Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.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: May 13, 2016

You should submit comments and suggestions regarding this draft document within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to http://www.regulations.gov. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. 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 Heike Sichtig Ph.D., Division of Microbiology Devices at 301-796-4574 or by email at Heike.Sichtig@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of *In Vitro* Diagnostics and Radiological Health
Division of Microbiology Devices

Preface

Additional Copies

Additional copies are available from the Internet. You may also send an e-mail request to CDRH-Guidance@fda.hhs.gov to receive a copy of the guidance. Please use the document number 1500016 to identify the guidance you are requesting.

Table of Contents

Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic.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 responsible for this guidance as listed on the title page.

I. Introduction

FDA is issuing this draft guidance to provide industry and Agency staff with recommendations for studies to establish the analytical and clinical performance characteristics of Infectious Disease Next Generation Sequencing Based Diagnostic Devices for Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers (hereafter referred to as "Infectious Disease NGS Dx devices"). Infectious Disease NGS Dx devices are for use as aids in the diagnosis (identification) of microbial infection and in selecting appropriate therapies. The next generation sequencing (NGS) technology can be used to detect the presence of clinically important pathogenic organisms in human specimens. In contrast to human sequencing diagnostics, infectious disease sequencing diagnostics generally require rapid and actionable results, sometimes within hours, as delayed or incorrect initial diagnoses can result in fatalities. Furthermore, the broad range of specimen types (e.g., urine, blood, cerebrospinal fluid (CSF), stool, sputum, etc.) and the large diversity of the infectious disease agents that can be present in the sample do not allow straightforward pre-analytical, biochemical, or bioinformatics processes. Each unique specimen type may require a different nucleic acid extraction procedure, a different library preparation protocol, and even a different bioinformatics algorithm to generate the final

clinical result. The opportunity for repeat testing is expected to be limited due to a frequently small specimen quantity (e.g., CSF) and the necessity to make a prompt and timely infectious disease treatment decision for the patient.

This draft guidance provides detailed information on the types of data FDA recommends be submitted in support of a Class II premarket submission. This document does not apply to devices that are intended to screen donors of blood and blood components or donors of human cells, tissues, and cellular and tissue-based products (HCT/Ps) for communicable diseases. The inclusion of certain targets (e.g., Hepatitis B, Hepatitis C, HPV and HIV) could elevate the classification of the device to Class III, and FDA encourages you to contact the Agency for additional guidance.[1] In addition, FDA recommends that sponsors contact the Agency prior to undertaking any clinical or analytical validation studies to discuss whether additional recommendations are available due to new advancements in this fast moving field.

FDA's guidance documents, including this draft guidance, 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

During FDA's Microbial Sequencing workshop held on April 1, 2014 (http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm386967.htm), scientific and clinical community leaders emphasized the benefits of regulatory oversight of Infectious Disease NGS Dx devices due to challenges these devices pose to patient management. Similarly, due to rising interest in Infectious Disease NGS Dx devices, the American Society for Microbiology (ASM) held a colloquium on April 13, 2015, entitled "Applications of Clinical Microbial Next-Generation Sequencing," where the group identified the need for oversight as a top challenge.

Input from stakeholders at these meetings stressed that detection and identification of infectious disease organisms, antimicrobial resistance, and virulence markers have progressed from culture-based methods to molecular methods using nucleic acid amplification and hybridization technologies. High-throughput or next generation sequencing has the capability to replace previous methods with a single approach to accomplish what might have required several different tests in the past.

An Infectious Disease NGS Dx device differs from traditional diagnostic devices in that it may be targeted to detect specific organisms or markers and could simultaneously detect

[1] Certain targets are inherently high risk, including some that are insufficiently understood to authoritatively identify the risks, and therefore belong in class III. A device that tests multiple targets takes on the classification of the highest class target.

multiple organisms present in a sample during a single run. This ability calls for an approach that leverages methods from systems science and is described in the following section. Thus, the data and information submitted to support a regulatory submission should be tailored to the specific NGS technology used.

A. Systems Approach for Infectious Disease NGS Dx Devices

This draft guidance is intended for targeted or agnostic (metagenomic) sequencing, to identify the presence or absence of infectious disease organisms, and/or to detect the presence or absence of antimicrobial resistance and virulence markers. For the purposes of this draft guidance document, Infectious Disease NGS Dx devices' capabilities include testing multiple pathogens and markers using targeted (preferential amplification of specific regions that target a specific organism(s) or marker(s) *a priori* by any lab or bioinformatics method) or agnostic (without target bias) approaches in a single sample through a common process, such as: specimen collection, specimen preparation for sequencing, sequencing/chemistry/data collection, data storage, or report of clinically actionable data. Infectious Disease NGS Dx devices are complex systems, mainly due to the diversity of infectious disease agents, different specimen types, and the entire sequencing data pipeline. Similar to the approach FDA uses for other molecular based diagnostic devices, FDA is proposing to use a "one system" approach for the evaluation of Infectious Disease NGS Dx devices – from sample collection through the output of clinically actionable data (see Figure 1). Further, FDA is proposing to use methods from the discipline of systems science[2] to evaluate these devices. This approach will evaluate, in parallel, the system as a whole (including generation of clinically actionable data), and each individual step in the sequencing data pipeline as part of that system, from specimen collection to results report.

Figure 1: Sequencing process for an Infectious Disease NGS Dx.

The solid green box in Figure 1 depicts the areas under FDA's regulatory oversight. It is important to note that Part 6a (databases) and some aspects of Part 6b (genome assembly,

[2] George J. Klir, Facets of Systems Science, Springer; 2nd edition (October 31, 2001); John N. Warfield, "A proposal for Systems Science", Systems Research and Behavioral Science, 20, 2003, pp. 507–520.

genome annotation, genome finishing) displayed on the right side of Figure 1 may fall under regulatory oversight if they are used as part of the data analysis pipeline to generate the final diagnostic report. In this context, a database is an organized collection of data managed by computer software applications that interact with the user, other applications, and the database itself to capture and analyze data. Below are recommendations on the information that FDA would expect to see in a submission for an Infectious Disease NGS Dx device. See Section VI(D), Instrumentation and Software for any device that uses a proprietary database, and Section VIII, Appendix.

B. FDA-ARGOS: FDA dAtabase for Regulatory Grade micrObial Sequences

FDA, in collaboration with various federal agencies, has developed the database entitled "FDA-ARGOS: FDA dAtabase for Regulatory Grade micrObial Sequences; BioProject 231221." To promote a least burdensome regulatory approach for devices that incorporate Infectious Disease NGS Dx technology, FDA proposes the use of an alternative comparator method for clinical evaluation that relies heavily on public databases populated with regulatory-grade target sequences. This database supplies a set of validated regulatory-grade microbial genomic sequence entries which is available at the National Center for Biotechnology Information's (NCBI) website (http://www.ncbi.nlm.nih.gov/bioproject/231221 (update with FDA web portal link)). Regulatory-grade microbial sequences are near complete high quality draft genomes with metadata requirements. For more information see Section VIII, Appendix.

III. Scope

FDA intends to regulate Infectious Disease NGS Dx devices as systems, including all of the components necessary to generate a result. The components of the system generally include: a specimen collection device, instruments, reagents, software (if applicable) used to generate the sequencing library or otherwise prepare the specimen for sequencing, the sequencing instruments along with the associated reagents and data collection elements that generate the raw sequence reads, and the data analysis pipeline (i.e., assembly, annotation, variant calling, as applicable). As an alternative comparator method to existing culture or composite methods, clinically valid identification of pathogens and antimicrobial resistance and virulence markers may be accomplished through FDA-ARGOS database. For more information see Section VIII, Appendix.

The scope of this draft guidance includes Infectious Disease NGS Dx devices that employ targeted or agnostic sequencing approaches. These approaches are as follows:

- *Targeted Infectious Disease NGS sequencing*: Targeted sequencing requires *a priori* knowledge of the target sequence; thus, its scope is limited to specific targets. For the purposes of this document, targeted sequencing refers to preferential amplification of

defined regions that target a specific organism(s) or marker(s) selected for analysis *a priori* by any lab or bioinformatics method (e.g., amplicon sequencing or a k-mer signature database) based on the diagnostic device's intended use. Design of targeted sequencing diagnostics may use database-driven algorithms requiring a robust regulatory-grade sequence database housing reliable genomic target sequences.

- *Agnostic Infectious Disease NGS sequencing*: Agnostic infectious disease sequencing does not use *a priori* knowledge of sequence targets and generally can identify all constituents (e.g., infectious agent(s) or marker(s) of interest; novel, emerging agent(s) or marker(s); microbiota; human background; and contaminants) in a clinical metagenomic sample (direct genetic analysis from a multi organism sample). Agnostic sequencing approaches rely heavily on bioinformatics approaches and expertise needed to enable correct computational analysis to identify sequence targets after wet lab generation of sequence data. Clearance or approval of Infectious Disease NGS Dx devices using agnostic sequencing technology will be evaluated pending the chosen intended use and a panel-based approach (infectious agents grouped together based on specific intended use or public health need, e.g., a Filovirus panel). Validation of genomic sequence target(s) in an agnostic clinical metagenomic sample (without target bias) will require regulatory-grade reference target sequences and near neighbors to make diagnostic calls. Performance metrics – analytical and clinical – on the detection of sequence targets need to be established in the regulatory submission, such as stating the rationale for making only a genus-level call at a specified confidence cut-off and justifying a diagnostic call benefiting the physician and their patient.

An algorithm should be specified to correlate a diagnostic call to an existing regulatory-grade target sequence. Documentation of the locked-down bioinformatics pipeline, including all required steps, from handling the "raw" sequencing data to producing the diagnostic output, should be provided and should demonstrate robustness for clinical microbiology use. This draft guidance does not address discovery of emerging or novel pathogens or other research applications.

