *in silico* analysis of published genome sequences in well-established publicly available sequence databases (e.g., NCBI) using the test primers and probe(s). If the *in silico* analyses reveal ≥ 80% identity between the cross-reactive microorganism(s) or the microbial interferent and the combination of test primers and probe(s) for a given target, wet testing should be conducted with the applicable organism(s). If there is sufficient justification as to why the performance of the test would not be impacted (e.g., due to a limiting number of primer(s)/ probe(s) included in the master mix), wet testing may not be needed.

For antigen tests, *in silico* analysis is generally not appropriate and wet testing should be conducted. Further, for lateral flow immunoassay tests, FDA has observed significant cross-reactivity (leading to false positive results) with different brands and types of VTM, which has resulted in erroneous patient results. As a result, FDA generally does not recommend VTM for use with lateral flow immunoassay tests.

Wet testing should typically use live microorganisms spiked into the most challenging, natural, clinical matrix included in the labeling at high clinically relevant microorganism levels. FDA generally considers a high clinically relevant level to be a minimum of 10^6 CFU/mL or higher for bacteria/fungi and 10^5 PFU/mL or $TCID_{50}$/mL or higher for viruses. It is generally acceptable to test a minimum of 1 strain per microorganism evaluated. Test specimens should either be real clinical specimens or be prepared by spiking cultured isolates into pooled negative clinical matrix. *In silico* analyses alone may be acceptable for certain microorganisms, such as those that are difficult to obtain. If specific microorganisms are not available, we recommend you contact FDA to discuss potential options and labeling mitigations.

If the test will be used with multiple extraction methods and/or multiple instruments, this study should be performed with the *most sensitive* extraction/instrument combination with the best LoD (i.e., the LoD with the lowest analyte concentration). Cross-reactivity and microbial interference should be determined based on using at least three replicate samples. If any false positive or false negative results occur when testing each microorganism using three replicates, then a minimum of 10 additional replicates should be tested. If results indicate cross-reactivity or microbial interference with any of the tested microorganisms, a plan for addressing false results should be provided.

The interferent or potentially cross-reactive microorganisms can be tested individually or as a pool (e.g., a pool of 4-5 microorganisms). If pooling, the concentration of each individual microorganism should be maintained. If a pool shows interference or cross-reactivity, each microorganism of a pool should be tested individually. If interference or cross-reactivity is seen, an additional titration study should be performed to determine the highest microorganism level the test can tolerate.

The non-target microorganisms that should be evaluated for these studies depends on the target pathogen: the target pathogen's genetic family, the disease etiology and symptoms, and how the test will be used, including the clinical specimen(s) used for detection.

Examples of recommended microorganisms to test for cross-reactivity and microbial interference for common respiratory specimens include: Human coronavirus 229E, Human coronavirus OC43, Human coronavirus NL63, Human coronavirus HKU1, MERS-coronavirus (if available), SARS-coronavirus (e.g., SARS-CoV-1, SARS-CoV-2), Adenovirus (e.g., C1 Ad. 71), Human Metapneumovirus (hMPV), Parainfluenza virus 1-4, Influenza A & B, Enterovirus, Respiratory syncytial virus, Rhinovirus, *Haemophilus influenzae*, *Streptococcus pneumonia*, *Streptococcus pyogenes*, *Candida albicans*, Pooled human nasal wash (negative clinical matrix): representative of normal respiratory microbial flora, *Bordetella pertussis*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, *Legionella pneumophila*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Mycobacterium tuberculosis*, *Pneumocystis jirovecii (PJP*), *Pseudomonas aeruginosa*, and *Streptococcus salivarius*.

Examples of recommended microorganisms to test for cross-reactivity and microbial interference for saliva and oral specimens include: Human coronavirus 229E, Human coronavirus OC43, Human coronavirus NL63, Human coronavirus HKU1, MERS-coronavirus (if available), SARS-coronavirus (e.g., SARS-CoV-1, SARS-CoV-2), Adenovirus (e.g., C1 Ad. 71), Human Metapneumovirus (hMPV), Parainfluenza virus 1-4, Influenza A & B, Rhinovirus, Respiratory syncytial virus, Herpes simplex virus type 1 (HSV-1), Epstein-Barr virus (EBV), Cytomegalovirus (CMV), *Moraxella catarrhalis, Porphyromonas gingivalis, Bacteroides oralis, Nocardia sp., Streptococcus mutans, Streptococcus mitis*, or other *Strep viridans, Eikenella sp., Neisseria sp., Candida albicans, Pseudomonas aeruginosa, Staphylococcus epidermidis, Streptococcus salivarius,* and *Lactobacillus sp.*

