Establishing the Performance Characteristics of In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Human Papillomaviruses

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.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: August 14, 2015

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draft document within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance.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 regarding this document contact Natalia Comella, Ph.D., at (301) 796-6226 or by email at natalia.comella@fda.hhs.gov or Marina V. Kondratovich, Ph.D., at (301) 796-6036 or by email at marina.kondratovich@fda.hhs.gov.

When final, this guidance will supersede FDA's guidance entitled "Establishing the Performance Characteristics of In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Human Papillomaviruses," dated November 28, 2011.

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

DRAFT

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 1740 to identify the guidance you are requesting.

DRAFT

Table of Contents

DRAFT

DRAFT

Establishing the Performance Characteristics of In Vitro Diagnostic Devices for the Detection or Detection and Differentiation of Human Papillomaviruses

Draft Guidance for Industry and Food and Drug Administration Staff

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic.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 facilitate study designs to establish the performance characteristics of in vitro diagnostic devices (IVDs) intended for the detection, or detection and differentiation, of human papillomaviruses (HPVs). These devices are used in conjunction with cervical cytology to aid in screening for cervical cancer or as first-line primary cervical cancer screening devices. These devices include those that detect a group of HPV genotypes, particularly high risk HPVs, as well as devices that detect more than one genotype of HPV and further differentiate among them to indicate which genotype of HPV is present. More than 100 HPV genotypes have been identified, approximately 40 of which can infect the genital tract [Ref. 1]. Infection with 'high-risk' types of HPV is considered a necessary cause of virtually all cervical cancer [Ref. 2]. Approximately fourteen HPV genotypes are considered carcinogenic or "high risk" [Ref. 3 & Ref. 20]. For the remainder of this document, "HPV" refers to a "high risk" HPV, except where otherwise noted. A "high risk HPV test" refers to an HPV IVD device that detects, but does not differentiate between different types of HPV; while a "HPV genotyping test" refers to an HPV IVD device that detects and further differentiates HPV types (some HPV tests provide individual HPV genotyping results in addition to the results of pooled probes).

DRAFT

This draft guidance, when finalized, provides detailed information on the types of studies the Food and Drug Administration (FDA) recommends to support a premarket application (PMA) for these devices. It is recommended that you contact FDA prior to beginning your studies to discuss specific study proposals and performance goals for your device.

This draft guidance is limited to studies intended to establish the performance characteristics of in vitro diagnostic HPV devices that are used in conjunction with cervical cytology for cervical cancer screening or as first-line primary cervical cancer screening devices. This draft guidance specifically addresses devices that qualitatively detect HPV nucleic acid from cervical specimens, but many of the recommendations will also be applicable to devices that detect HPV proteins. See Section III "Scope" for more details on what is covered by this draft guidance document.

FDA's guidance documents, including this 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

This document, when finalized, provides guidance for establishing the performance characteristics of in vitro diagnostic devices for the detection, or detection and differentiation, of human papillomaviruses in cervical specimens. These recommendations apply to PMAs for HPV IVDs.

A manufacturer who intends to market an IVD device for detection, or detection and differentiation, of human papillomaviruses must conform to the requirements of the Federal Food, Drug, and Cosmetic Act (the FD&C act) and obtain premarket approval prior to marketing the device (sections 513 and 515 of the FD&C Act; 21 U.S.C. 360c and 360e). Because HPV diagnostic devices are postamendment devices, they are automatically classified as class III under section 513(f)(1) of the FD&C act. Devices that have been classified by section 513(f)(1) into class III require premarket approval in accordance with section 515 of the FD&C act. See section 515(a)(2) of the FD&C act (requiring premarket approval for devices classified into class III by section 513(f)); see also section 513(a)(1)(C) of the FD&C act (defining a class III device as one that "is to be subject, in accordance with section 515, to premarket approval to provide reasonable assurance of its safety and effectiveness").

Further information on device testing can be found in FDA's guidance entitled "In Vitro Diagnostic (IVD) Device Studies – Frequently Asked Questions" at (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071230.pdf), and FDA's guidance entitled "Guidance on Informed Consent for *In Vitro* Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable" at

DRAFT

(http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071265.pdf).

III. Scope

This document recommends studies for establishing the performance characteristics of in vitro diagnostic devices for the qualitative detection, or detection and differentiation, of HPV. This guidance is limited to studies intended to establish the performance characteristics of in vitro diagnostic HPV devices that are used in conjunction with cervical cytology for cervical cancer screening, or as first-line primary cervical cancer screening devices. It does not address HPV testing from non-cervical specimens, such as pharyngeal, vaginal, penile or anal specimens, or testing for susceptibility to HPV infection. It does not address quantitative or semi-quantitative assays for HPV.

As postamendment devices, HPV diagnostic devices are automatically classified as class III devices under section 513(f)(1) of the FD&C act. To date, two product codes have been established for HPV nucleic acid detection devices: MAQ (HPV DNA detection devices) and OYB (HPV RNA detection devices). Both of these product codes are class III. The recommendations in this guidance apply to HPV diagnostic devices that detect HPV nucleic acid (HPV DNA and RNA). Many of the recommendations will also apply to HPV detection devices that utilize targets other than HPV nucleic acid (such as HPV protein). This guidance therefore may encompass future HPV product codes beyond the ones listed. This guidance does not apply to HPV-associated biomarkers (e.g., p16).

IV. Risks to Health

Failure of devices for the detection, or detection and differentiation, of human papillomaviruses to perform as expected, or failure to correctly interpret results may lead to incorrect patient management decisions in cervical cancer screening and treatment. False negative results may lead to delays in the timely diagnosis of cervical cancer and treatment, allowing an undetected condition to worsen and potentially increasing morbidity and mortality. False positive results could lead many women to unnecessarily undergo more frequent screening and potentially invasive procedures such as colposcopy and biopsy. False positive results for the highest risk types of HPV, such as HPV 16 and/or 18, could lead to unnecessarily aggressive treatment of cervical lesions that may impair fertility. Because cervical cancer screening is recommended for virtually all sexually active women and a substantial number of these women will be tested for HPV, the risk scale for potential harm to public health from false negative and false positive HPV results is significant. Therefore, establishing the performance of these devices and understanding the risks that might be associated with the use of these devices is critical to their safe and effective use.

The studies that are submitted in a PMA to establish the performance of HPV detection devices are a key factor for determining the safety and effectiveness of these devices.

DRAFT

V. Device Description

You should provide in your PMA a device description that includes information sufficient to understand what the proposed device is and how it works, such as:

- A description of the device in text and with pictures, diagrams, and/or engineering drawings, as applicable.
- An explanation of the mechanism of action and principles of operation.
- Characteristics of the device output (i.e. whether or not genotypes can be differentiated, genotypes assessed simultaneously vs. individually in a well or channel, etc.).
- A detailed technical description of the device including instruments, reagents, components, software, and accessories.
- The proposed indications for use of the device (including sample type(s) and collection devices).