Pathogen or marker discovery should not be part of an original regulatory submission. If claims of the Infectious Disease NGS Dx device are sought for identification of emerging or novel infectious agents or detection of emerging or novel resistance and virulence markers, the addition of these new sequence targets should be reported to the Agency at the time of emergence discovery and before diagnostic use. If beneficial to physicians and their patients, claims for initial genus identification can be sought but should be accompanied by sufficient performance data and an appropriate benefit-risk analysis. An understanding of the clinical significance of these emerging or novel sequence targets is often limited, because there is no specific *a priori* knowledge available. However, stakeholders have asserted that information regarding novel or emerging sequence targets may have some value to physicians and their patients in clinical decision-making. In these circumstances there may not be adequate evidence to demonstrate clear clinical significance, but rather evidence towards establishing a likely association. In order to make information available for these novel or emerging

sequence targets, sponsors should submit a rationale for including them in a device submission, detailing the following: 1) the value of conveying information about novel or emerging sequence targets, 2) a description of how the novel or emerging target with associative limitations is reported (e.g., genus call with subsequent validation of emergent/novel infectious agent or marker after discovery), and 3) how information on the novel and emerging sequence targets is effectively communicated (i.e., maximum benefit medical decision-making with minimal risks to patients). After discovery and validation, these novel or emerging sequence targets are known *a priori* and regulatory-grade target sequences should be qualified for clinical diagnostic use. We encourage developers to contact the Agency regarding adding emerging and novel targets to their existing cleared or approved device.

FDA notes that Infectious Disease NGS Dx devices have the potential to detect multiple infectious agents and/or resistance and virulence markers in a single human clinical specimen. To promote a least burdensome regulatory approach for devices that incorporate Infectious Disease NGS Dx technology, FDA proposes the use of an alternative comparator method for clinical evaluation that relies heavily on public databases populated with regulatory-grade target sequences. For this application, FDA has developed FDA-ARGOS (FDA dAtabase for Regulatory Grade micrObial Sequences, BioProject 231221) containing a set of validated regulatory-grade genomic sequence entries. Section VIII Appendix summarizes FDA's framework of a public regulatory-grade microbial reference database. FDA proposes the use of regulatory-grade genomic sequences as an alternative comparator for clinical evaluation. We note that device performance should be established prior to using this alternative comparator. In order to use the alternative comparator method, microorganisms as well as resistance and virulence markers claimed in the intended use or panel (e.g., a Filovirus panel) should be available as regulatory-grade references before clinical evaluation.

This draft guidance is not intended to address devices that utilize any other type of device technology.

This draft guidance recommends the studies that sponsors should conduct to establish the analytical and clinical performance characteristics of Infectious Disease NGS Dx devices for microbial identification and detection of antimicrobial resistance and virulence markers used in conjunction with a patient's clinical presentation and other laboratory tests to aid in the diagnosis of pathogenic microorganism infections. For the assays addressed in this draft guidance, positive results do not rule out potential co-infection with other pathogens. Also, negative results should not be used as the sole basis for diagnosis, treatment or patient management decisions.

The following is an overview of the information that FDA would expect to see in a submission for an Infectious Disease NGS Dx device. Details for each part of the submission are addressed in more detail in the following sections. Below is an overview of such information:

- Device Description
- Device Validation
    - Analytical Performance
    - Instrumentation and Software
    - Clinical Evaluation

FDA encourages sponsors to use the pre-submission program[3] to discuss the premarket submission strategy for their specific device.

IV. Benefit-Risk Analysis

Sources of risks to health associated with an Infectious Disease NGS Dx device include the risk of incorrect identification of a pathogenic microorganism or marker, which can lead to individual and public health consequences. Such risks to health warrant specific consideration in a premarket submission. As previously mentioned, infectious disease sequencing diagnostics carry an absolute need for immediate and actionable results, sometimes within hours, as an incorrect initial diagnosis potentially leads to fatalities.

Infectious Disease NGS Dx devices can be used in conjunction with the patient's clinical presentation and other laboratory tests to aid in the diagnosis of infection. However, potential risks to patient health management decisions associated with Infectious Disease NGS Dx devices persist. Some of these risks could include failure of the device to perform as indicated, leading to inaccurate results or lack of results, and ultimately incorrect interpretation of results by the user. These potential risks may lead to devastating consequences in patient management decisions.

Specifically, false positive identification of a pathogenic microorganism may lead to an incorrect diagnosis with concomitant inappropriate or delayed antibiotic treatment and erroneous patient isolation precautions. Consequently, this may potentially lead to a more serious infection. Additionally, false positive results in the context of a public health emergency could lead to misallocation of resources used for surveillance and prevention. Similarly, false negative results, or lack of results, could lead to failure to provide a diagnosis and correct treatment, or lead to incorrect patient management to prevent transmission of infection.

Additional risks may arise based on the intended use or technological characteristics of the Infectious Disease NGS Dx device. Premarket submissions for clearance or approval of each specific device should address the potential for and consequences of:

[3] The Pre-Submission program is described in the guidance document titled "Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff" found on FDA's website at: http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm311176.pdf.

- Incorrect or missed identification of microbial target.
- Incorrect detection of antimicrobial resistance marker.
- Incorrect detection of virulence marker.
- Inability to differentiate between colonization and infection.
- Missed identification of contaminant (defined based on intended use).

The guidance document entitled "Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approvals and De Novo Classifications" (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm267829.htm) provides information on FDA benefit-risk determinations. Premarket submissions should include a discussion of the potential benefits and risks associated with the device that is being assessed, the analytical strengths and weaknesses of the technology, and the clinical information that is available demonstrating device effectiveness.

V. Device Description

You should include the following descriptive information to adequately characterize your Infectious Disease NGS Dx device.

A. Intended Use

Intended use applies to targeted and agnostic sequencing approaches and should specify: sequence target or group of sequence targets (infectious agent, antimicrobial resistance and virulence markers), the nature of the target detected (e.g., RNA, DNA, or both), sequencing technology specimen types, the clinical syndrome, and the specific population(s) for which the test is intended. The intended use also should specify any specific conditions of use and state that the identification or detection of a sequence target is presumptive.

In your submission, you should clearly include the following information related to the intended use of your product:

- The identity, phylogenetic relationship (if applicable), or other recognized characterization of the sequence targets (pathogens or genetic markers) that your device is designed to identify or detect.
- How the device results might be used in a diagnostic algorithm.
- Additional measures that might be needed for a laboratory identification and diagnosis of the infection.
- Additional measures that should be instituted if infection with a novel or emerging infectious agent is suspected based on current clinical and epidemiological screening criteria.

B. Test Methodology

You should describe in detail the methodology used by your device. You should describe, at a minimum, the following elements, as applicable to the device:

- Sequencing strategy (i.e., targeted Infectious Disease NGS sequencing or agnostic Infectious Disease NGS sequencing).
- Information and rationale for selection of strategies for: (1) preferential amplification of specific regions that target a specific organism(s) or marker(s) *a priori* by any lab or bioinformatics method (e.g., probe design), (2) sequencing protocol and bioinformatics algorithm from "raw" sequence data to clinically actionable data for agnostic approach (without target bias).
- Description of sequencing technology
- Specimen collection and handling methods (e.g., swabs, viral culture media, positive blood culture, stabilization, etc.).
- Specimen matrix (e.g., blood, sputum, stool, etc.).
- All pre-analytical methods and instrumentation for collection, stabilization, and concentration of specimens.
- Specificity of the claimed sequence targets detected (i.e., methodologies used in addition to the evaluation of clinical specificity to demonstrate that the target sequence is found only in the infectious agent or viral and resistance marker of interest).
- Limiting factors of the Infectious Disease NGS Dx device (e.g., saturation level, de-multiplexing, number of indices, etc.).
- Reagent components provided or recommended for use and their function within the locked down system (e.g., buffers, enzymes, bar codes, sequencing reagents, oligonucleotides, other signaling or amplification reagents, etc.).
- The potential for specific and non-specific interference effects from reagents or device material.
- Internal controls and a description of their specific function in the system.
- External controls recommended or provided to users.
- Instrumentation necessary for using the device, including the components and their function within the system.
- The computational path from raw data to the reported result (e.g., how raw signals are processed and converted into a clinically actionable result). This would include sufficient software controls for identifying and dealing with visible problems in the dataset. It would also include adjustment for background noise and normalization, if applicable.
- Illustrations, photographs, and a detailed description of non-standard equipment or methods, if available.
- Design inputs and outputs with a risk analysis and traceability matrix.

When applicable, you should include descriptions of design control specifications that address or mitigate risks associated with Infectious Disease NGS Dx devices. For example, design controls may be needed:

- To prevent cross-contamination of samples during the sample indexing and bar coding process.
- For manufacturing process procedures that may affect quality.
- To minimize false positive results due to contamination or carryover of samples.
- To enable detection of emerging variants due to mutations within the target organism.
- To detect and correct long term signature instability in device performance due to inherent genetic drift or selective pressure.

C. Ancillary Reagents

Ancillary reagents are those reagents that a manufacturer of Infectious Disease NGS Dx devices specifies in device labeling as "required but not provided." These reagents are required in order to carry out the assay as indicated in its instructions for use and to achieve the performance characteristic claimed in the device labeling. For the purposes of this document, "specific ancillary reagents" are those that the sponsor specifies. Specifics should include a catalog, product number or other designation as necessary for the device to achieve its labeled performance characteristics. For example, for the purposes of this document, if the device labeling specifies the use of Brand X or other amplification enzyme that has been cleared by FDA for this use in this specific device, then Brand X DNA amplification enzyme is a specific ancillary reagent.[4] Moreover, the use of any other DNA amplification enzyme may alter the performance characteristics of the device from that reported in the labeling. For example, the ancillary reagent is general if the device requires the use of 95% ethanol and any type of 95% ethanol will allow the device to achieve the performance characteristics provided in the labeling.

If the instructions for use of the device specify one or more specific ancillary reagents, you should describe in detail how you will ensure that the results of testing with the device and these specific ancillary reagents are in accordance with the instructions for use. In this context, results should be consistent with the performance established in your application for premarket submission. Your plan may include application of quality systems approaches, product labeling, and other measures.

FDA will evaluate whether the plan will help mitigate the risks presented by the device to offer reasonable assurance of the safety and effectiveness of the device and establish its substantial equivalence. Your plan should contain the following elements in detail:

[4] Even if you establish that one or more alternative ancillary reagents may be used in the assay, each of those named alternatives may still be an ancillary reagent. If you are unsure whether this aspect of the guidance applies to your device, we recommend that you consult with the Agency.