(4) Endogenous/Exogenous Interference

The purpose of an endogenous/exogenous interference study is to assess the effects of endogenous and exogenous substances on test performance. Endogenous substances include those found at elevated levels in the type(s) of clinical specimens the test will be used with, such as blood in a nasal swab sample. Exogenous substances can sometimes be introduced into specimens before or during specimen collection, such as toothpaste in a saliva sample, including commonly prescribed or over-the-counter clinically relevant medications, treatments, or topical applications for treating symptoms associated with specific infections. This study is designed to demonstrate that a substance does not cause false positive results in specimens known to be negative for the target analyte or lead to false negative results in specimens known to be positive for the target analyte.

The potential interfering substance should be spiked into the most challenging applicable negative clinical matrix, either alone or with acceptable target material at or near the test LoD. FDA generally considers use of pooled negative clinical matrix as acceptable for this study. Live samples of each target analyte included in the intended use is preferred, but use of inactivated stocks or genomic nucleic acid may be acceptable if supported by an LoD study. Positive specimens can be prepared by spiking negative clinical matrix at a challenging concentration (e.g., ≤ 3-fold of the LoD), for example, spiking negative clinical matrix with live virus,

inactivated virus, or viral genomic RNA (if applicable). Please refer to CLSI EP07 (3rd edition) *Interference Testing in Clinical Chemistry* (Section 3.4.2)[21] which references CLSI EP37 *Supplemental Tables for Interference Testing in Clinical Chemistry,* for the recommended concentrations for testing common endogenous substances. Testing in triplicate is recommended. The evaluation should be conducted over the expected clinical range of the potential interfering substance concentrations. If interference is observed during these studies, the interferent should be further tested at serial dilutions to determine the lowest interfering concentration.

Examples of potentially interfering substances for respiratory specimens include: throat lozenges, oral anesthetic, and analgesic (active ingredients Benzocaine, Menthol), Mucin: bovine submaxillary gland, type I-S or pooled mucous (active ingredient Purified mucin protein), Blood (human), Leukocytes, FLUMIST QUADRIVALENT, Zinc (common ingredient in many nasal sprays), Nasal sprays or drops (active ingredients Phenylephrine, Oxymetazoline, Sodium chloride with preservatives), Nasal corticosteroids (active ingredients Beclomethasone, Dexamethasone, Flunisolide, Triamcinolone, Budesonide, Mometasone, Fluticasone), Nasal gel (active ingredients *Luffa opperculata*, sulfur), Homeopathic allergy relief medicine (active ingredients *Galphimia glauca*, *Histaminum hydrochloricum*), Anti-viral drugs (active ingredient Zanamivir), Antibiotic, nasal ointment (active ingredient Mupirocin), and Antibacterial, systemic (active ingredient Tobramycin).

Examples of potentially interfering substances for saliva and oral specimens include toothpaste, tobacco product, oral rinse, and Nicotine.

(5) High-Dose Hook Effect

The high-dose hook effect, where false negative results occur due to the presence of very high levels of the target analyte in the patient specimen, is most commonly an issue for antigen tests. This is particularly applicable in primary sandwich-based immunoassays and secondary sandwich-based immunoassays without wash steps. The hook effect occurs when an excessive amount of target analyte present in the tested specimen interferes with the binding ability of the capture antibody, leading to potential false negative results.

Evaluation of whether a hook effect occurs should be done by testing increasing analyte concentrations. Contrived specimens should be prepared by spiking the most challenging pooled negative clinical matrix with live or inactivated pathogen (e.g., heat treated, chemically modified, or irradiated pathogen). You should evaluate 3-5 replicates per pathogen concentration. If results indicate the test is susceptible to a high-dose hook effect, the lowest concentration where performance is impacted should be identified.