VI. Test Methodology

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

- Test platform.
- Information and rationale for selection of specific target sequences and the methods used to design detection elements.
- Specimen collection and handling methods.
- All pre-analytical methods and instrumentation for collection, stabilization, and concentration of specimens, as appropriate.
- Limiting factors of the assay (e.g., saturation level, maximum cycle number, etc.).
- Reagent components provided or recommended for use, and their function within the system (e.g., buffers, enzymes, fluorescent dyes, chemiluminescent reagents, oligonucleotides, other signaling/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 that you recommend or provide to users.
- Instrumentation inherent to using your 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 useable result). This would include adjustment for background and normalization, if applicable. Show how results are reported and interpreted.
- Illustrations, photographs, and a detailed description of non-standard equipment or methods, as appropriate.

DRAFT

VII. Establishing Performance Characteristics

A. Analytical Studies

You should provide in your PMA analytical studies that conform to the following recommendations.

(1) Limit of Detection

FDA (or "we") recommends that you determine the limit of detection (LoD) of your device using serial dilutions of HPV genomic DNA or RNA transcripts, as appropriate, in sample collection buffer. Genomic DNA or RNA transcripts, or both, can be cloned or synthesized material, since HPV cannot be cultured. We recommend that you determine the LoD for each HPV genotype and each specimen collection media tested by the device.

If your assay is indicated for testing with liquid-based cytology (LBC) specimens, and involves centrifugation of the cervical cytology sample and removal of the LBC collection media (supernatant) prior to processing for HPV testing, you should perform your LoD studies in whatever matrix or buffer the cells are re-suspended in after the centrifugation step. If you use LBC mock-samples containing HPV-infected cell lines in any of your analytical studies (as recommended under Section VII(A)(2) "Precision" below), then you should also perform LoD studies with these types of samples. A human HPV-negative cell line is recommended to serve as a surrogate for non-HPV infected cells in LBC samples contrived from HPV-infected cell lines (i.e., SiHa and HeLa cell lines). You should conduct a paired sample LoD study with at least one of these HPV-infected cell lines showing that you get the same LoD results in both a pooled negative clinical and simulated background matrix (i.e., an HPV negative cell line in LBC media). If these two samples demonstrate equivalence then a negative cell line can be used as the background in other analytical studies.

We recommend that you first define a cutoff for the numerical signal (i.e., the limit of blank (LoB)) such that a signal above the LoB in a patient sample indicates that the virus was detected. You should also estimate the level of virus that gives a 95% detection rate (the LoD). There are two different types of devices to consider when establishing the cut-off. One type covers devices for which a distribution of numeric signals are obtained when repeatedly testing samples known to have zero concentration (true absence of targeted analyte). For this device type, LoB is a threshold for numeric signal with a pre-defined type I error (typically 5%), such that samples with a numeric signal above the LoB are considered as "HPV detected". For the second type of device, the ultrasensitive devices, samples with zero analyte concentration almost always have "HPV not detected" results (type I error is close to zero).

We suggest that you refer to the Clinical and Laboratory Standards Institute (CLSI) document EP17-A2 [Ref. 4] for the basic concepts, design, and statistical analysis of your LoD studies. For the first type of device described above, you can use the approach described in CLSI EP17-A2, and by Linnet and Kondratovich [Ref. 5] to estimate the LoD using the standard deviation of

DRAFT

samples with very low concentrations. For the second type of device, the LoD can be estimated from Probit analysis using hit rates (percent of virus detected) of different dilutions [Ref. 6]. Hit rates for these dilutions should cover a large part of the range of detection (0% detection to 100% detection). The LoD should be confirmed by preparing at least 20 additional replicates at the LoD concentration and demonstrating that the virus was detected 95% of the time. The LoD study should include serial dilutions of each targeted HPV genotype, cell line, or specimen type. In both approaches to LoD estimation, the appropriate sources of variability should be included in the LoD study by testing 3-5 samples over 3-5 days with 2-3 lots of your device.

Please note that the clinical cutoff, which defines positive and negative results for the HPV test on clinical samples, can be higher than the LoB, which analytically defines whether the HPV virus is present or absent. The C_{95} concentration is the concentration of analyte just above the clinical cutoff such that results of repeated tests of this sample are positive approximately 95% of the time. When the LoB is used as a cutoff, then the concentration C_{95} is the same as the LoD. For an HPV assay in which the clinical cutoff is higher than the LoB, the concentration C_{95} may differ from the LoD concentration.

(2) Precision

a. Samples for within-laboratory precision/repeatability studies

For establishing the precision of HPV tests, you should create 10-20 precision panel members with defined analyte levels and HPV genotype(s). You should establish performance using specimens with analyte levels that challenge medical decision points, in addition to specimens with moderate analyte levels (as described below). Since HPV cannot be cultured, HPV-infected human cell lines (as well as human HPV-negative cell lines) can be used to create panel members that mimic clinical specimens and contain a targeted level of HPV analyte. Utilizing cell lines is important for LBC specimens since this helps to account for some of the variability that arises due to the sampling and processing of a heterogeneous suspension of cells. When an HPV genotype that you intend to claim to detect is not readily available as an infected cell line, you may also use contrived panel members derived from HPV DNA plasmids or RNA transcripts, as appropriate. In addition to these contrived samples with defined levels of HPV infected cells or HPV nucleic acid, you should include four or more real clinical samples with signal levels that challenge the assay clinical cutoff, plus at least one clinical sample negative for HPV in your precision study panels. Real clinical samples should be used since cell lines and plasmids alone cannot address all the variability present in clinical samples. Clinical samples can be pooled to create sufficient volume and to achieve desired levels of virus concentrations. In some instances, pooling of clinical samples may significantly increase the variability observed; in that event, you should contact FDA to discuss alternative study designs using individual clinical samples. Viral load cannot be defined for clinical specimens, but you should challenge medical decision points [i.e., clinical cutoff(s)] by including specimens that test positive and/or negative only a fraction of the time (the exact value of this fraction is not critical, anything from 5 to 95% positivity is acceptable). This way, the end user can see what output signal levels have a degree of variability associated with their qualitative results. Panel members derived from cell lines and/or real clinical samples should be processed as real LBC specimens,

DRAFT

starting from suspension in LBC media before the nucleic acid extraction step. If you wish to utilize a precision panel composed entirely of real clinical specimens (without any simulated specimens), please contact FDA to discuss recommendations for additional characterization of such specimens.