1. A risk assessment addressing the use of specific ancillary reagents. This should include:

    - Risks associated with Management of reagent quality and variability.
    - Risks associated with Inconsistency between instructions for use that come with the specific ancillary reagent and your instructions for use of that specific ancillary reagent.
    - Risks you have encountered for the device.
    - Any other issues that could present a risk of obtaining incorrect results with your device.

2. A description of how you intend to use your risk assessment to mitigate risks through implementation of any necessary controls over ancillary reagents should be addressed using your risk assessment. These may include, where applicable:

    - User labeling to assure appropriate use of ancillary reagents.
    - Plans for assessing user compliance with labeling instructions regarding specific ancillary reagents.
    - Plans for alerting users in the event of an issue involving specific ancillary reagents that would impact the performance of the Infectious Disease NGS Dx.
    - Material specifications for specific ancillary reagents.
    - Identification of reagent lots that will allow appropriate performance of your device.
    - Stability testing.
    - Complaint handling protocols.
    - Corrective and preventive actions.
    - Any other issues that should be addressed in order to assure safe and effective use of your device in combination with named ancillary reagents, in accordance with your device's instructions for use.

In addition, you should submit testing data with your regulatory submission to establish that the quality controls you supply or recommend are adequate to detect performance or stability problems with the specific ancillary reagents.

For questions regarding identification, use or control of specific ancillary reagents, contact FDA for advice.

D. Controls

FDA recommends that you run appropriate controls every day of testing for the duration of the analytical and clinical validation studies. This includes any positive and negative

controls intended for use with the assay as well as appropriate external controls recommended, but not necessarily provided, for the assay. In a clinical laboratory setting, laboratories are responsible for following their state and local regulations for running appropriate controls.

Ideally, the assay should include external controls for each target identified by the device. Due to the high number of targets detected by Infectious Disease NGS Dx devices, a rotating control scheme may be considered whereby a panel of representative control organisms or markers (reflective of each claimed organism in the assay menu) is designed and used throughout the evaluation process. External controls should also monitor the organism extraction for each assay run, if applicable. For example, one organism can be used for multiple days, then the next organism for multiple days as long as the stability studies substantiate the time frame of use. Controls should approximate the composition and quality of a clinical specimen in order to adequately challenge the system.

FDA recommends that you provide the following information about the calibrator and control materials, if applicable:

- The nature and function of the various controls included with, or recommended for, the system. These controls should enable the user to determine if all steps and critical reactions have proceeded properly without contamination or non-specific interference.
- Protocols and acceptance criteria for value assignment (relative or absolute) and validation of control and calibrator material.
- The control parameters that could be used to detect failure of the instrumentation to meet required specifications.
- The library calibration controls for cross-talk matrix generation, phasing and pre-phasing.

The daily run of external positive and negative controls during the analytical and clinical studies are used to monitor the ongoing performance of the entire testing process. External controls should be designed to cover low diversity samples and unbalanced genomes. Controls should provide information about: (1) specimen quality, (2) nucleic acid quality, and (3) process quality. FDA generally recommends that you include the following types of controls: negative controls, positive controls, and internal controls.

(1) Negative Controls

Blank or No-Template Control (NTC)

The blank or no-template control contains buffer or specimen transport media and all of the assay components except nucleic acid. This control is used to rule out contamination with target nucleic acid or increased background in the amplification reaction. Negative controls

should be run at a justifiable frequency (i.e., number per shift, day, week) as determined by the laboratory in keeping with state and local recommendations to control for contamination. FDA recommends you run blank or NTC controls with any multiplexed Infectious Disease NGS Dx device to determine bleeding of indices. NTC controls provide a mechanism of tracking the evolution of background laboratory contamination.

Negative Specimen Control

The negative specimen control contains non-target nucleic acids. It reveals non-specific detection and indicates that signals are not obtained in the absence of target sequences, when applicable. Examples of acceptable negative specimen control materials could include:

- Patient specimen from a non-infected individual that has been tested to exclude any of the pathogens detected by the Infectious Disease NGS Dx device.
- Specimens containing a non-target organism.
- Surrogate negative control (e.g., packaged RNA).

(2) Positive Controls

Positive Control for Complete Assay

The positive control is designed to mimic a patient specimen, contains target nucleic acids, and is used to control the entire assay process, including nucleic acid extraction, amplification (when applicable), and detection. Positive controls are run as a separate assay, concurrently with patient specimens. For the clinical and analytical studies, FDA recommends running a minimum of one positive and one negative external control daily during the evaluation. Positive controls can be a subset of the larger assay menu and can be rotated through a pre-defined schedule. In the case of a single use/test consumable with an internal control, periodic external control testing may need to be performed with every new lot, taking into consideration state and local recommendations. If you find that different specimen types require different sample processing, each processing method should be represented by the controls for each day of use. Some examples of acceptable external positive assay controls include:

- Attenuated viral or bacterial vaccine strains.
- Low pathogenic virus or bacteria.
- Inactivated virus or bacteria. Note that some inactivation strategies (i.e., irradiation) provide very poor positive controls due to nucleic acid shearing.
- Packaged RNA/DNA containing target sequences (as appropriate).

Positive Control for Amplification or Detection

The positive control for amplification or detection can contain purified target nucleic acid near the limit of detection for a qualitative assay. It controls the integrity of the device and the reaction components when negative results are obtained.

(3) Internal Controls

The appropriateness of the internal control will depend on the nature of the control and how it is used. The workflow of Infectious Disease NGS Dx devices is complex and involves several steps where significant sample loss, modification, or contamination could occur. It will be important to identify all the sources of sample loss, modification, or contamination in the complex workflow and identify internal controls that allow a determination of whether any of those have occurred. An internal control is usually a non-target nucleic acid sequence that is co-extracted and co-amplified (when applicable) with the target nucleic acid. It controls for integrity of the reagents, equipment functionality, and the presence of inhibitors in the specimen. Use of these internal controls is specifically critical for Infectious Disease NGS Dx devices to control the potential for cross-contamination of samples during the sample indexing, bar coding process, library preparation, and sequencing procedure. An example of an acceptable internal control material includes a packaged non-target genome that is spiked in at sufficient concentration to each clinical specimen before any pre-analytical steps and is analyzed simultaneously with the clinical targets.

FDA recommends that you consult with the Agency when designing specific controls for the device, including the selection and design of control constituents. Sponsors should use the assay's sequencing/data analysis pipeline with all controls.

E. Interpreting Test Results and Reports

In the premarket submission, you should describe the computational pipeline from raw sequencing data generated to positive, negative, indeterminate or invalid organism or marker identification in the final report. You should provide:

- Identity of software packages, databases and versions used in the pipeline.
- Computation method and cut-off values for calling an organism positive or marker identification.

If applicable, please also provide the following:

- If the interpretation of results involves re-testing, you should provide the following information: (1) a recommendation whether re-testing should be repeated from the same nucleic acid preparation, a new extraction, or whether a new patient specimen should be obtained and tested, and (2) an algorithm for defining a final result by combining the initial result and the results after re-testing. Note that this algorithm

should be developed before the pivotal clinical study that evaluates the clinical performance of the assay.

- If the assay has an invalid result, you should provide a description of how an invalid result is defined. If internal controls are part of the determination of invalid results, you should provide the interpretation of each possible combination of control results for defining the invalid result and include recommendations for how to follow up any invalid result (i.e., whether the result should be reported as invalid or whether re-testing is recommended). If re-testing is recommended, you should provide information similar to that for the re-testing of indeterminate results (i.e., whether re-testing should be repeated from the same nucleic acid preparation, a new extraction, or a new patient specimen).

For agnostic sequencing, you should address if and how results for near neighbors and emerging or novel pathogens are reported. Emerging or novel pathogens are not part of the initial premarket submission, but part of the pathogen discovery process. However, FDA notes the importance of capturing this information and suggests contacting the Agency to discuss current policy.

VI. Device Validation

FDA recommends that you contact the Agency prior to undertaking any clinical or analytical validation studies to discuss whether additional recommendations are available due to new advancements in this fast moving field.

It is essential to evaluate standard sets of pre-analytical and analytical protocols for optimization and operation of Infectious Diseases Dx devices in a clinical setting. This is important because there are many variables that markedly influence the performance characteristics of these devices. As previously stated, FDA is proposing to conceptualize an Infectious Disease NGS Dx device as one "system" and to use methods from systems science to evaluate these devices.

Your performance claims established through premarket testing reflected in your device's labeling should be based on the particular test configuration described in the labeling, including all pre-analytical steps. In addition, if your product labeling indicates the use of multiple extraction methods, the premarket performance testing submitted to support the submission should use all extraction methodologies specified in the instructions.

Evaluation of assay performance should include appropriate controls for the duration of the analytical and clinical studies. This includes any internal assay controls as well as the appropriate external controls recommended by the manufacturer, but not necessarily provided with the assay. Furthermore, the evaluation should show how these controls were tested during the analytical and clinical studies.

In the submission, you should detail the study design used to evaluate each of the performance characteristics outlined below.

A. Pre-analytical Factors

Consideration of pre-analytical factors is critical for an Infectious Disease NGS Dx. In your submission, you should clearly address the following issues regarding pre-analytical factors.

(1) Specimen Collection and Handling

Performance of an Infectious Disease NGS Dx device is highly dependent on the quality and quantity of the isolated nucleic acid, therefore specimen type, collection methods, and storage plays important roles in reaching successful results. For the purposes of this document, there are three main specimen types associated with Infectious Disease NGS Dx: (1) clinical isolates, where microbes are grown as pure clonal cultures on a defined media; (2) enriched complex cultures, such as blood tube culture; and (3) direct human clinical specimens, where potential infectious agents may be present in a complex environment, potentially with commensal organisms or host background.