(6) Carry-Over/Cross-Contamination

Many tests utilize automated liquid handling systems to process and test specimens, which can pose a risk of contamination within or between test runs. All workflows (including all

[21] FDA recognizes the importance of updating consensus standards to reflect current knowledge on device performance and safety issues. In general, FDA actively assesses the impact of new consensus standards and revisions of existing standards on the premarket review process and recognizes these standards, as appropriate. For the most up-to-date list of FDA-recognized consensus standards, see the FDA Recognized Consensus Standards Database.

instruments) should be evaluated to determine whether carry-over or cross contamination from high positive specimens could generate false positive results in other specimens. If there is significant manual manipulation of specimens and/or reagents, multiple operators should be used.

The experimental design should be based on risk, taking into consideration all aspects of the workflow, including pre-processing, and run set up. Carry-over specimens should be prepared by spiking live or inactivated pathogen in the most challenging negative individual or pooled clinical matrix with which the test will be used. High positive specimens and negative specimens should be alternated based on the operational function of the device. For example, high concentration and negative specimens should be evaluated in a checkerboard pattern for plate-based assays. At least 5 runs with alternating 8 high positive (prepared at the highest clinically relevant level) and 8 negative specimens should be evaluated. If any false positives are observed, we recommend investigating the source of cross contamination by performing a root cause analysis.

(7) Specimen Stability

Degradation of a specimen prior to testing can lead to false results. The stability of specimens collected and stored should be evaluated in real-world conditions including, for example, the expected environmental conditions at the recommended storage and/or shipping specifications (e.g., temperature and time specifications). Acceptable specimen stability conditions are typically required by the Centers for Disease Control and Prevention (CDC).[22] No further data are likely needed where the specimen stability is based on CDC recommendations; additional or extended specimen stability should be validated in an appropriate specimen stability study.

The study should include several time points throughout the duration of the recommended storage time and at least one time point beyond the stability included in the labeling, as well as temperatures at the upper and lower limits of the recommended temperature ranges for storage and transportation. For example, when storage at room temperature is indicated, both extremes of the temperature range should be evaluated (e.g., 15°C and 30°C). When a test is intended to be performed on the specimen immediately or shortly after obtaining the specimen, the specimen stability testing timeframe should reflect a short storage time (e.g., 2 hours at room temperature).

The study should include contrived positive specimens prepared by spiking live or inactivated pathogen into an individual or pooled negative clinical matrix around the LoD (e.g., 30 replicates at < 2-fold of the LoD and 10 replicates at < 5-fold of LoD) and a minimum of 10 negative specimens. If live or inactivated pathogen are not available, we recommend you contact FDA to discuss potential options.

If a test is intended for use with multiple transport methods (e.g., VTM/UTM, saline, dry swabs), specimen stability should be demonstrated for each method. If a test is intended for use with atypical specimen types (e.g., saliva, oral fluid, and buccal swabs for respiratory viruses), specimen stability should be demonstrated for each specimen type. For this purpose, specimens from the same anatomic site but transported in different ways (i.e., liquid transport media vs. dry

[22] See CDC Infectious Diseases Laboratories Test Directory, available at: https://www.cdc.gov/laboratory/specimen-submission/list.html

swabs, viral transport media vs. saline) are considered different specimen types and each should be evaluated. If a test is intended for use with multiple commonly used specimen types, specimen stability could be demonstrated using only the most challenging specimen type included in the intended use (e.g., NP swabs for common upper respiratory types, sputum for common lower respiratory types).

(8) Reagent Stability

Degradation of the reagents used in a test can lead to false results. The stability of reagents used in a test, such as those that may be shipped as part of a collection kit or test kit, should be demonstrated.

For test kits, the reagent stability studies should be designed to support the shipping and storage conditions outlined in the instructions for use (IFU). This typically includes:

- Evaluation of unopened kits stored at the storage temperature included in the labeling;
- Evaluation of unopened kits when exposed to shipping/transport time and environmental conditions (e.g., temperature, humidity, light exposure and/or environmental factors) expected during normal distribution to end users;
- Evaluation of reagents once the kit has been opened (e.g., storage at 2-8°C for 7 days) and once reagents have been placed on an instrument, if applicable;
- Evaluation of reagents that have undergone the specific number of freeze-thaw cycles[23] indicated as acceptable in the IFU, if applicable.