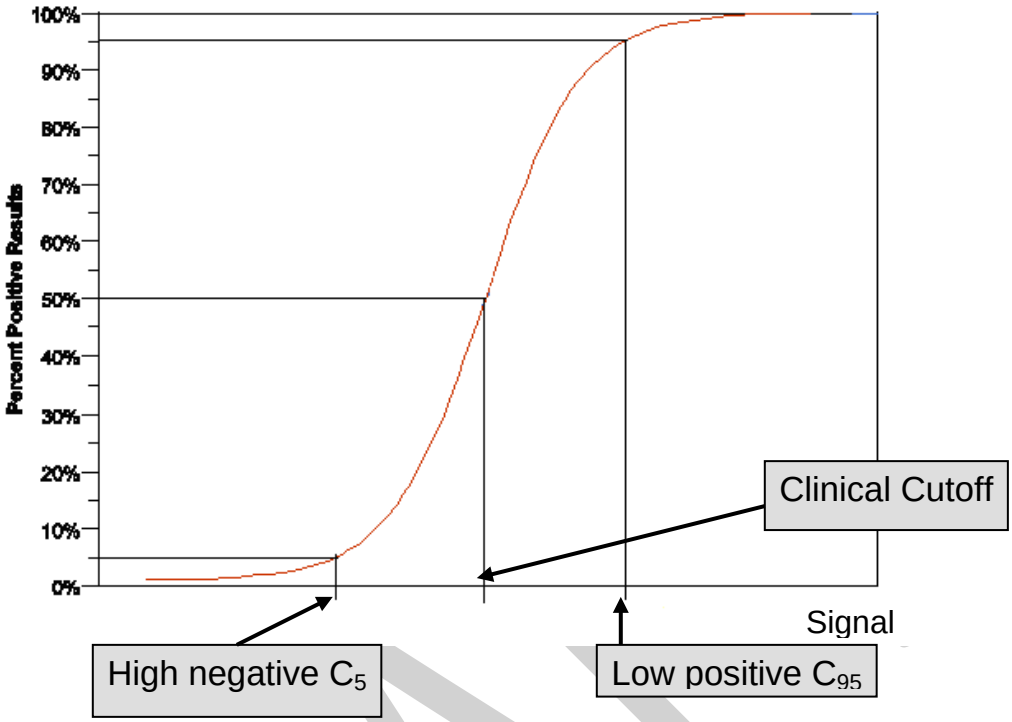
We recommend that you conduct within-laboratory precision studies for devices that include complex instruments or automated components. You should include sources of variability (such as operators, days, instruments, assay runs, etc.) encompassing a minimum of 12 days (not necessarily consecutive), with two runs per day, and two replicates of each sample per run. You should assess precision between three reagent lots; there should be no redundancies in the individual reagent lots evaluated within each kit lot, or in any calibrators or controls that are sold separately. Between-instrument imprecision can be evaluated during your in-house precision study, but is more commonly assessed during a sponsor's reproducibility study (as described below in Section VII(A)(2)(b) "Reproducibility").

For simulated precision panel members, the test panel should include at least six samples (two HPV genotypes) at three levels of viral load as described below (also see Chart 1):

- **A "zero concentration" sample** with no analyte present.
- **A "high negative" sample** aiming to represent the analyte concentration below the clinically established cut-off such that results of repeated tests of this sample are negative *approximately* 95% of the time and results are positive *approximately* 5% of the time, C_5 concentration (e.g., for real–time PCR assays, a sample with an analyte concentration not more than 10 fold below the clinical cutoff of the assay).
- **A "low positive" sample (C_{95} concentration)** with a concentration of analyte just above the clinical cut-off such that results of repeated tests of this sample are positive *approximately* 95% of the time.
- **A "moderate positive" sample** with a concentration at which one can anticipate positive results approximately 100% of the time (e.g., approximately two to three times the concentration of the clinical cut-off).

DRAFT

Chart 1. Three Levels for Precision Studies



When the LoB is used as a clinical cutoff, then the concentration C_{95} is the same as the LoD and the zero concentration (no analyte present in sample) is C_5 [Ref. 4]. CLSI documents EP05-A3 [Ref. 7] and EP12-A2 [Ref. 8] contain further information about designing and performing precision studies.

For precision studies, it is not necessary to have the high negative and low positive samples at exactly C_5 or C_{95}. If the high negative and low positive samples in the precision study are close enough to the cutoff that the standard deviation (or percent coefficient of variation (%CV)) is approximately constant over the range around the cutoff, the C_5 and C_{95} can be evaluated from this within-laboratory precision study.[1] The objective of estimating the C_5 and C_{95} concentrations in this manner is to ensure that your precision panel members are adequately challenging your medical decision points.

b. Reproducibility

[1] If the standard deviations (SD) in the precision studies for concentrations around the cutoff value (C_{50}) are almost constant, then: $C_{95} = C_{50} + 1.645 \times SD$, and $C_5 = C_{50} - 1.645 \times SD$. If the coefficient of variation (CV) in the precision studies for concentrations around the cutoff value are almost constant, then $C_{95} = C_{50} + 1.645 \times CV \times C_{95}$ and $C_5 = C_{50} - 1.645 \times CV \times C_5$. From here, $C_{95} = C_{50} / (1 - 1.645 \times CV)$ and $C_5 = C_{50} / (1 + 1.645 \times CV)$.

The protocol for the reproducibility study may vary slightly depending on the assay format. We recommend the following protocol:

- Evaluate the reproducibility of your test at three testing sites (e.g., two external sites and one internal site).
- Use a five day testing protocol, including a minimum of two runs per day (unless the assay design precludes multiple runs per day) and three replicates of each panel member per run.
- Each day have at least two operators at each facility perform the test.
- Each sample in the reproducibility study should have at least 90 measurements.
- Use the sample panel as described in the within-laboratory precision study above (include C_5 and C_{95} samples estimated from the within-laboratory precision internal study in your reproducibility study). For your reproducibility study, the cell line panel members and the clinical sample panel members should be processed for each run starting from the nucleic acid extraction step with an independent extraction for each run.
- Between-instrument imprecision is often assessed as part of a sponsor's reproducibility study (rather than during in-house precision testing) by having each site conduct testing using a different instrument. With this design, instrument precision is confounded with site precision and if significant differences in precision are observed between the different sites, it is the sponsor's responsibility to conduct another study to determine if this imprecision is attributable to the sites or the instruments.

c. Presentation of precision studies results

For each sample tested in the precision studies (within-laboratory internal precision study and reproducibility study), we recommend you present the mean value of the signal with variance components (standard deviation and percent CV). In addition, you should include the percent of values above and below the cutoff for each sample in the precision studies. For the reproducibility study, present the mean value with variance components and percent of values above and below the cutoff for each site separately and for the combined data.

We recommend you consult the CLSI documents EP05-A3 [Ref. 7] and EP15-A3 [Ref. 9] for additional information on reproducibility study design and statistical analysis.

(3) Cross-Reactivity

We recommend that you test your device for potential cross-reactivity with other organisms known to colonize the genital tract, including human pathogens that are transmitted by sexual contact. We recommend that you test medically relevant levels of viruses and bacteria (usually 10^5 pfu/ml or higher for viruses and 10^6 cfu/ml or higher for bacteria). We recommend that you confirm the virus and bacteria identities and titers. Titers in particular are usually estimated by suppliers but are not guaranteed. The microorganisms recommended for cross-reactivity studies are listed below in Table 1. Specific species are recommended according to prevalence, clinical relevance, or both, but additional species may also be tested at the discretion of the sponsor. Any additional species selected should be known to colonize the genital tract. Additional organisms

should be tested if there is reason to suspect that cross-reactivity may occur (i.e., clinical evidence of cross-reactivity, homology to chosen probe/primer sequences, etc.).

For devices that target a group of HPV genotypes but do not differentiate among them, you should test the most closely related and/or clinically significant non-targeted HPV genotypes for cross-reactivity. For devices that detect more than one genotype of HPV and further differentiate among them, you should test for cross-reactivity among targeted genotypes. Since HPV cannot be readily cultured, HPV genotypes may be tested as cloned genomic HPV DNA in plasmids or in vitro transcripts, depending upon your targeted analyte.