You should validate all specimen types for which the Infectious Disease NGS Dx device is intended to be used. Appropriate specimen types depend on a variety of factors, including the site of infection and the infectious agent or resistance and virulence marker nucleic acid to be detected. Specifically, a clinical specimen should be collected from the appropriate anatomical site or source at the appropriate time in the clinical progression of disease. Appropriate specimen types will vary according to clinical syndrome. Many different specimen types have the potential to be used for validation studies and we suggest that you consult FDA to determine which specimens are considered appropriate for the device platform's intended use, and if certain specimen types could be considered equivalent and combined.

The quality and quantity of extracted nucleic acids can be affected by multiple factors such as specimen source, collection method, and handling (e.g., transport, storage time, temperature). The acceptance criteria for all specimen stability parameters should be clearly indicated and justified and should include the following:

- Validation of any nucleic acid extraction method to be indicated for use with the system.
- Validation that sample collection methods provide adequate and appropriate nucleic acid for all sequence targets detected by the Infectious Disease NGS Dx device, if applicable.
- Validation that the device maintains acceptable performance under all specimen handling conditions claimed.

Specimens for pathogen identification and antimicrobial or viral marker detection should be collected and handled using all applicable state and federal biosafety guidelines. For standard precautions for handling of specimens, refer to the most current editions of the related Clinical and Laboratory Standards Institute (CLSI) documents.[5]

Prior to any signal generation through an Infectious Disease NGS-based technology, the nucleic acid should be prepared. Given the significant differences sample preparation methods can entail and the impact on the overall performance that they could have, validation data should be provided for each method used with the assay. We note that extraction kits should be properly labeled as in vitro diagnostic devices, including meeting the requirements of Section 502(a), (c), and (f) and Section 519(f) of the Federal Food, Drug, and Cosmetic Act (the FD&C Act) as well as the implementing regulations found at 21 CFR Parts 801 and 809 and registered and listed in accordance with the requirements of 21 CFR Part 807. If you have questions, you may discuss them with FDA.

(2) Specimen Preparation for Sequencing

Following specimen collection or clonal isolation, nucleic acid extraction and purification represent the next steps in the process. A number of methods are available for preparing purified nucleic acids and there are several commercially available kits, as well as automated systems. The integrity and purity of the extracted nucleic acids are especially important for successful identification and detection by an Infectious Disease NGS Dx device. However, we note that determination of integrity and purity of the extracted nucleic acids may not be possible for metagenomics specimen types. Similarly, the presence of inhibitors and interfering substances can impact the performance of Infectious Disease NGS Dx devices. However, successful sequencing results depend on the availability of a sufficient amount and quality of sample material for the specific type of sequencing. You should submit your minimum requirements and cutoffs used as quality control for your material, including but not limited to the following factors:

- Sample Amount (µg)
- Sample Volume (µl)
- Concentration of nucleic acid in the sample (ng/µl) (typically used to address unspecific loss)
- Quantification Method
- 260/280 ratio
- 260/230 ratio
- Agarose gel
- Total ng of nucleic acid

[5] Biosafety in Microbiological and Biomedical Laboratories 1999. Richmond, J.Y. and McKinney, R.W. eds., HHS Publication Number (CDC) 93-8395; and CLSI. Protection of Laboratory Workers from Infectious Disease Transmitted by Blood, Body Fluids, and Tissue. CLSI document M29-A. Wayne, PA: Clinical and Laboratory Standards Institute; 1997.

The next major step in the process involves the preparation of the nucleic acids for sequencing, and library preparation. This step can also be performed using a number of different methods. While the exact methods employed are usually platform specific, they tend to share many similar features. For each Infectious Disease NGS Dx device platform, the methods employed during this step should be well documented and in a final "locked-down" configuration for manufacturing.

For devices that use library preparation methods, the sponsor should address the variability on assay performance for all claimed preparation methods and reagents used. Different library preparation methods may yield nucleic acids of varying quantity and quality, and thus the preparation method can be crucial to a successful result. Steps involved in the construction and normalization of the specimen libraries should be considered, which could impact the reproducibility and reliability of the sequences generated (e.g., sample enrichment, sequencing strategy, primers, amplification efficiency, reagent lots, hybridization, etc.). Moreover, an analysis of potential inhibitors from the clinical specimen or methods employed to extract the nucleic acids should also be considered during the validation of a library preparation.

(3) Sequencing, Chemistry and Data Collection

Infectious Disease NGS Dx device's platforms employ a number of sequencing mechanisms, including, but not limited to: sequencing by synthesis that is based on DNA polymerase dependent methods such as cyclic reversible termination (CRT), single-nucleotide addition, and real-time sequencing; sequencing by ligation (SBL) that uses DNA ligase; and single molecule sequencing without prior amplification.

The majority of Infectious Disease NGS Dx device platforms use optical-based imaging for detection, measuring either bioluminescent or fluorescent signals generated when labeled nucleotides are sequentially incorporated into the template. In addition, there are platforms that use non-optical methods for detection, such as the ion-sensitive field-effect transistor semiconductor chip.

The common feature across all of these technologies is that they generate sequences of multiple DNA\RNA fragments in parallel that comprise the sequencing reaction. A minimum number of quality metrics should be applied to Infectious Disease NGS Dx device to evaluate the performance of the instrument runs and quality of the data generated. Specific recommendations on metrics to be submitted are provided in Subsection B – Infectious Disease NGS Dx Device Performance Metrics.

(4) Data Storage

The sequences of the multiple DNA\RNA fragments that comprise the signal outputs of the reaction should be stored in a suitable format that allows subsequent bioinformatics analysis.

There are a number of data formats applied to Infectious Disease NGS; however, the most common are the text-based formats FASTA (stores the biological sequence format used to search National Center for Biotechnology Information BLAST database) and FASTQ (stores both the biological sequence and its corresponding quality scores). All data generated for the evaluation of your Infectious Disease NGS Dx device during the review process should be securely stored and kept on file using clear naming conventions for data sets. FDA may request such data for independent verification purposes.

(5) Clinical Call Determination

The informatics package or data analysis pipeline provided by the manufacturer for use with the sequencing platform is the final step in the process to obtain clinically actionable data. It is important to note that the data analysis pipeline should be in a "locked-down" configuration prior to device validation.

The breadth of the data analysis pipeline ranges from detection of a signal indicating the presence of a specific nucleotide to a final call based on the sequence targets. This analysis relies heavily on informatics components that are intrinsic parts of the analysis pipeline and requires supporting validation data. This could include information from the following specific areas of the data analysis pipeline:

- Signal to base call transformation.
- Alignment via classical sequence alignment methods or via statistical analysis of k-mers (short subsequences of length k).
- Clinical call determination (algorithm and specific clinical regulatory-grade measurement used to determine identification or detection of sequence targets claimed in the intended use or panel).
- Database, if applicable (e.g., databases such as FDA-ARGOS – FDA dAtabase for Regulatory Grade micrObial Sequences (BioProject 231221) discussed in Section VIII Appendix).

For the signal to base call transformation component, the platform pipeline, including base caller and version, and the quality score rationale should be provided.

For alignment and mapping to regulatory-grade reference target sequences (single or multiple), a protocol outlining the steps of the "locked-down" pipeline from raw sequence data (i.e., reads) to the actionable final target sequences should be provided. The protocol should list the specific alignment and mapping tool, version and parameter settings, and the reference sequence with adequate source information. Any emerging and novel sequence targets should be qualified as clinically actionable and as regulatory-grade target sequences before inclusion in a device's intended use or a panel. Note that sequencing a novel microbial genome for the first time (de novo sequencing) is part of pathogen discovery and hence, not within the scope of this draft guidance document.

Detailed information on the algorithm and version used should be submitted for the clinical call determination for validation studies, including genomic coverage requirements, trimming logic and other potential factors. There may be different considerations for how a clinical call is made and validated. These determinations depend on the sequencing format, ranging from a targeted enrichment approach to an agnostic sequencing approach. All assay specific software optimization should be addressed and properly validated. We encourage early discussions with the Agency on these issues.

B. Infectious Disease NGS Dx Device Performance Metrics

FDA recommends including the following items in your premarket submission.

(1) Infectious Disease NGS Dx Device Data Sets

A detailed description of all analytical and clinical data sets used for validation, including study protocols for sample collection, when applicable, should be provided. Also, all processed data used directly in the clinical determination should be included.

(2) Sequencing Strategy

A detailed description of your sample processing, NGS library construction, library quantitation/validation, and whether a targeted or agnostic Infectious Disease NGS approach is applied should be provided.

(3) Selected Targets and Reference Sequences Used for Target Identification

A detailed description regarding each selected target(s) as well as target(s) used that are publicly available or are proprietary reference target sequence(s) should be provided. For each publicly available or proprietary reference sequence used in clinical determination, please provide the following metrics:

- NCBI accession ID
- Expected size of the genome and feature
- Number of contigs
- Number of open reading frames (ORFs)
- Estimated percent of genome covered

We recommend using regulatory-grade genomic reference sequences for the identification of targets and the development of Infectious Disease NGS Dx devices. Regulatory-grade microbial sequences are near complete, high quality draft genomes with metadata requirements (see Section VIII Appendix). Microbial quality metrics are organism specific

and several community guidelines outline recommendations for "High Quality" status [6,7]. The regulatory-grade genome sequences should provide sufficient coverage for the assay's indication for use. Regulatory-grade microbial reference genome sequences require coverage at a minimum of 20X over 95% of the core genome at Phred like quality score[8] => Q30 or provide an adequate justification for why a lower level of coverage is acceptable. FDA has developed the database entitled "FDA-ARGOS – FDA dAtabase for Regulatory Grade micrObial Sequences (BioProject 231221)" that supplies a set of validated regulatory-grade microbial genomic sequence entries which is available at the NCBI's website (http://www.ncbi.nlm.nih.gov/bioproject/231221 (update with FDA web portal link)). If a reference sequence for the development of your target is not available, you should contact FDA to discuss further steps. The appropriateness of the reference sequence(s) is dependent on the design of the assay and contingent on the microbial organism, and should be determined with input from FDA.

(4) Clinical Call Informatics Pipeline

A description of the "locked-down" informatics pipeline should be submitted detailing programs, parameters and reference databases used from signal generation to clinical call determination (e.g., positive, negative, indeterminate). This should include the mapping algorithm settings (e.g., percent of read matching regulatory-grade genomic reference sequence) and percent identity setting for each pathogen/marker target. We recommend providing diagrams/ pictures displaying the flow of information.