CLSI EP25-A *Evaluation of Stability of In Vitro Diagnostic Reagents Approved Guideline* is recognized by FDA and should be considered when designing the reagent stability study. In some cases, it may be appropriate to temporarily rely on results from accelerated stability studies to support a six-month shelf life. In such cases, you should seek FDA's agreement on a proposed real-time study design and start the study immediately after agreement to avoid relying on accelerated stability data longer than necessary. Extension of expiration dates can be considered once real-time data becomes available.

(9) Fresh/Frozen Specimens

If the test will be used on frozen specimens or if the clinical performance evaluation used some frozen specimens, it is recommended comparable performance between fresh and frozen specimens should be demonstrated, where applicable. The freeze-thaw conditions tested should reflect the actual conditions (e.g., temperature) expected for frozen archived specimens used in a clinical performance evaluation.

Either natural clinical specimens or contrived specimens could be used for this study. Contrived specimens should be prepared by spiking live or inactivated pathogen into a negative pooled clinical matrix at different levels of pathogen concentration including concentrations close to the test LoD. A minimum of 50 specimens should be evaluated for each sample type (fresh and frozen), taking into consideration both transport methods and clinical matrix as described below.

[23] The number of cycles should be counted following the first thaw of a frozen reagent.

If live or inactivated pathogen are not available, we recommend you contact FDA to discuss potential options.

If a test is intended for use with multiple transport methods (e.g., VTM/UTM, saline, dry swabs), performance with frozen specimens should be demonstrated for each method. If a test is intended for use with atypical specimen types (e.g., saliva, oral fluid, and buccal swabs), performance with frozen specimens should be demonstrated for each specimen type. For this purpose, specimens from the same anatomic site but transported in different ways (i.e., viral transport media vs. saline) are considered different specimen types and each should be evaluated. If a test is intended for use with multiple commonly used specimen types, performance with frozen specimens can be demonstrated using only the most challenging specimen type included in the intended use (e.g., NP swabs for common upper respiratory types, sputum for common lower respiratory types).

Results should demonstrate at least 95% positive agreement between performance of the test with fresh and frozen specimens.

(10) Flex Studies

Flex studies demonstrate the robustness of a test, including the test's ability to maintain performance through environmental and usage variations under conditions of stress. These studies are primarily recommended for home use and point of care test systems. First, a thorough hazard risk analysis should be conducted to identify the most common or likely sources of error based on the use locations and test procedure. Flex studies should be conducted to evaluate the impact of errors, or out-of-specifications conditions, identified in the risk analysis on test performance. In general, the flex studies should be conducted to the point of failure to determine the maximum deviation that will still generate accurate results. If erroneous results are observed during these studies, adequate mitigation(s) should be identified.

Flex studies should include testing negative specimens and low positive specimens near cut-off (e.g., < 2-fold of the LoD) prepared in negative clinical matrix for each condition being evaluated and include three replicates for each condition under evaluation. Flex studies should be conducted with trained operators at an internal testing site. Each study should be performed using a pre-defined study protocol that includes the objective of the study, detailed test procedure, and materials used. Examples of some conditions that could be evaluated as potential user errors and anticipated environmental stresses include, but are not limited to:

- Reading Time: Evaluating test results at multiple reading times four-fold below and three-fold above the recommended reading time. For example, where the recommended read time is 20 minutes, evaluating read times of 5, 10, 15, 20, 30, and 60 minutes, at a minimum.
- Specimen Volume: Evaluating test results at specimen volumes two times below and two times above the recommended specimen volume, and the maximum possible added. For example, where the recommended specimen volume is 10 µL, evaluating specimen volumes of 5, 10, and 20 µL, as well as at the maximum specimen volume. If incorrect results are observed at either 5 or 20 µL, additional testing at 7.5 and/or 15 µL may be appropriate. The amount of diluent/buffer added should be specified in the IFU.