Table 1. Microorganisms Recommended for Analytical Specificity (Cross-reactivity) Studies.

Organism	
Bacteria:	**Human Papillomaviruses**:
Lactobacillus acidophilus	All non-targeted alpha-HPV genotypes. Alpha HPV genotypes include the following: HPV 16, 18, 26, 30, 31, 33, 34, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 69, 70, 73, 82, 85
Staphylococcus epidermidis	
Staphylococcus aureus	
Streptococcus faecalis	HPV 6, 11
Streptococcus pyogenes	Any non-targeted genital HPV genotypes that are likely to cross-react with your assay based on probe-homology analysis (such as blast search results).
Streptococcus agalactiae	
Corynebacterium spp.	
Chlamydia trachomatis	
Neisseria gonorrhoeae	**Other Viruses:**
Escherichia coli	Adenovirus
Enterococcus spp.	Cytomegalovirus
Clostridium spp.	Epstein Barr virus
Peptostreptococcus spp.	Herpes simplex virus 1
Klebsiella spp.	Herpes simplex virus 2
Enterobacter spp.	
Proteus spp.	
Pseudomonas spp.	
Bacteroides spp.	**Other:**
Bifidobacterium spp.	*Candida albicans*
Fusobacterium spp.	*Trichomonas vaginalis*

(4) Interference

We recommend that you conduct a comprehensive interference study using medically relevant concentrations of the interferent and at least one of the most clinically relevant HPV genotypes (such as HPV 16 or HPV 18) to assess the potentially inhibitory effects of substances encountered in cervical specimens.

DRAFT

Potentially interfering substances include, but are not limited to, the following: whole blood (human), leukocytes, contraceptive and feminine hygiene products. The active ingredients and brand names of selected products and tested concentrations should be provided in your labeling. Examples of potentially interfering substances are listed in Table 2 below. We recommend that you test for interference using specimens with analyte levels that challenge medical decision points around the clinical cutoff (e.g., C_{95}). We also recommend that you evaluate each interfering substance at its potentially highest concentration (i.e., "the worst case"). One way of accomplishing this is to dip a specimen collection device directly into the potentially interfering substance and subsequently place the collection device into one aliquot of a split test specimen. The other aliquot would be tested without the potential interferent so that the signal between the paired samples can be compared. In this approach, both aliquots (with and without the potential interferent) are tested in the same manner as patient specimens with adequate replication (at least four to seven replicates) within one analytical run. An estimate of the observed interference effect as the difference between the means of the two aliquots is computed and the 95% two-sided confidence interval for the interference effect is calculated. If no significant clinical effect is observed, no further testing is indicated. We recommend that you refer to the CLSI document EP07-A2 [Ref. 10] for additional information on interference testing.

Table 2. Substances Recommended for Interference Studies.

Substance
Whole blood (human)
Leukocytes ($1x10^6$ cells/ml)
Contraceptive jelly
Douche
Anti-fungal cream
Spermicide
Vaginal lubricant
Feminine spray
Intravaginal hormones
Mucus

(5) Carry-Over and Cross-Contamination Studies (for devices with automated liquid handling systems)

We recommend that you demonstrate that carry-over and cross-contamination will not occur with your device under your recommended instructions for use. In a carry-over and cross-contamination study, we recommend that high positive samples be used in series alternating with negative samples in patterns dependent on the operational function of the device. At least five runs with alternating high positive and negative samples should be performed. We recommend that the high positive samples in the study be high enough to exceed 95% or more of the results

obtained from specimens of diseased patients in the intended use population. The carry-over and cross-contamination effect can then be estimated by the percent of negative results for the negative samples that are adjacent to high positive samples in the carry-over study compared to the percent of negative results in the absence of adjacent high positive samples (i.e., only negative samples are run on the plate). For additional detail, see Haeckel [Ref. 11]. For devices that are indicated for HPV testing of residual cytology samples, an analysis of the carryover effects of any upstream automated cytology processing system(s) should be provided.

(6) Specimen Storage and Shipping Conditions

For your recommended specimen storage conditions, you should demonstrate that your device generates equivalent results to time zero for the stored specimens at several time points throughout the duration of the recommended storage. Storage temperatures evaluated should represent each extreme of your recommended temperature range. You should establish your specimen storage and shipping conditions utilizing a panel of real clinical samples that represent the specimen types claimed in your intended use and analyte levels that challenge the medical decision point(s) of your assay. The percent change in signal (when compared to time zero) should be presented for each panel member at each time point tested, as well as for all panel members combined. Using regression analysis, each sample should be analyzed separately such that the absolute and percent difference in signal between the recommended storage time and time zero (T_0) should be calculated with a 95% confidence interval. Similar regression analysis should be performed for all panels combined. For these studies, detailed information about the samples used should be recorded and included in your submission. In particular, we recommend you submit the date the specimens were collected from patients in relation to their test date in the clinical study, and also in relation to when they are utilized to establish T_0 for your stability studies (among those used to establish stability).

(7) Reagent Storage and Shipping Conditions

For your recommended reagent storage conditions, you should demonstrate that your device generates equivalent results to time zero utilizing the stored reagents at several time points throughout the duration of the recommended storage. Storage temperatures evaluated should represent each extreme of your recommended temperature range. We recommend that you refer to the CLSI document EP25-A [Ref. 12] for additional information. Accelerated stability studies are appropriate for estimating reagent stability, but the data provided in your submission should show real-time performance. You should establish your reagent storage and shipping conditions utilizing the specimen types claimed in your intended use and analyte levels that challenge the medical decision point(s) of your assay. The percent change in signal (when compared to time zero) should be presented for each panel member at each time point tested, as well as for all panel members combined. Using regression analysis, each sample should be analyzed separately such that the absolute and percent difference in signal between the recommended reagent storage time and T_0 should be calculated with a 95% confidence interval. Similar regression analysis should be performed for all panels combined.

DRAFT

(8) Evaluation of HPV Detection in the Clinical Dataset

We recommend that you provide an evaluation of your device's ability to detect the targeted HPV genotypes in your clinical dataset. One way to do this is to perform an FDA-approved HPV test that detects the same genotypes as your test, or you may perform PCR followed by sequencing of the amplicon (PCR/Sequencing) on your clinical specimens and compare these results to the results of your device. Use of an FDA-approved HPV test is recommended whenever feasible. Use of a composite HPV comparator that incorporates multiple FDA-approved HPV test(s) and/or PCR/Sequencing is also an option. The nucleic acid amplification method used in the composite reference method should be targeted to genomic regions different from the one probed by your assay. You should provide published literature or laboratory data in your submission in support of the primers used for amplification.

For PCR followed by Sanger sequencing, we recommend that you perform the sequencing reaction on both strands of the amplicon (bidirectional sequencing) and the generated sequence should meet all of the following acceptance criteria:

- Sequence contains a minimum of 100 contiguous bases,
- Bases have a Quality Value of 20 or higher as measured by PHRED, Applied Biosystems KB Basecaller, or similar software packages (this represents a probability of an error of 1% or lower), and
- Sequence matches the reference or consensus sequence, e.g. Expected Value (E-Value) < 10-30 for the specific target for a BLAST search in GenBank, (http://www.ncbi.nlm.nih.gov/Genbank/).