(5) Subtraction Rationale

If applicable, details should be provided in narrative form on how the genetic material of the infectious agent(s) or resistance and viral marker(s) of interest is accurately separated from the genomes of the host and other microbes, either physically or bioinformatically, if applicable.

(6) Quality Controls

Evaluation of assay performance should include appropriate controls for the duration of the analytical and clinical studies. The results should also include any positive and negative controls provided with your assay as well as appropriate external controls recommended but not necessarily provided with the assay. If a rotating control scheme is used throughout the evaluation process, the results should be presented for each control panel member.

---

[6] Ladner et al., "Standards for Sequencing Viral Genomes in the Era of High-Throughput Sequencing, mBio," June 17, 2014: Vol. 5 no. 3.

[7] Chain et al., "Genome Project Standards in a New Era of Sequencing, Science," October 9, 2009:Vol. 326 no. 5950 pp. 236-237.

[8] Ewing B., Green P. (1998): "Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res.," 8 (3): 186–194.

(7) Sequencing and Read Mapping

Summary information and statistics on sequence run per sample should be provided, which should include:

- Narrative on trimming and filtering logic (e.g., minimum Q-Score, minimum length, etc.).
- Total number of reads generated.
- Total number of unique reads generated.
- Range of read length.
- Total number of mapped reads and percent identity.
    - Per target mapped reads and percent identity.
    - Per target positive, negative or indeterminate clinical call designation. (Note: Coverage requirements are dependent on specific assay intended use and typically will vary depending on infectious agent(s) or resistance and viral marker(s) of interest, specimen type and the read quality and number of reads generated. Typically, the range is a minimum of 30-200x coverage for specific infectious agent or maker signatures in an Infectious Disease NGS Dx device using culture isolate sequencing. The appropriateness of the target coverage requirements is dependent on the design of the assay and can be discussed with FDA.)

For agnostic NGS sequencing, additional metrics should be provided detailing the number of reads and percent identity of infectious agent(s) and marker target(s) to a qualified regulatory-grade genomic reference sequence. Additional metrics should be provided after human host reads are subtracted, if clutter mitigation is applied. Details and definitions concerning nonpathogenic microbiota, contaminants and controls should be included in performance metrics pending device's intended use.

(8) Contaminant Analysis

A detailed description of how potential contaminants (e.g., carry-over, read misidentification due to barcode demultiplexing) and mitigation procedures are identified should be provided. And, outline the mitigation procedures.

(9) Sample to Result Turn-Around Time (TAT)

The sample-to-result turn-around time should be provided for the Infectious Disease NGS Dx device. Please include data demonstrating turn-around times for "locked-down" Infectious Disease NGS Dx for both: (1) laboratory workflow from clinical sample to sequence, and (2)

the subsequent computational analysis of Infectious Disease NGS data from sequence to actionable clinical result.

(10) Data Storage

All the data used during the evaluation process should be provided. This data should be kept on file and available upon request by FDA.

Below are additional references and resources to help guide assay development and to provide more in-depth information on performance metrics for Infectious Disease NGS Dx devices consistent with FDA's current thinking on regulating these devices:

- "FDA-ARGOS – FDA dAtabase for Regulatory Grade micrObial Sequences (BioProject 231221)" (http://www.ncbi.nlm.nih.gov/bioproject/231221 (update with FDA web portal link)) that supplies a set of validated regulatory-grade genomic sequence entries.
- CLSI Molecular Methods Standard MM09-A2[9]
- Ladner et al., Standards for Sequencing Viral Genomes in the Era of High-Throughput Sequencing[10]
- Chain et al., Genome Project Standards in a New Era of Sequencing[11]

The flow chart below depicts the studies which support targeted and agnostic sequencing approaches. Detailed descriptions of analytical and clinical studies are addressed in the following sections.

[9] Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline—Second Edition (MM09-A2).

[10] Ladner et al., Standards for Sequencing Viral Genomes in the Era of High-Throughput Sequencing, mBio 17 June 2014: Vol. 5 no. 3.

[11] Chain et al., Genome Project Standards in a New Era of Sequencing, Science 9 October 2009:Vol. 326 no. 5950 pp. 236-237.

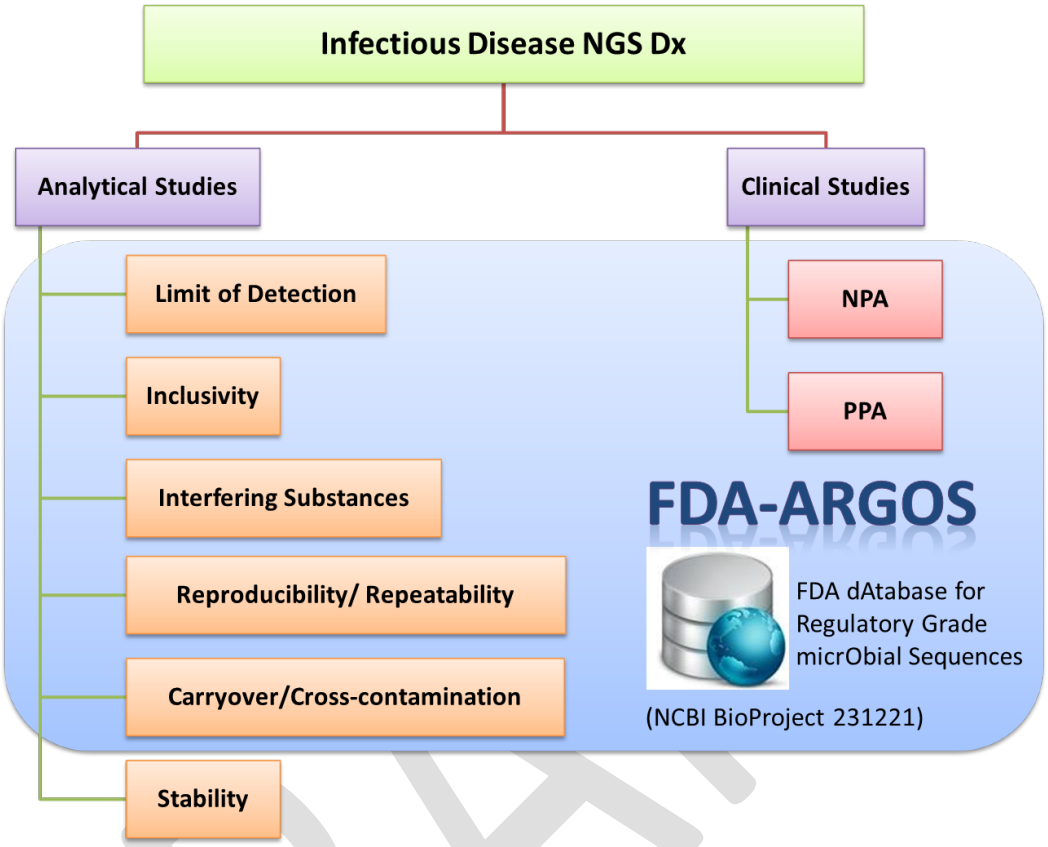

Figure 2: Overview of analytical and clinical studies that support FDA clearance or approval of Infectious Disease NGS Dx.

In general, for targeted Infectious Disease NGS, validation concepts similar to those of other multiplexed devices could be applied. More information on the multiplex device validation concepts is available in the FDA guidance entitled "Highly Multiplexed Microbiological/Medical Countermeasure In Vitro Nucleic Acid Based Diagnostic Devices" (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM327294.pdf). For agnostic Infectious Disease NGS, a representative number of targets (certain agreed on organisms or markers) based on intended use or chosen panel should be validated in the analytical and clinical studies.

C. Analytical Performance

Figure 2 demonstrates the analytical performance studies you should conduct for your assay and include in your submission, which are as follows:

(1) Limit of Detection

Limit of Detection (LoD) provides a measure of assay analytical sensitivity for a particular target, and is defined as the lowest concentration of a target that can be sequenced reliably and distinguished from negative specimens with consistent detection in ≥95% of the specimen replicates. Proper determination of the LoD is critical as microbial pathogens may be present in a patient specimen at very low levels. Depending on the sequencing format, ranging from a targeted to agnostic sequencing approach, there may be different considerations for how the LoD is established and validated. If a targeted multiplexed sequencing panel approach is used, then validation concepts similar to those of other multiplexed devices could be applied. For more information refer to the FDA guidance entitled "Highly Multiplexed Microbiological/Medical Countermeasure In Vitro Nucleic Acid Based Diagnostic Devices" (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM327294.pdf).

In brief, the LoD determination can be performed with a pool of different claimed targets in claimed specimen matrix. This pooled approach is also applicable for inclusivity and reproducibility studies. The sensitivity and specificity determination can be determined with limited prospectively performed clinical studies.

In contrast, if an agnostic sequencing approach is applied, determination of LoD for every target sequence included in the intended use may not be feasible. The appropriateness of the LoD determination is dependent on the design of the Infectious Disease NGS Dx device and type of sequencing employed.

For example, one approach for agnostic Infectious Disease NGS Dx devices is to design a feasibility study to approximate the LoD range. Mock samples can be designed to simulate the Infectious Disease NGS Dx device samples as closely as possible under controlled conditions. Mock samples should contain human background DNA at an acceptable clinical level expressed in genome equivalents/mL. A set of representative pathogen and marker targets for the assay's intended use should be spiked in at clinical levels expressed in genome equivalents/mL. Consult literature to determine the appropriate clinical levels for each pathogen and marker target of interest. You can consult with FDA prior to commencing these studies. Assurance should be provided that the device can detect the clinical range of the targets from potentially single copy to highest documented levels. We recommend the use of "spike-ins" for internal quality control (e.g., National Institute of Standards and Technology (NIST) RNA spike-in control standard reference materials (SRM) 2347). An estimate of the LoD range can be determined by examining the sequencing and read mapping statistics as described in Subsection VI(B) – Infectious Disease NGS Dx Device Performance Metrics.

The initially established preliminary LoD determined by testing a small number of replicates at each concentration should be confirmed by testing a minimum of 20 independent

replicates at the lowest concentration that can produce a positive result greater than 95% of the time.