- Specimen Diluent/Buffer Volume: Evaluating test results at diluent/buffer volumes at two times below and two times above the recommended diluent/buffer volume specified in the IFU and the maximum volume. For example, where the recommended buffer/diluent volume is 2 drops, evaluating specimen diluent volumes of 1, 2, 3, 4 drops and the whole bottle.
- Specimen Elution: Evaluating how mixing the swab in elution buffer (or other reagent) affects test results. Evaluating all extremes from not-mixing to vigorous shaking, including generating bubbles and intermediate mixing (e.g., swirling 1 or 2 times).
- Temperature and Humidity: Evaluating test results at temperature and humidity extremes that are likely to occur in the United States (e.g., 40°C and 95% relative humidity (RH) to mimic a hot and humid climate and 5°C and 5% RH to mimic a cold and dry climate).
- Light: Evaluating test results in different lighting conditions that would be expected during use (e.g., fluorescent, incandescent, and natural lighting mimicking the outside environment.)
- Disturbance during analysis: Evaluating the effect of moving the test while it is running. This could include dropping the test while it is being run, moving the test to another surface, unplugging the test, receiving a phone call while the mobile software application is running, etc.
- Device Orientation: Evaluating unique device characteristics, as determined by a robust risk analysis. For example, if the test is intended to be run upright, evaluating the test if it is run horizontally, or vice versa.

Sample size should be sufficient to establish that the tested conditions reliably produce the expected result. Any result that is not expected (e.g., a negative result when testing a positive sample) should be considered a failed result and that test case should be considered a failed test case. Additional information on flex studies may be found in the FDA guidance document "Recommendations for Clinical Laboratory Improvement Amendments of 1988 (CLIA) Waiver Applications for Manufacturers of In Vitro Diagnostic Devices" and CLIA Waiver by Application Decision Summaries.[24]

(11) Usability and User Comprehension

For home collection and home use tests, a usability study should be conducted to ensure lay users can complete all steps of the workflow in an actual or simulated use environment. It may be possible to combine the usability study with the clinical performance evaluation study. We recommend you contact FDA for advice prior to initiating this approach.

Additionally, a user comprehension study should be conducted to assess risks associated with misinterpretation and misuse of test results. This study should evaluate the lay user's understanding and comprehension of critical elements and concepts in the labeling, including the intended use of the test, the IFU, the warnings and precautions, and comprehension of the test results (e.g., positive, invalid, and negative results and the impact of each). The user comprehension study can be conducted as a stand-alone, or as part of the usability validation of the user interface.

[24] Available at https://www.fda.gov/about-fda/cdrh-transparency/clia-waiver-application-decision-summaries

Additional information about conducting usability studies can be found in the FDA guidance document "Applying Human Factors and Usability Engineering to Medical Devices."

(12) Analytical Equivalency

In some cases, for test systems with optional components or workflows (e.g., multiple thermocyclers, multiple extraction methods), an analytical equivalency study may reduce the need to perform clinical performance evaluation with multiple configurations or workflows. Analytical equivalency can be evaluated by performing an LoD study with each configuration. If the configurations are analytically equivalent (e.g., the difference in LoD is within 3-fold for each configuration), then the clinical performance evaluation can be conducted using any of the analytically equivalent configurations.

If one or more configurations are non-equivalent (e.g., more than 3-fold differences in LoD), we recommend conducting the remaining analytical validation and clinical performance evaluation with the configuration having the least sensitive LoD.

An analytical equivalency study can sometimes, depending on the specific change(s) made to the system, also be used to support additional component options that were not evaluated during the clinical performance evaluation (e.g., different collection media, extraction and/or PCR instruments).

(13) Software Validation and Cybersecurity

Test systems that include device software functions[25] that have not been previously cleared/approved/authorized by the FDA should be validated to ensure that:

- The inputs and outputs of the software are appropriate to fulfill the system and assay requirements;
- All expected inputs produce the expected outputs for all functions important for proper test system operation and for defined user needs and intended uses (e.g., verification and validation); and
- The system will be provided to the customer free of defects, or defects will be known and mitigated to an acceptable level (e.g., risk assessment).

The following FDA guidance documents and resources include additional information on software validation and documentation and can be referenced to help support and prepare an EUA request:

- General Principles of Software Validation
- Content of Premarket Submissions for Device Software Functions
- Device Software Functions Including Mobile Medical Applications
- Off-The-Shelf Software Use in Medical Devices
- 21 CFR 820.30

[25] Device software functions are software functions that meet the definition of a device under section 201(h) of the FD&C Act. Device software functions may include software as a medical device (SaMD) and software in a medical device (SiMD).

The cybersecurity[26] of test systems with any external wired and/or wireless communication interfaces (e.g., Wired: USB, ethernet, SD, CD, and RGA; Wireless: Wi-Fi, Bluetooth, Radio Frequency, inductive communication, Near Field Communication (NFC), and Cloud) should be evaluated to ensure user and patient safety in the intended use environment.