As Next Generation Sequencing (NGS), also known as High Throughput Sequencing (HTS), technologies evolve and mature, they may also be used in the composite reference method. Comparator methods based on these technologies should be validated and should meet pre-specified quality metrics. Please contact the FDA Division of Microbiology Devices for additional information on the use of the NGS/HTS methods in evaluation of your clinical data.

A comparison against an FDA-approved HPV genotyping test, or PCR/Sequencing is especially important for HPV genotyping assays to establish that the correct HPV genotype has been identified by your device.

Please note that there are two scenarios in which the samples are found negative by the HPV test when the clinical cutoff is set above the LoB: 1) the HPV test detected some amount of analyte (analyte level is above the LoB) but this amount was below the clinical cutoff that is used to define positive and negative results ("Detected" in Table 3 below = "LoB<signal<clinical cutoff") or 2) the HPV test did not detect the analyte of interest ("Not Detected" in Table 3 below = signal≤LoB). For the comparison of the HPV test and an appropriate comparator discussed above, please describe whether the analyte was detected or not detected for the samples negative by the HPV test as defined above. You should present the comparison for ASC-US (Atypical Squamous Cells of Undetermined Significance) and NILM (Negative for Intraepithelial Lesion or Malignancy) ≥30 populations separately in tables. For a test with only a primary screening indication for population of women ≥25 years, present the data for NILM and

≥ASC-US separately and also for an all-comers population which includes NILM, ≥ASC-US and UNSAT (Unsatisfactory cytology results). An example of a data presentation format is provided below in Table 3.

Table 3. Example of a Data Presentation Format of HPV Detection in the Clinical Dataset.

		Comparator Result			Total
		High Risk Positive	High Risk Negative	Indeterminate	
HPV Positive					
HPV Negative	Detected				
	Not Detected				
Other (Invalid)					
Total					

Evaluation of HPV detection should be presented for each testing site separately and for each type of collection media separately. For the differentiation of HPV genotyping tests, you should present the data comparing all outputs of the HPV test versus the same outputs for the comparator in a table separately for the ASC-US and NILM ≥30 populations. For details, please see Section 9 of CLSI MM17-A [Ref. 16].

An example of a data presentation format for an HPV genotyping test with five possible outcomes: HPV16 Positive, HPV18 Positive, HPV16 & HPV18 Positive, Negative, and Invalid (Indeterminate), is provided in Table 4:

Table 4. Example of a Data Presentation Format for HPV Genotyping Test.

	Comparator Result								
	No High Risk Types	One High Risk Type			Two High Risk Types				Multiple High Risk Types
		16	18	Other	16&18	16&Other	18&Other	Other	
Pos:HPV16									
Pos:HPV18									
Pos:HPV16&18									
Negative									
Other Invalid/Indeterminate									
Total									

DRAFT

B. Clinical Performance Studies

You should provide in your PMA clinical performance studies that conform to the following recommendations.

(1) Consideration of the Cervical Cancer Screening Guidelines

Professional cervical cancer screening guidelines help define the role that an HPV device will play in the larger scheme of patient management and are therefore useful in assessing any intended use statement for an HPV device and its supporting data. The guidelines that will be considered in this guidance are the *2006 Consensus Guidelines for the Management of Women with Abnormal Cervical Cancer Screening Tests* (2006 consensus guidelines) [Ref. 13], along with the 2012 update to these guidelines (2012 consensus guidelines) [Refs. 23 and 24], which are the most current consensus guidelines available on cervical cancer screening to date. Consideration should be given to the latest version of the guidelines as the recommendations may change.

Although professional guidelines are considered in FDA's evaluation, intended uses given for an HPV test are supported primarily by the data submitted for test approval and are generally limited to the populations and sample types evaluated. Studies should be focused on establishing a woman's risk for cervical disease in a given population stratified by the HPV test outcomes. Intended uses for an HPV test may be written more generally (such as the "adjunct" intended use below) to allow clinicians the flexibility to utilize this risk information as they deem appropriate, particularly in the development of future cervical cancer screening guidelines.

(2) Intended Use

The intended use of your device should drive your clinical study design to assess performance, as the intended use will ultimately determine how FDA will review your data. Below is an example of an intended use statement that could be appropriate for a device for detection of HPV:

> The [trade name] HPV Test is a [technology or type of assay] assay for the qualitative detection of high-risk types of human papillomavirus (HPV) [indicate target, such as DNA, RNA transcript or protein] in cervical specimens. The HPV types detected by the assay are the high-risk HPV types [list types –indicate whether test can identify specific types]. Cervical specimens that may be tested with the [trade name] HPV Test include [insert sample types that may be tested by assay and types of collection devices which may be used to collect the samples].
>
> The use of this test is indicated:
>
> 1. To screen patients 21 years and older with atypical squamous cells of undetermined significance (ASC-US) cervical cytology results to determine the need for referral to colposcopy.

DRAFT

2. In women 30 years and older the [trade name] Test can be used with cervical cytology to adjunctively screen to assess the presence or absence of high-risk HPV types. This information, together with the physician's assessment of cytology history, other risk factors, and professional guidelines, may be used to guide patient management.

Below is an example of an intended use statement that could be appropriate for a device for detection and differentiation of HPV:

3. In women 25 years and older, the [trade name] Test can be used as a first-line primary cervical cancer screening test to detect high-risk HPV, including genotyping for 16 and 18. Women who test negative for high-risk HPV types by the [trade name] Test should be followed up in accordance with the physician's assessment of screening and medical history, other risk factors, and professional guidelines. Women who test positive for HPV genotypes 16 and/or 18 by the [trade name] Test should be referred to colposcopy. Women who test high-risk HPV positive and 16/18 negative by the [trade name] Test should be evaluated by cervical cytology to determine the need for referral to colposcopy.

The first intended use will be referred to as the "ASC-US triage" intended use, the second will be referred to as the "adjunct" intended use and the third will be referred to as the "primary screening" intended use throughout this guidance. Study design considerations for specific intended uses are described below, following the more general study design recommendations.

(3) Study Design Considerations Common to ASC-US Triage, Adjunct and Primary Screening Intended Uses (and likely any other intended uses):

For general study design guidance, see FDA's guidance entitled "Design Considerations for Pivotal Clinical Investigations for Medical Devices" (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM373766.pdf) and FDA's guidance entitled "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests" (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm).

a. Use of study sites outside the United States (21 CFR 814.15)

If you rely on clinical data gathered in a study conducted outside the United States to support your PMA and not conducted under an investigational device exemption (IDE), you must ensure that the data are scientifically valid and that the rights, safety, and welfare of human subjects have been protected in accordance with 21 CFR 814.15. To serve as the sole basis for marketing approval, your data must be applicable to the intended population and the United States medical practice (21 CFR 814.15(d)(1)). Areas of concern for studies conducted outside the US include prevalence of specific high-risk HPV strains, patient screening intervals, average age of onset of screening and sexual activity, cervical cancer risk, cervical sampling methods, differences in

DRAFT

medical or clinical practice, and ethnicity. We encourage contacting FDA through the pre-submission process if you intend to seek approval based on foreign data, thus reducing the risk that the foreign study will not support your intended uses.