The use of Probit analysis may also be used to establish LoD, provided the study is appropriately designed. The CLSI document entitled "EP17-A2, Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline; 2012" provides additional information about the Probit approach.

(2) Inclusivity

Validation of inclusivity or analytical reactivity should be conducted based on the intended use of the device and the sequencing strategy. Depending on the diagnostic claims made by the manufacturer, the studies should be designed to validate the ability to specifically detect potential genetic variation among the pathogens and resistance and virulence markers included in the intended use. The approach to establish inclusivity should use intact cultured organisms that undergo all pre-analytical steps. In certain circumstances, such as rare organisms, non-culturable organisms or BLS3 and BSL4 organisms pre-extracted and defined nucleic acids could be used. The targets used in this evaluation should be tested at or very near the LoD of the device. Note that the LoD of the device depends on the target tested in the device and could be different for each target assessed. The evaluation could use test panels designed to reflect the different genetic elements on which any conclusions would be based.

The inclusivity and reactivity evaluation can be performed with panels of organisms. These panels should be designed to include different strains, laboratory isolates, serotypes, and other closely related subspecies relevant to the specimen type. It is important to note that the panel design for inclusivity should incorporate a diverse and clinically relevant specimen set. To ensure the highest quality materials are used in this analysis, the identity and titer of the original stock should be confirmed (e.g., genome equivalent). For example, if your assay detects and identifies *Salmonella enterica*, we recommend that you demonstrate that the test can detect all frequently reported serotypes by testing at or near the specific LoD or cut-off value.

When you cannot acquire sufficient organisms to present an adequate diversity, we recommend that you contact FDA to discuss your study. When strain availability is limited, laboratory testing can be augmented through *in silico* analysis of target sequences. *In silico* analyses should include sequences of clinically relevant organisms and represent temporal, geographical, and phylogenetic diversity for each claimed target. In these cases, the *in silico* approach will be used to guide the inclusion of pathogens for traditional analyses and empirical testing of these isolates should be noted in the intended use of the device. For example, an approach whereby an *in silico* analysis guides laboratory testing could be based on read mapping identity. With this approach, representative organisms selected from groups with decreasing levels of identity to the target region will be selected for further laboratory testing. We recommend that you provide a clear rationale for the inclusion of the selected

strains, the metrics used to assess inclusivity, and a clear presentation of the read mapping in the specific regions of interest for each pathogen and marker target evaluated. This panel-based approach applies to targeted sequencing approaches based on amplification strategy.

For agnostic sequencing, we recommend the use of panels designed to cover adequate diversity of the assay's proposed intended use. Phylogenetic trees should be considered for panel design. If the assay's intended use proposes to identify targets that are closely related to each other (i.e., one base variation), inclusivity testing should include representative panel members with these base variations.

(3) Interfering Substances

An evaluation of interfering substances and near neighbors found in the clinical specimen that could interfere with signal generation and sequencing should be considered. Potential sources of interfering substances from the clinical specimens include exogenous substances (i.e., prescription/non-prescription drugs, anticoagulants, etc.) and endogenous substances (i.e., proteins, lipids, hemoglobin, bilirubin, etc.). The CLSI document entitled "EP07-A2, Interference Testing in Clinical Chemistry; Approved Guideline; 2005" provides additional information about how to design interference studies. The selection of inhibitors in the device validation studies would be determined by the indicated clinical specimen type. Additionally, a thorough evaluation of potentially interfering substances that could be introduced by the sequencing instrument should be considered during the validation process and could include residual chemicals from previous treatments or wash cycles.

For targeted sequencing, you should submit:

- Interference by contaminants when targets are present.
- Interference by other microorganisms when targets are present (microorganisms known to be present in types of specimen tested by assay for specific indication of use (clinical syndrome)).
- Interference by human background, if applicable.
- Cross-reactivity when targets are not present (near neighbors).
- Interference by polymerase chain reaction (PCR) inhibitors.
- Competition of amplifying primers, if applicable.

For agnostic sequencing, you should submit:

- Interference by human background.
- Interference by contaminants when targets are present.
- Interference by other microorganisms when targets are present (microorganisms known to be present in types of specimen tested by assay for specific indication of use (clinical syndrome)).
- Cross-reactivity when targets are not present (nonpathogenic microbiota, near neighbors).

- Interference by PCR inhibitors.

Please note that targeted and agnostic sequencing should include a detailed description of contaminants based on the intended use of the Infectious Disease Dx (e.g., skin biota in a blood sample).

(4) Precision (Reproducibility and Repeatability)

The reproducibility of the Infectious Disease NGS Dx device should be evaluated to assess the variability when the same material is repeatedly tested and multiple variables are introduced. For example, evaluation of reproducibility could be done using instruments at multiple sites with different operators running the instruments on different days. The evaluation should also determine the effect of multiple reagent lots on the variability of the performance of the device and any impact it may have on the final results. The CLSI document entitled "EP12-A2, User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline; 2008" provides additional information about how to design reproducibility studies. Microbial standard reference materials (SRM's) are under development by NIST and will be a valuable tool for use in this evaluation.

Similarly, repeatability should be evaluated to assess the precision of the assay when a standard material is analyzed multiple times at fixed conditions. This evaluation should be performed at a single site, evaluating as many non-assay related variables to determine the impact, if any, that the device has on the precision of the sequence outputs. The CLSI document entitled "EP12-A2, User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline; 2008," provides additional information about how to design repeatability studies. Similar to the evaluation of reproducibility, the evaluation of repeatability could also employ the SRM's that are currently under development by NIST.

(5) Carryover and Cross-contamination

Evaluation of the effects from carryover contamination should be considered. This should include evaluation of the entire device, including sample preparation and library preparation, where known positive samples (at a high target concentration) and negative samples are alternated. The carryover rate from previous runs should be calculated and reported. This information should be included in the device labeling to caution the end user. Furthermore, depending on the rate of carryover, there may need to be additional information included in the package labeling, such as warnings, precautions and cleaning instructions, to direct the end user on how to reduce or eliminate this effect.

(6) Stability

You should describe your study design for determining the real-time stability of the reagents and instrument, and if applicable, a description of stress test conditions and results. For each study, you should describe your acceptance criteria values and how you selected them.

(7) Additional Analytical Studies

We note that depending on device intended use, specimen type and study design, the following additional studies might be needed:

- Matrix equivalency study.
- Fresh vs. frozen study.
- Specimen stability study.
- Mixed infection study evaluating specimens with multiple targets.

D. Instrumentation and Software

The following referenced regulations are related to Infectious Disease NGS Dx devices and contain information applicable to these devices. These regulations are:

- 21 CFR 862.2265 – High-throughput DNA sequence analyzer for clinical use. The decision summary for the MiSeqDx Platform device intended for targeted sequencing of human genomic DNA from peripheral whole blood samples is available under submission number K123989 (http://www.accessdata.fda.gov/cdrh_docs/reviews/K123989.pdf).

- 21 CFR 862.2570 – Instrumentation for clinical multiplex test systems. Information concerning such instrumentation is available in FDA's guidance entitled "Class II Special Controls Guidance Document: Instrumentation for Clinical Multiplex Test Systems" (http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm077819.htm).

If the system includes software, information on the computational pipeline (e.g., programs, versions, etc.) should be submitted, from raw sequence data to final clinical call. Furthermore, software information detailed in accordance with the level of concern should be submitted. More information can be found in the document entitled "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices" (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm). The level of concern should be determined prior to the mitigation of hazards. *In vitro* diagnostic devices of this type are typically considered a moderate level of concern because software flaws could indirectly affect the patient and potentially result in injury due to inaccurate information.

For any device that uses a proprietary database to define the outcome of a signal generated by their device, FDA recommends that the quality criteria for establishing the accuracy of

regulatory-grade reference databases as well as the methods for curating, maintaining, and updating the databases be included in the submission. The regulatory-grade genomic target sequence entry for each claimed organism in the database should be constructed using a minimum of five well characterized isolates. You should provide the procedures and acceptance criteria of how correct species designations for each entry are evaluated. Please provide details of organism or marker identification and how sequence quality is assessed.

In your submission, you should provide a detailed table representing the composition of your database to include the number of isolates per claimed organism, summary data of how each isolate in the database was characterized (e.g., sequencing, biochemicals, certificate of analysis) and all applicable regulatory-grade quality metrics outlined in Section VIII Appendix for Comparator Database Quality Criteria for Regulatory-Grade Genomic Entries, for any device using the public database resource "F-ARGOS – FDA dAtabase for Regulatory Grade micrObial Sequences (BioProject 231221)" (http://www.ncbi.nlm.nih.gov/bioproject/231221(update with FDA web portal link)).

If the database contains more organisms and markers than the sponsor is seeking, you should specify whether your matching algorithm searches for matches against all organisms and markers in the regulatory-grade genomic target sequence reference database or only against the claimed database organisms. If the number of regulatory-grade genomic target sequences for a particular claimed organism or marker is limited, then more unique isolates should be tested in the clinical trial to verify the target reference sequences. Further, the sponsor should evaluate the matching algorithm on how an isolate identification is determined (e.g., matching against the regulatory-grade validated organism and marker database only or against the entire regulatory-grade).

If biothreat organisms are included in the database, please contact Heike Sichtig Ph.D., Division of Microbiology Devices at 301-796-4574 or by email at Heike.Sichtig@fda.hhs.gov prior to initiating studies.