(14) Basic Safety and Essential Performance of Instruments

Basic safety hazards such as electrical hazards (e.g., electrical shock to the operator and/or patient), fire hazards, and mechanical hazards should be addressed for test systems that include instrumentation that has not been previously cleared/approved/authorized by the FDA. We recommend you consider International Electrotechnical Commission (IEC) 60601-1 *Medical electrical equipment – Part 1: General requirements for basic safety and essential performance*, which defines basic safety as freedom from unacceptable risk directly caused by physical hazards when medical electrical equipment is used under normal condition and single fault condition.

(15) Electromagnetic Compatibility (EMC) Testing

For test systems that are electrically-powered or have functions or sensors that are implemented using electrical or electronic circuitry and that have not been previously cleared/approved/authorized by the FDA, Electromagnetic Compatibility (EMC) testing should be conducted to ensure the test system can function safely and effectively in its intended electromagnetic (EM) environment, including immunity to EM disturbances (i.e., interference), without introducing excessive EM disturbances (i.e., emissions) that might interfere with other equipment.

FDA partially recognizes International Electrotechnical Commission (IEC) 61326-1 *Electrical equipment for measurement, control and laboratory use - EMC requirements - Part 1: General requirements* and IEC 61326-2-6 *Electrical equipment for measurement, control and laboratory use - EMC requirements - Part 2-6: Particular requirements - In vitro diagnostic (IVD) medical equipment* and recommends using the test methods from these standards. Additionally, we recommend using test levels specified by ANSI/AAMI/IEC 60601-1-2 *Medical electrical equipment – Part 1-2: General requirements for basic safety and essential performance – Collateral Standard: Electromagnetic disturbances – Requirements and tests* or, alternatively, determining the reasonably foreseeable maximum levels of the electromagnetic phenomena in the device intended use environments (e.g., through study of published literature or environmental measurements). Acceptance criteria should be specific to the test system's functions and intended use.

For more information on EMC testing, consult the FDA guidance document "Electromagnetic Compatibility (EMC) of Medical Devices."

C. Predetermined Change Control Plans

[26] See FDA Guidance document "Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions."

Manufacturers seeking an EUA might consider developing a predetermined change control plan (PCCP) for potential future modifications. When a PCCP is included in the initial authorization, changes implemented pursuant to the change plan are considered to be covered by the initial authorization. PCCPs should include the types of anticipated modifications, the steps that will be taken to validate the modifications, and the performance metrics that would be considered an indication of successful validation (e.g., acceptance criteria). All modifications included in a PCCP should maintain the device within the device's intended use. Examples of modifications that might be in a PCCP include adding new instruments and extending the shelf-life/expiration date.

VI. Additional Considerations for Certain Test Types

A. Multi-Analyte Panels

During an outbreak it may be beneficial to have a multi-analyte panel that can detect and differentiate between pathogens that cause multiple diseases with similar symptoms from a single specimen. Taking just one specimen from a patient may help alleviate the need for multiple samplings, which means less discomfort for the patient and faster and more comprehensive results. In addition, multi-analyte tests need fewer supplies, such as swabs and personal protective equipment, and reduce pressure on the supply chain for test reagents.

In general, each analyte of a multi-analyte test should be validated as discussed throughout this guidance. You should also address the potential for cross-reactivity and microbial interference (including competitive inhibition) between the multiple analytes.
Generally, the validation needed for multi-analyte panels depends on several factors, including, but not limited to:

- State of scientific knowledge for each pathogen;
- Whether target analytes have been previously FDA cleared/approved/authorized;
- Whether the test is a modification of a multi-analyte test previously FDA cleared/approved/authorized for other pathogens (e.g., adding a new respiratory pathogen analyte which is the subject of the outbreak to the design of an existing FDA cleared/approved/authorized test);
- Test format (e.g., individual wells used to test for each target analyte or one single well used to test for all target analytes together (i.e., multiplex reaction));
- Types of specimens the test will be used with (e.g., upper respiratory specimens, lower respiratory specimens, or atypical specimen types such as saliva, oral fluid, and buccal swabs for respiratory viruses); and
- Disease prevalence and associated availability of clinical specimens with the target analytes for a prospective clinical performance evaluation.