For additional information about the pre-submission process, see FDA's guidance entitled "Requests for Feedback on Medical Device Submissions: The Pre-Submission Program and Meetings with Food and Drug Administration Staff" (http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM311176.pdf) prior to beginning the study.

b. Histology review

FDA considers results of colposcopy and biopsy (if necessary) to be the clinical reference standard (i.e., gold standard) for the disease assessment of subjects in the clinical study. You may choose to use histology results generated at each of your clinical sites, but we recommend a centralized three expert pathologist review (CPR) panel that will likely generate a more consistent and accurate disease assessment for your study. The three pathologists should distinguish between Cervical Intraepithelial Neoplasia (CIN) 2 and 3, and should not combine these two categories together for reporting purposes (i.e., results of "CIN2/3" should not be reported). If you choose to use a centralized panel and/or clinical sites that utilize the new 2-tiered Lower Anogenital Squamous Terminology (LAST) recommendations [Ref. 25] for reporting the results of cervical histology (as low-grade squamous intraepithelial lesion (LSIL) and high-grade squamous intraepithelial lesion (HSIL), such that immunohistochemical staining for the biomarker p16 is used to clarify any considered intermediate category (CIN2) into either LSIL or HSIL, then the 3-tiered result (i.e. CIN1, CIN2, or CIN3) based on the review of hematoxylin and eosin (H&E) stained slides should be denoted along with the 2-tiered diagnosis which will be based on the review of the H&E and p16 stained slides for every patient. You should also provide FDA with the analytical validation data for any non-FDA approved p16 assay utilized for cervical histology in your study. Please notify FDA prior to beginning a study using the LAST recommendations or any other histology reporting that utilizes the p16 biomarker.

We recommend that the CPR panel establish the clinical reference standard (i.e., clinical truth) for the subject and that two of the three expert pathologists review the slide independently in a masked fashion. If the two pathologists agree, the diagnosis should be considered the clinical reference standard. If there is no agreement, the third expert pathologist should read the slide independently in a masked manner. If there is agreement among any of the three expert pathologist diagnoses, this should be considered the clinical reference standard for the subject. If there is no agreement after the third pathologist review, all three expert pathologists should review the slide together at a multi-headed microscope (or equivalent technology) to try and reach a consensus diagnosis (with majority rule of 2 of the 3 if a complete consensus cannot be reached). When submitting your data, you should provide information on how discordant histology results were resolved. If your CPR panel and/or clinical sites are utilizing the new 2-tiered LAST recommendations, a consensus should be reached among 2 of the 3 pathologists

DRAFT

regarding what the preliminary histopathology diagnosis is prior to p16 staining (i.e., CIN1, CIN2, CIN3) using the same method for establishing clinical truth among the pathologists as when p16 is not used, as described above. A final consensus should be reached after the p16 staining is conducted.

c. Cytology reporting terminology

Collection sites should utilize cytology reporting terminology that can be translated to The 2001 Bethesda System for Reporting Cervical Cytology (2001 Bethesda System), or a more current Bethesda system if and when available [Ref. 14]. Cytology results should be converted to the 2001 (or more current) Bethesda system before reporting the results to FDA.

d. Blinding

Investigators, patients, and clinicians (including those conducting colposcopy and histology) should be blinded to a patient's HPV status until colposcopy/histology is completed to avoid bias in the study. Additional blinding is recommended for HPV primary screening studies as described under "Primary Screening Intended Use" in Section VII(B)(8) below.

Also, the protocol should clearly specify what test results will ultimately be released to the physician and patient as well as under what circumstances the cytology and HPV results will become unblinded as that could inadvertently bias your follow-up study.

e. Human papillomavirus genotypes

For an assay to detect high risk human papillomavirus, the following genotypes categorized as "carcinogenic" by the World Health Organization International Agency for Research on Cancer (IARC) should be targeted: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59 [Ref. 15]. If your assay does not target any of these recommended HPV genotypes, you should explain why. Additional genotypes, such as those deemed "probably carcinogenic" or "possibly carcinogenic" by IARC (i.e. types 66, 68) may also be included. We recommend that you discuss with FDA the benefits and risks of inclusion of any other human papillomavirus genotypes prior to beginning your studies.

f. Specimen collection media

We recommend you perform the described analytical and clinical studies for each type of specimen collection media (i.e., specific brand of liquid-based-cytology collection fluid) claimed in your intended use. Clinical performance should be presented for each collection media separately.

g. Specimen collection devices

The list of collection devices that may be used to collect specimens for testing by your device should be described in the intended use statement and should be approved for use with your

DRAFT

indicated cytology method(s). Each claimed collection device (i.e., brush/spatula vs. broom) need not be evaluated in your analytical studies. However, each indicated collection device should be evaluated in your clinical studies. Clinical performance should be presented for each collection device separately.

h. Specimen collection – general

For each specimen collected in your clinical study you should capture the date the specimen was collected from the patient, the date it was shipped to and received by the testing laboratory, and the date the specimen was tested.

i. Biopsy methods

The biopsy methods utilized should be consistent for all patients and all sites within each study. If separate studies are conducted for distinct indications (e.g., ASC-US triage vs. adjunct), then different biopsy methods may be used for each study. If the biopsy method is not consistent within a dataset for a given indication, it may lead to bias in your study that may prevent proper establishment of your performance characteristics for that indication. A standardized biopsy method can have variables associated with it, but these variables should be associated with the appearance of the cervix upon visualization during colposcopy, such as the presence or absence of visible lesions, or the visibility of the squamocolumnar junction (SCJ). If additional variables are desired, you should discuss them with FDA prior to beginning your studies. Note that biopsies taken from lesioned and non-lesioned areas should be denoted differently on your case report forms.

j. Cytology sample aliquoting

Sponsors pursuing intended uses for HPV testing from cytology samples should consider, when designing their studies, whether they should be testing from pre-aliquoted cytology samples (aliquot taken prior to slide processing) or working from residual cytology samples (aliquot taken after slide processing). Pre-aliquoting of cytology samples can only occur if the cytology collection system has been approved for aliquot removal prior to cytology slide processing. This will ensure that patient cytology test results are not compromised by inappropriate processing of their cytology specimens.

Alternatively, sponsors who seek a claim to work from residual cytology specimens should analytically assess the effects of carryover during cytology slide processing (see Section VII(A)(5) "Carry-Over and Cross-Contamination Studies"). Sponsors with amplification assays who have concerns about contamination may need to work with alternative specimen collection systems or systems approved for pre-aliquoting to address their contamination issues. Sponsors who perform their clinical studies on pre-aliquoted cytology specimens who subsequently seek a claim to test residual cytology specimens should compare the results of a panel of paired pre-aliquoted and residual real clinical cytology specimens, in addition to conducting the analytical carryover and cross contamination studies discussed above.

DRAFT

k. Reporting results for HPV genotyping assays

Results should be reported in a manner readily interpretable by clinicians. Groups of HPV genotypes with similar risk levels should be reported in groups, instead of individually, where appropriate.