Below is a list of additional references to help you develop and maintain your device under good software life cycle practices consistent with FDA regulations. These references are as follows:

- "General Principles of Software Validation; Final Guidance for Industry and FDA Staff" (http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm085281.htm)
- "Off-the-Shelf Software Use in Medical Devices" (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073779.pdf)
- "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices"

(http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm089543.htm)

- 21 CFR 820.30, Subpart C – Design Controls
- ISO 14971-1; Medical devices - Risk management - Part 1: Application of risk analysis
- AAMI SW68:2001; Medical device software - Software life cycle processes

E. Clinical Evaluation

Determination of the clinical sensitivity (or positive percent agreement) and specificity (or negative percent agreement) of an Infectious Disease NGS Dx device can be done using many of the same principles applied to other microbial diagnostic devices. The evaluation should be done at multiple geographically and demographically diverse study sites in the intended use environment using specimens indicated for the subject device, and with operators trained at the appropriate level. The intended use population should be defined appropriately using recognized clinical definitions (e.g., IDSA, EORTC). Please note that only one site may be located outside of the United States. However, given the number of potential pathogens and resistance and virulence markers that NGS technologies may be able to detect in a single clinical specimen, the application of more traditional regulatory strategies may hinder approval or clearance of these devices by requiring extensive evaluation of every detected organism (genomic sequence) from a single specimen, or in the case of device specificity all of those that were not detected, using expensive reference methods.

Therefore, to promote a least burdensome regulatory approach, we are proposing an alternative validation process that will rely heavily on public databases that are populated with high-quality genomic sequences that meet certain regulatory quality criteria (see Appendix VIII). The genomic sequence outputs from the subject device, when compared against the high quality database with sufficient coverage, should provide adequate information to determine the specificity of the device. Clearly, there may not be adequate representation of every organism in the public domain to employ this approach in its entirety at the present time; however, there may be pathways where certain facets of this strategy can be employed until such a time as there is adequate coverage in the public domain, especially if a panel-based approach is utilized.

In addition to regulatory-grade reference sequences, the implementation of Infectious Disease NGS Dx devices relies heavily on a robust analytical validation of the LoD (in the appropriate matrix). Moreover, you should provide information relating the analytical sensitivity of the device to the clinically relevant range of the pathogen load in the indicated disease state.

(1) Evaluation of Negative Percent Agreement

Evaluation of negative percent agreement should be conducted using prospectively collected specimens and analysis, at a minimum, of three clinical testing sites, two of which should be in the U.S. Patient enrollment in the study should be based on signs and symptoms and meet any additional inclusion criteria for the study. In general, the use of healthy blood donors is not acceptable; however, in some circumstances (e.g., contrived specimen), these specimens may be appropriate for these studies, and we encourage developers to contact FDA to discuss when these types of specimens are appropriate.

Generally, for the evaluation of the negative percent agreement, 1500 prospective samples should be collected and analyzed by the subject device in order to obtain sufficient statistical power for FDA to make a substantial equivalence determination or to determine the reasonable assurance of safety and effectiveness of the device. Depending on the number of organisms and specimen types to be used with an Infectious Disease NGS Dx device, negative percent agreement evaluation can be done using the regulatory-grade genomic target sequence database as a comparator. If possible, negative percent agreement (NPA) should be evaluated with patients from the intended use population.

If the tested organism or marker is not available in the database for evaluation with other acceptable comparator methods (CMs), we recommend consulting the FDA guidance entitled "Highly Multiplexed Microbiological/Medical Countermeasure In Vitro Nucleic Acid Based Diagnostic Devices - Guidance for Industry and Food and Drug Administration Staff" (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm327293.htm) for a description of acceptable CMs that are applicable to targeted sequencing approaches and to discuss with FDA before conducting the study. If specimen volume becomes prohibitive to run all comparator tests, a clearly defined randomized approach should be taken such that a minimum of 100 of each CM for each detected organism or marker would be analyzed. In addition, provisions should be made so that an adequate number of specimens can be analyzed for biothreat organisms in order to meet the specificity performance criteria.

Using the example of a targeted microbial device detecting 20 different organisms by employing 100 Amplicons (5 Amplicons per organism) where each CM needs equal test volumes and allows five CM tests, the first specimen could be tested with comparative methods (CM1, CM2, CM3, CM4, CM5), the second specimen tested with CMs as (CM6, CM7, CM8, CM9, CM10) and so on. After testing four specimens, each CM would have been applied one time. After testing the first four specimens, a new array of integer numbers from 1 to 20 in a random order could be generated and the next four specimens could be tested with comparative methods according to this new array. Developers should power the study to establish clinical specificity with a point estimate and lower bound of the 95% CI to exceed a level that has been agreed upon through FDA feedback. For biothreat organisms[12],

[12] For clarification of pathogens considered to be biothreats please see the National Select Agent Registry (http://www.selectagents.gov/SelectAgentsandToxinsList.html).

clinical specificity would be demonstrated to achieve a point estimate 99.9% with a lower bound of the 95% CI greater than 99%.

Also note, for each specimen from the prospective study that has positive results by the subject targeted Infectious Disease NGS Dx device for a pathogen, this specimen will also require testing by the corresponding CM. Information about CM results that were driven by a positive result from the Infectious Disease NGS Dx device should not be used directly in the calculation of sensitivity and specificity as it introduces bias into estimation of the Infectious Disease NGS Dx device performance. However, this information is useful to understand the overall performance of the targeted Infectious Disease NGS Dx device, especially in terms of co-detections, and should be presented in a separate table. Comparative performance of the specificity of the Infectious Disease NGS Dx device should be established using FDA cleared or approved devices, if available. Use of cleared Infectious Disease NGS Dx devices, when appropriate, is recommended. When FDA cleared or approved devices are not available or appropriate, a composite reference method of two well-validated PCR based assays followed by bidirectional sequencing could be used.

(2) Evaluation of Positive Percent Agreement

The analysis of positive percent agreement will include a minimum of 50 positive specimens per claimed organism or marker. Initially, a culture-based or PCR-based reference method (preferably FDA cleared or approved) should be used for claimed targets which should at least contain representative targets for an assay's intended use. A regulatory grade confirmatory database can be used to potentially confirm closely related targets. For agnostic sequencing approaches, panels of representative organisms could be designed (confirmed positives by an acceptable CM) and the menu of all claimed organisms should need to be tested using the confirmatory microbial reference database. The number of positive specimens for each pathogen or marker to include will be driven by the point estimate of positive percent agreement and the lower bound of the 95% two-sided confidence interval. These values can vary depending on the intended use of the device. You should discuss with FDA to determine the appropriate clinical sensitivity levels for each pathogen or marker indicated by the Infectious Disease NGS Dx device.

For example, an Infectious Disease NGS Dx device with a panel menu composed of bacteremic organisms should include a sufficient number of archived and retrospective specimens for each claimed pathogen or marker to generate a result with at least 90% positive percent agreement with a lower bound of the two-sided 95% confidence interval (CI) greater than 80%. Assuming a point estimate of 90.2% is achieved; a minimum of 61 positive specimens (55/61) will need to be included to surpass the indicated lower bound of the 95% CI of greater than 80%. Indeed, for 61 specimens, 55 out of 61 yields a point estimate of 90.2% with 95% CI: 80.2% to 95.4%. However, using the example of 60 specimens, with a performance of 54 out of 60 yielding a point estimate of 90.0%, the CI does not meet the minimum performance bar with 95% CI being 79.9% to 95.3%.

All positive archived or retrospective specimens (as determined by the reference method prior to banking the specimens) will be analyzed with the corresponding CM and the subject device. Verification by the CM is imperative to ensure that specimens were properly archived, that no specimen degradation occurred during storage, and that the specimens are properly identified. Any specimens that are not confirmed as positive by the CM should not be included in the initial performance evaluation for claimed organisms. However, results from the subject device testing the unconfirmed specimens can be used when expanding the claimed organism panel using the confirmatory microbial reference database. Additionally, any positive determination for any other pathogen by the Infectious Disease NGS Dx device should also be verified by the CM as this provides additional information about NGS device performance, especially in cases of potential co-infection. Alternative approaches to confirming the positivity of specimens can be considered; however, we encourage discussions with FDA before executing the study. Retrospective positive specimens should be the same specimen type as listed in the intended use of the device and should have been collected from the appropriate intended use population. The specimens selected for inclusion in this study should represent the clinically relevant range of concentrations for the particular pathogen or marker. In cases where extracted nucleic acids from positive clinical specimens have been archived, they can be considered for inclusion in the analysis provided that the appropriate intended use population was used, the indicated specimen type was collected and processed using the indicated pre-analytical steps, and confirmation was done by the corresponding CM.

We recognize that actual clinical human specimens, archived or otherwise, may not be readily available for biothreat organisms. The use of mock clinical specimens, prepared by spiking cultured pathogen into individual negative clinical specimens may be used. For this analysis, 50% of the spiked specimens would be made at the LoD concentration, while the remaining 50% would span the expected clinical range of pathogen concentrations. For non-biothreat pathogens with extremely low prevalence, mock specimens should reflect the relevant clinical range. Justification of the expected clinical range through peer-reviewed literature references or feedback from subject matter experts should be provided by the developers for each specimen type indicated. Given the restrictions associated with the handling of many of the biothreat organisms, arrangements to validate the clinical performance at qualified institutions with the capability to conduct the proper studies should be made. For biothreat organisms, due to the logistical issues with this aspect of the positive percent agreement validation, analysis can be conducted at a single site. Alternatively, if the biothreat pathogen does not involve a special facility, a multiple testing site approach can be used to evaluate positive percent agreement, and the archived specimens (positive and negative) should be randomly and evenly distributed among three testing sites for analysis.

Prior to conducting any studies using mock specimens, you should consult FDA for feedback. Your protocols should include a detailed test plan and justification.

(3) Data Presentation

You should present positive percent agreement (PPA) and NPA separately for each pathogen or marker identified by the Infectious Disease NGS Dx device. Each agreement should have a 95% CI. Also, you should present:

- The results of your device for the specimens that have co-infections as obtained by the reference method. Note: This information may not be available for some of the prospective specimens due to prohibitive specimen volume;
- The results of the CM for the specimens that have co-infections as obtained by the Infectious Disease NGS Dx device; and
- The results for CM measurements that were driven by a positive result from the subject Infectious Disease NGS dx device.

All specimens in the clinical study should be tested with the Infectious Disease NGS Dx device as described in the instructions for use of your device. For example, if specimens with initial indeterminate or invalid results are re-tested according to the instruction for use for the Infectious Disease NGS Dx device then the final result obtained from the indicated testing procedure for these specimens should be used in your statistical analysis. For the specimens in your clinical study, you should provide the following: 1) the percent of re-tested specimens because of initial indeterminate results (if applicable), and 2) the percent of re-tested specimens because of initial invalid results (if applicable). In addition, you should present the percent of final invalid and final indeterminate results (if applicable) for each. You should provide numerical result distributions of the Infectious Disease NGS Dx device for all prospectively collected fresh, prospectively collected archived, and banked pre-selected specimens shown separately, for each pathogen and for all pathogens combined.