B. Home Collection Kits

Home collection of clinical specimens can be beneficial during an outbreak because it provides increased patient access to testing and protects others from potential exposure. FDA recommends that developers of home collection kits consider the incorporation of design features that would increase accessibility for users of differing abilities (e.g., vision or hearing deficits) in their device.

Collection kits intended for home use should use only non-invasive specimen collection that requires no specialized training to be safely and correctly performed. The collection device (e.g., nasal swab) should be appropriate for collection of specimens from the intended anatomical site and safe for home use. Collection kits that contain hazardous or irritating materials (e.g., guanidinium salts) are generally not appropriate for home use unless the test has specific safety features to mitigate the risk of patient exposure. The components of the collection kit should be assessed for toxicology and labeling should inform users of the risks associated with use of the kit, as well as any recommendations for personal protective equipment. The IFU should be written for lay users at no higher than a 7th grade reading level, be in the format of Quick Reference Instructions (QRI) that are limited to one to two pages, and include pictures and diagrams to facilitate use by a lay user.

The risk of inadequate specimen collection by a lay user at home should be mitigated. Inclusion of an internal control in the test design can indicate that adequate human specimen was collected and placed into the test for analysis. This may not be necessary in some cases, such as for specimen types that have generally been shown to be appropriate for lay user self-collection (e.g., anterior nasal swabs). The risk of inadequate specimen collection can also be mitigated in other ways, such as video observation of the user by a trained healthcare professional or other design features of the collection device.

The home collection testing workflow starts with distribution of the home collection kit to an individual who then collects and stores a clinical specimen at home using the materials provided. The individual then sends the specimen to a specific CLIA-certified clinical laboratory for testing. Home collection kits can be paired with a single test or multiple tests and validation should support the proposed intended use. Usability, user comprehension, reagent stability, and specimen stability[27] studies should be conducted.

Where home collection kit and test manufacturers separately seek EUAs, a right of reference[28] shared between the manufacturers may help streamline the review process by allowing data from each EUA request (the home collection kit and the assay) to be incorporated by reference into the other.

C. Point-of-Care (POC) Tests

Near-patient or Point-of-Care (POC) tests are intended for use in near patient settings, such as hospitals, urgent care centers, and emergency rooms. POC tests are beneficial during an outbreak because they provide more immediate results compared to testing in laboratories.

[27] The specimen stability study should be designed to simulate home specimen collection and shipping/transport (e.g., storage of specimens before the home user ships the specimen, specimen stored in a mailbox or drop box waiting for pick-up, shipping conditions after pick-up when the specimen is shipped to the testing lab).

[28] A manufacturer that has provided data to the FDA may grant a right of reference to other manufacturers, either broadly or to individual manufacturers, to leverage that data. A right of reference provides a manufacturer the ability to rely upon, and otherwise use, existing information in one regulatory submission for the purpose of supporting a different regulatory submission. In these cases, if the data is applicable to the new manufacturer's test, the new manufacturer may not have to repeat that validation for its submission to FDA, or FDA may recommend only a bridging study. Any manufacturer seeking to leverage data regarding another manufacturer's EUA-authorized assay should obtain a right of reference from that manufacturer.

Clinical performance evaluation of POC tests should be conducted at one or two U.S. sites representative of anticipated real-world settings with four to six operators without laboratory training and representative of intended operators. For example, this may include: using the device in a healthcare setting, such as at a hospital bedside, by non-laboratorian healthcare professionals; in a non-traditional healthcare setting, such as at a school, by untrained users who are not healthcare professionals; or in a temporary testing site setting, such as a tent set up at a non-healthcare workplace, by users who have limited or no training or hands-on experience in conducting laboratory testing. To help support emergency use authorization for use in settings operating under a CLIA Certificate of Waiver, the test should be validated in such settings.

The clinical performance evaluation should include specimen collection and handling, including addition of the specimen to the specimen port/well of the test, both of which could introduce error. Testing should be done in real time immediately after specimen collection. Operators should *only* rely on Quick Reference Instructions and have received no training on how to use the device. The Quick Reference Instructions should be written for untrained users at no higher than a 7th grade reading level, limited to one to two pages, and include pictures and diagrams to facilitate use. As this study is intended to mimic a worst-case scenario, any supplemental materials provided with the device (e.g., a video or a mobile application that can be easily accessed by the user) should not be used in the study.