In your PMA, you should describe how each of the reported results or invalid results are determined, and how they should be interpreted. You should indicate the cut-off values for all outputs of the assay.

If the assay has an invalid result, you should describe 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. You should provide 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).

l. HPV vaccination and study populations

When making sample size estimations, you should consider that as the number of HPV-vaccinated individuals increases, this will lead to a decrease in the overall prevalence of cervical disease in the United States. Current estimates of vaccine rates and disease prevalence should be taken into account when estimating study sample size. Inclusion of study sites with higher than average levels of non-vaccinated individuals may eventually become advisable as the number of vaccinated individuals across the US increases. Please note that, in this scenario, study sites with average levels of vaccinated individuals should also be evaluated. Sponsors considering this type of design should discuss this option with FDA before beginning their studies.

(4) ASC-US Triage Intended Use

You should conduct prospective clinical studies using specimens representing the intended use population, i.e. patients with ASC-US cervical cytology results, to determine clinical performance of your device for all specimen types and specimen collection devices you claim in your labeling. The clinical performance of a qualitative test (test with two outcomes, Positive or Negative) is described by its clinical sensitivity and specificity, positive and negative predictive values, and prevalence. The clinical sensitivity of your device is the proportion of individuals who have precancer or cancer [greater than or equal to Cervical Intraepithelial Neoplasia 2 (≥CIN2)[2]] that are positive by your test. The clinical specificity of your device is the proportion of individuals who do not have precancer or cancer (<CIN2) that are negative by your test. These performance characteristics should be established in prospective clinical studies conducted

[2] Please note that for all places in this document where data for the ≥CIN2 target condition is requested, data for the ≥CIN3 target condition should also be presented since ≥CIN3 is more likely to progress to cervical cancer. If you utilize the 2-tiered LAST system, you should also present data for HSIL and above. HSIL in the LAST system is different from HSIL in The 2001 Bethesda System for Reporting Cervical Cytology and this distinction should be made clear by indicating for each table containing "HSIL" results whether cytology or biopsy results are being reported.

at a minimum of three study sites that are representative of clinical sites in the United States. For a test with more than two outcomes, clinical performance is described by likelihood ratios, percentage of subjects in each outcome and prevalence.

a. Specimen collection and processing

Proper specimen collection and processing is critical for establishing the performance characteristics of an HPV test. For an ASC-US triage intended use, the population of women studied should be recruited from Ob/Gyn clinics. Please note that colposcopy clinics are not good sources of patients for an ASC-US triage evaluation, as the women who present at colposcopy clinics have already been determined to be in need of colposcopy (i.e., have already been determined to be HPV positive by other tests, or repeat ASC-US by cytology). Since women who are already known to need colposcopy are not the target population for the ASC-US triage intended use, this population should not be used for your study, as the performance estimates derived would be inaccurate. The population of women who present at a colposcopy clinic has a higher prevalence of both HPV infection and cervical disease and, due to verification bias, device sensitivity would be overstated.

For tests that are to be performed directly from liquid-based cytology (LBC) specimens, all investigative HPV test results should be performed on the same LBC sample that was used to generate the cytology result. This will enable you to avoid any sampling bias in your study (i.e., infections that may resolve between the time the original cytology sample and investigative sample are taken, removal of a large portion of the HPV infected cells in the first sample, etc.). Although one approach to mitigating sampling bias when collecting an extra sample is to randomize the test procedures performed on the two samples (i.e., cytology and HPV testing), this is not an acceptable approach for generating a cytology result in patients. The first cytology sample taken from a patient should always be the sample utilized to generate a cytology result, so that this result (and subsequently, the health of the patient) is not compromised. Therefore, randomizing testing on two cytology samples would not mitigate sampling bias for HPV studies.

One challenge in enrolling patients from Ob/Gyn clinics as opposed to colposcopy clinics is fielding the large number of women who are not part of the intended use population. If you are conducting a large study to support multiple HPV testing intended uses, it may be advisable to enroll all women, regardless of cytology status, into your study. Another option, if the ASC-US triage intended use is to be pursued in a separate study, is to enroll only patients with ASC-US cervical cytology results into your prospective clinical study. When utilizing the latter approach, it is important to establish a procedure for obtaining the original cytology sample that was originally used to generate the enrollment ASC-US result in order to avoid sampling bias as described above.

b. Clinical reference ("Gold") standard

Your study should be designed such that all women with ASC-US cytology from Ob/Gyn clinics will proceed to colposcopy, regardless of HPV status or other factors. Investigators, patients and

clinicians (including those conducting colposcopy and histology) should be blinded to a patient's HPV status until colposcopy/histology is completed to avoid bias in the study.

Time elapsed between collection of a screening cervical cytology specimen and subsequent colposcopy procedures should not exceed 12 weeks. Allowing too much time between these procedures could result in higher than normal rates of spontaneous regression of HPV infections and their associated cervical lesions, which will adversely affect your estimates of clinical sensitivity and specificity.

You should describe details of the colposcopy procedures used in your clinical study and the results of the colposcopy procedures should be categorized as (Negative Colposcopy/No Biopsy), Negative Biopsy, CIN 1, CIN 2, CIN 3 and Cancer. If you utilize 2-tiered LAST system, in addition, present the results of the colposcopy procedure as (Negative Colposcopy/No Biopsy), Negative Biopsy, LSIL, HSIL, and Cancer.

c. Clinical performance evaluation of HPV tests

The clinical performance of a test for the detection of HPV (a qualitative test) is described by its clinical sensitivity and specificity, and by its positive and negative predictive values, along with the prevalence of the target condition in the intended use population.

An example of an acceptable data presentation format for a qualitative test with two outcomes (Positive and Negative) is provided below in Table 5:

Table 5.—Table for Qualitative HPV Test Results

		Central Histology				
	Neg Colpo	Neg	CIN1	CIN2	≥CIN3	
HPV Pos	A_1	A_2	A_3	A_4	A_5	$A_1+A_2+A_3+A_4+A_5$
HPV Neg	B_1	B_2	B_3	B_4	B_5	$B_1+B_2+B_3+B_4+B_5$
Total	A_1+B_1	A_2+B_2	A_3+B_3	A_4+B_4	A_5+B_5	N

The clinical performance of your device for the target condition "CIN2 and above" (≥CIN2) should be evaluated as follows:

Sensitivity = $(A_4+A_5)/(A_4+A_5+B_4+B_5)$;

Specificity = $(B_1+B_2+B_3)/(A_1+A_2+A_3+B_1+B_2+B_3)$

Positive Predictive Value (PPV)=$(A_4+A_5)/(A_1+A_2+A_3+A_4+A_5)$

Negative Predictive Value (NPV)=$(B_1+B_2+B_3)/(B_1+B_2+B_3+B_4+B_5)$

Prevalence of ≥CIN2=$(A_4+A_5+B_4+B_5)/N$

Since CIN3 lesions are more likely to progress to cervical cancer than CIN2 lesions [Ref. 17], the clinical performance of your device for the target condition "CIN3 and above" (≥CIN3) should also be presented:

Sensitivity = $A_5/(A_5+B_5)$;

Specificity = $(B_1+B_2+B_3+B_4)/(A_1+A_2+A_3+A_4+B_1+B_2+B_3+B_4)$
PPV=$A_5/(A_1+A_2+A_3+A_4+A_5)$
NPV=$(B_1+B_2+B_3+B_4)/(B_1+B_2+B_3+B_4+B_5)$
Prevalence of ≥CIN3=$(A_5+B_5)/N$

The estimates of sensitivity and specificity as well as positive and negative predictive values should be provided along with 95% two-sided confidence intervals. For the 95% confidence intervals for sensitivity and specificity, a score method is recommended. For more details about score confidence intervals, see Section IX Appendix "Statistical Analysis" and CLSI EP12-A2 [Ref. 8]. The confidence intervals for the predictive values can be calculated (when prevalence is constant) based on the confidence intervals of the corresponding likelihood ratios (an estimate of the likelihood ratio is a ratio of two independent proportions; therefore, the confidence intervals for a ratio of two independent proportions can be used; see Section IX Appendix "Statistical Analysis").