(4) Study Specimens and Specimen Types

You should use clinical specimens from all specimen types and matrices you claim in your intended use to demonstrate that correct results can be obtained from clinical material. For specimens you use in your clinical studies, you should provide data demonstrating that storage and transport of any banked specimens have not affected assay results as well as the methods used to bank the specimen as positive for a specific organism. For example, if archived specimens are previously frozen, you should perform an analytical study to demonstrate that your assay provided equivalent results for fresh and frozen specimens. If you have questions regarding the choice of appropriate specimen type(s) as well as specimen types that can be pooled, please contact FDA.

VII. Device Modification

The following information defines a pathway to incorporate new targets on an existing platform device in response to public health needs or an emergency situation and to ensure that performance characteristics of a cleared or approved device are consistent over time. Addition of a new sequence target to an Infectious Disease NGS Dx device may result in a

major change or modification in the intended use of the device, or a new intended use, and thus may require the submission of a traditional regulatory submission (e.g., a new premarket notification (510(k)) or premarket approval application (PMA) submission) in accordance with 21 CFR 807, Subpart E along with validation data to establish performance. Since many studies have been conducted to establish the performance of the previously cleared or approved device and we presume that the assay's performance has not changed, only a subset of the evaluations may need to be repeated for the new submission. In your submission, you should also provide a detailed procedure for adding new species to your device. These procedures include: acceptance criteria, risk analysis and validation testing. We note that the addition of target sequences to your cleared or approved organism claims can be performed using the data generated from the original clinical study against the newly expanded database. In your submission, include how database updates will be issued. We encourage device developers to contact the Agency for assistance.

In cases where the inclusion of additional targets to address a public health need or emergency is necessary, the studies to substantiate performance will focus primarily on the additional sequence target. Certain types of evaluations may not be needed when adding a new sequence target or modifying a device, including stability studies and the evaluation of carry-over and cross contamination. Additionally, the scope of the reproducibility study and clinical evaluation should be focused on the new or modified sequence target and a representative panel for performance confirmation.

Furthermore, modification or an update of a library and bioinformatics pipeline should be communicated to FDA prior to use and implementation.

VIII. Appendix: Comparator Database Quality Criteria for Regulatory-Grade Genomic Sequence Entries

Infectious Disease NGS Dx devices have the potential to detect multiple infectious agents and resistance and virulence markers in a single human clinical specimen. To promote a least burdensome regulatory approach for devices that incorporate Infectious Disease NGS Dx technology, FDA proposes the use of an optional alternative comparator method for clinical evaluation that relies heavily on public databases populated with regulatory-grade target sequences. For this application, FDA, in collaboration with various federal agencies, has developed a resource entitled "FDA-ARGOS – FDA dAtabase for Regulatory Grade micrObial Sequences (BioProject 231221)" containing a set of validated regulatory-grade genomic sequence entries (http://www.ncbi.nlm.nih.gov/bioproject/231221(update with FDA web portal link)). This Appendix summarizes FDA's framework of a public regulatory-grade microbial reference database.

FDA proposes the use of regulatory-grade genomic sequences as an alternative comparator for clinical evaluation. In order to use the alternative comparator method, microorganisms as well as resistance and virulence markers claimed in the intended use or panel (e.g., a Filovirus panel) should be available as regulatory-grade references before clinical evaluation. We continue to expand the database by adding new entries or by qualifying existing entries. Please contact the Agency if you have specific requirements for representation and limited resources to develop these regulatory-grade genomic sequence entries.

The following sections highlight the areas of information that FDA intends to capture so that genomic sequence depositions in the public domain or in proprietary databases can be evaluated and qualified for regulatory purposes. To qualify as a regulatory-grade genomic sequence entry, the microbial organism or resistance and virulence marker has to be explicitly identified prior to sequencing. The regulatory part of the database should include validated regulatory-grade genomic sequence entries for all organisms claimed in the Infectious Disease NGS Dx device's intended use. As previously stated, depending on the intended use (e.g., genus, species or marker level ID) or panel (e.g., a Filovirus panel), only the regulatory-grade identified microbial agent(s) and marker(s) can be included in the final report.

Quality metrics for regulatory-grade genomic sequence entries:

A. Extracted Genomic DNA (gDNA)

Extracted gDNA should be of high quality and purity, and at sufficient concentration to achieve a suitable yield to assure adequate depth and breadth of genomic coverage for the type of sequencing method employed.

B. BioSample Metadata

A minimal description of the sample source material (e.g., clinical, environmental, public health need) is necessary for traceability. We are using the following descriptors as outlined below. (Note: Minimal metadata is modeled in part after NCBI's minimal pathogen template.)

(1) Clinical Sample

Clinical Sample	Description
1. Unique ID	Unique Database ID for the sample
2. Organism	Organism genus and species
3. Identification Method	Sample identification method(s) (biochemical, MicroScan, Vitek)
4. Isolation Source	Anatomical sampling site (e.g., skin, wound, urine catheter)/ Specimen type (e.g., blood, stool, urine)
5. Host Disease	Relevant clinical syndrome (e.g., sepsis, meningitis, bacteremia)
6. Collection Date	Date of sampling (month and year)
7. Collected By	Place or Lab of origin for clinical sample collection
8. Geographic Location	Geographical origin of the sample
9. Age Category	Age group (in years, FDA categories)
10. Gender	Gender (male, female) (recommended)
11. AST Method*	Antimicrobial susceptibility testing method (recommended)
12. AST Method Manu.*	Manufacturer of AST Method (recommended)
13. Antimicrobial Susc.*	For each antibiotic (e.g., Vancomycin, Oxacillin) (recommended)

*It is important to note that not every entry will have the associated antimicrobial susceptibility testing (AST) data; however, the lack of the AST data will not be used as a criteria for exclusion. The purpose of this information is to create a link between the phenotypic traits of particular organisms and their genomic sequence. Moreover, this information is becoming increasingly critical as diagnostic technologies begin to migrate away from more traditional culture based formats.

(2) Environmental Sample (for clinical next neighbor evaluation and rule out)

Environmental Sample	Description
1. Unique ID	Unique Database ID for the sample
2. Organism	Organism genus and species
3. Identification Method	Sample identification method(s) (biochemical, MicroScan, Vitek)

4. Isolation Source	Sampling site (e.g., zoonotic, air filter, river bed)/ Specimen type (e.g., tick pool, water, soil)
5. Collection Date	Date of sampling (month and year)
6. Collected By	Place/Institute of origin for environmental sample collection
7. Geographic Location	Geographical origin of the sample
8. Host/Reservoir	Specimen host (e.g., mosquito, cow) (recommended)
9. AST Method*	Antimicrobial susceptibility testing method (recommended)
10. AST Method Manu.*	Manufacturer of AST Method (recommended)
11. Antimicrobial Susc.*	For each antibiotic (e.g., Vancomycin, Oxacillin) (recommended)

*It is important to note that not every entry will have the associated antimicrobial susceptibility testing (AST) data; however, the lack of the AST data will not be used as a criteria for exclusion. The purpose of this information is to create a link between the phenotypic traits of particular organisms and their genomic sequence. Moreover this information is becoming increasingly critical as diagnostic technologies begin to migrate away from more traditional culture based formats.

(3) Clinical Public Health Need Sample

Clinical Public Health Need Sample	Description
1. Unique ID	Unique Database ID for the sample
2. Organism	Organism genus and species
3. Identification Method	Sample identification method(s) (biochemical, MicroScan, Vitek)

Also include all available and applicable clinical or environmental descriptors. We note that entries are included only on a case by case basis due to existing restrictions on obtaining this data; however, inclusion should be validated by reason of public health need (e.g., samples from an outbreak that have clinical relevance but have very limited or no descriptive metadata).

C. Sequencing Data

The minimum requirement for sequencing data is that the generated raw reads should be deposited in NCBI's Sequence Read Archive (SRA) and assemblies should be deposited at NCBI's Assembly division. The availability of raw reads and assemblies will provide a pathway to re-analyze the data as newer technologies emerge. Furthermore, annotation data should be deposited when available.

Sequencing Data	Description
1. SRA	Deposit raw reads at NCBI's Sequence Read Archive (SRA) division
2. Assembly	Deposit assemblies at NCBI's Assembly division
3. Annotation*	Deposit annotations at NCBI's Annotation division (recommended)

* Genome annotations should be deposited at NCBI's Annotation division when available and should be requested to be added using NCBI Prokaryotic Genome Annotation Pipeline (PGAP).

D. Sequencing Metadata

A minimal description of the sequencing process is necessary for traceability. We are using seven descriptors as outlined below including bioinformatics tool information for assembly and annotation, and genomic coverage information.

Sequencing Metadata	Description
1. Library	Library manufacturer, strategy, source, selection and layout of library
2. Platform	Platform manufacturer and instrument model
3. Submitted by	Name of person or sequencing center that submitted the clinical or countermeasure isolate sequencing data
4. Fold coverage	Coverage of genome
5. Pipeline	Processing pipeline used to generate data, sequencer platform software and version
6. Assembler	Assembler and version
7. Annotation Tool	Annotation tool and version (recommended when available)

E. Suggested Phenotypic Metadata

A description of the phenotypic information is suggested to create a link between the phenotypic traits of particular organisms and their genomic sequence. We are recommending five descriptors as outlined below (descriptors 1-4 are also included in sections VIII(B) and (C) of this Appendix).

Suggested Phenotypic Metadata	Description
1. Annotation	Genome Annotation data
2. AST Method	Antimicrobial susceptibility testing method
3. AST Method Manuf.	Manufacturer of AST Method
4. Antimicrobial Susc.	For each antibiotic (e.g., Vancomycin, Oxacillin, Tetracycline, Tobramycin)

1726 1727	5. Addl. Phenotypic Data	Info on morphology, gram stain, virulence data, metabolic data