Clinical performance recommendations are discussed in Section V.A above. In addition to the clinical performance evaluation, the performance of POC tests around the LoD should be evaluated with contrived specimens in real clinical matrix. Testing should include 10 samples near the LoD, and 10 negative specimens per site. All contrived specimens should be blinded, randomized, and tested as part of the normal workflow of the site. Testing should be conducted by untrained operators, each of whom tests at least three positive samples near the LoD and three negative samples. Results that do not match the expected result (e.g., a negative test result from a sample with analyte above the test LoD) should be investigated. Testing should demonstrate positive and negative agreement of at least 95%. If this is not achieved, the LoD should be re-evaluated.

Flex studies, discussed in Section V.B(10), should be conducted to identify the maximum deviation in conditions reasonably expected for the POC settings that will still generate accurate results.

D. Home Use Tests

Tests for home use may be beneficial during an outbreak because they provide increased patient access to testing, typically provide quick results, and can help protect others from potential exposure. In general, a home use test should be simple to perform, and its results should be simple to interpret. Home use tests can be prescription use or over the counter (OTC). Home use tests can also be used in additional non-laboratory settings, such as offices, sporting events, airports, schools, etc., where an individual performs the test themselves, including reading the results. FDA recommends that developers of home collection kits consider the incorporation of design features that would increase accessibility for users of differing abilities (e.g., vision or hearing deficits) in their device.

Tests intended for home use should use only non-invasive specimen collection that needs no specialized training to be safely and correctly performed. The collection device included with the test should be appropriate for collection of specimens from the intended anatomical site and safe for home use. Tests that contain hazardous or irritating materials (e.g., guanidinium salts) are generally not appropriate for home use unless the test has specific safety features to mitigate the risk of patient exposure. The components of the test should be assessed for toxicology and labeling should inform users of the risks associated with use of the test, as well as any recommendations for personal protective equipment.

The risk of inadequate specimen collection by a lay user at home should be mitigated. Inclusion of an internal control in the test design can indicate that adequate human specimen was collected and placed into the test for analysis. This may not be necessary in some cases, such as for specimen types that have generally been shown to be appropriate for lay user self-collection (e.g., anterior nasal swabs). The risk of inadequate specimen collection can also be mitigated in other ways, such as video observation of the user by a trained healthcare professional or other design features of the collection device.

When using smartphone software applications to facilitate use of the test and/or to provide test results, such applications should be simple and easy to interpret (e.g., positive, negative, and invalid). Error messages should be readily understandable, and troubleshooting should be included in the IFU. The display should promote understanding of results and what lay users should do next, including how to care for themselves and when to seek follow up care. The software application should be capable of capturing and transmitting test results and associated diagnostic data when appropriate in accordance with local, state, and federal requirements. Automation, data harmonization, and integration of software in the diagnostic workflow should be optimized to lessen burden on the test user, minimize the potential for data entry errors, and improve the overall quality and utility of data captured. Software applications intended to interpret test results or otherwise function as part of the test system should be included in analytical validation and clinical performance evaluation and validated in alignment with the recommendations in Section V.B(13) of this guidance. The IFU should be written for lay users at no higher than a 7th grade reading level, limited to one to two pages, and include pictures and diagrams to facilitate use. Usability and user comprehension studies should be conducted as discussed in Section V.B(11).

The clinical performance evaluation of home use tests should be conducted at U.S. sites representative of the intended use setting (i.e., that mimic a home use environment) and with users representative of the intended use population (e.g., including different socioeconomic and educational backgrounds and range of ages). Generally, for OTC tests, the intended use patient population includes adults (and older pediatrics) who can perform self-collection and testing, pediatrics who may be able to self-collect and perform the test under supervision of an adult, and younger pediatrics (and some adults) who need their specimen collected and tested by an adult caregiver. Each of these patient populations, covering a broad age range, should be validated appropriately. The entire workflow should be performed by each individual participant including, as applicable, test registration, specimen collection, testing, and results interpretation. Testing sites should be set up in a way that precludes a user from seeing or hearing other users

performing the test (e.g., in separate rooms or areas partitioned with curtains). Specimens collected for use with the comparator methods should be collected by a health care provider.

Clinical performance expectations are discussed in Section V.A above. Flex studies, as discussed in Section V.B(10), should be conducted to identify the maximum deviation in conditions reasonably expected for the home use environment that will still generate accurate results.