The clinical performance for the target condition ≥CIN2 should be stratified by age. The prevalence of ≥CIN2, sensitivity, specificity, PPV and NPV along with 95% CI should be presented for each of the following age groups: 21-30, 30-39, and >39.

d. Sample size

When considering sample size for an ASC-US triage intended use, one should consider the number of samples from ASC-US patients needed to establish point estimates of clinical sensitivity and specificity, along with the lower limits of 95% two-sided confidence intervals. Clinical sensitivity for cervical disease (≥CIN2) is the most critical performance parameter for an HPV test, since a false negative HPV test result could lead to delays in cervical cancer detection and treatment [Ref. 13].

If the estimated clinical sensitivity, specificity and subsequent positive and negative predictive value(s) of your device do not meet current performance expectations for HPV testing [Ref. 23 and 24], panel review of your performance data may be necessary to allow assessment of the clinical effectiveness of your test.

e. Selection of appropriate clinical cutoff for HPV tests

Selection of the appropriate clinical cutoff can be justified by the relevant levels of sensitivity and specificity that are based on Receiver Operating Curve (ROC) analysis of pilot studies with clinical samples. The clinical performance of the HPV test at the selected clinical cutoff is ideally estimated using a pivotal clinical study. In some circumstances, the clinical cutoff can be determined during the pivotal clinical study using an unbiased procedure and appropriate sample size. If the level of sensitivity that is clinically acceptable is pre-specified (e.g., the level of sensitivity of 93%-95% is clinically acceptable in the intended use population), then the pivotal study can be used to establish the clinical cutoff corresponding to the pre-specified level of sensitivity and to obtain an unbiased estimation of the clinical performance of the HPV test with this selected cutoff [Ref. 18 & Ref. 19].

DRAFT

(5) ASC-US Population - HPV Tests for Detection and Differentiation (HPV Genotyping Tests)

The study principles described in the preceding Section VII(B)(4) "ASC-US Triage Intended Use" to establish clinical sensitivity and specificity for ≥CIN2 in women with ASC-US cytology apply to both dual outcome HPV tests (positive or negative for HPV) and multiple outcome HPV genotyping tests.

A test for the detection and differentiation of HPV genotypes usually has multiple outcomes (e.g., HPV16+, HPV18+, HPV16/18+, etc.). The clinical performance of a test for the detection and differentiation of HPV genotypes is described by the probabilities of a target condition for each outcome of the test, as well as the percent of study subjects with each outcome of the test along with disease prevalence for each outcome.

In addition to establishing clinical sensitivity and specificity for ≥CIN2 in an ASC-US population for an HPV genotyping test, likelihood ratios for each test outcome and the percent of study subjects with each test outcome should also be established as described below.
An example of an acceptable data presentation format for a HPV genotyping clinical study in the ASC-US population is provided in Table 6 (Example shown has outcomes: HPV16+, HPV18+, HPV16/18+, etc.):

Table 6.—Data Presentation for HPV Genotyping Test Results

		Central Histology				
	Neg Colpo	Neg	CIN1	CIN2	≥CIN3	
Pos:HPV16	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	$A_{11}+A_{12}+A_{13}+A_{14}+A_{15}$
Pos:HPV18	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	$A_{21}+A_{22}+A_{23}+A_{24}+A_{25}$
Pos:HPV16&18	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	$A_{31}+A_{32}+A_{33}+A_{34}+A_{35}$
……	……	……	……	…..	…..	……
Total						N

The clinical performance of such a test for the target condition ≥CIN2 is evaluated by the likelihood ratio for each test outcome X and the percent of study subjects with each test outcome. The likelihood ratio (LR) for the test outcome X, LR(T=X), summarizes how many times more (or less) likely subjects with the disease (≥CIN2) are to have that particular result X, Pr(T=X|D+), than subjects without the disease, Pr(T=X|D-): LR(T=X) = Pr(T=X|D+)/Pr(T=X|D-).

The following calculations should also be described in your PMA for HPV genotyping tests:

- The likelihood ratios for each of K outcomes (K is a number of different outcomes) should be calculated along with 95% confidence intervals.

- In addition to the likelihood ratios, probabilities that a patient has ≥CIN2 for each of the K outcomes of the test should be calculated along with 95% confidence intervals. As an illustrative example, the probability for Outcome HPV16 pos is evaluated as: Probability (≥CIN2|HPV16 pos)=$(A_{14}+A_{15})/(A_{11}+A_{12}+A_{13}+A_{14}+A_{15})$.
- Also present the percent of each of K outcomes in the clinical data set. For example, for outcome HPV16 pos: Probability (HPV16 pos)=$(A_{11}+A_{12}+A_{13}+A_{14}+A_{15})/N$ (where N is the total number of women with results).
- In addition, probability of ≥CIN2 for the combined outcomes HPV16/18+ (HPV16/18 is defined positive if either HPV16+ or HPV18+ or both) should be calculated as: Probability (≥CIN2| HPV16/18+)=$(A_{14}+A_{15}+A_{24}+A_{25}+A_{34}+A_{35})/(A_{11}+A_{12}+A_{13}+A_{14}+A_{15}+A_{21}+A_{22}+A_{23}+A_{24}+A_{25}+A_{31}+A_{32}+A_{33}+A_{34}+A_{35})$ and the percent of the subjects with HPV16/18+ results.
- Prevalence of disease (≥CIN2) should also be calculated.

The confidence intervals for the probabilities of ≥CIN2 can be calculated based on the confidence intervals of the corresponding likelihood ratios.

In a similar way, the clinical performance of the HPV test should be estimated for the target condition ≥CIN3 (or HSIL and above if 2-tiered LAST system is used).

(6) Adjunct Intended Use

a. General study design options

Per the 2012 consensus guidelines [Ref. 23], in women 30 years and older, HPV testing is recommended as an adjunct to cytology primarily in women with normal cytology. Establishing the clinical sensitivity and specificity of your device in a population of women with normal cytology is complicated by the fact that these women are not typically sent for colposcopic examination at the time when HPV testing is done due to their low incumbent risk of cervical cancer. However, a subset of women with normal cytology will have cervical abnormalities (≥CIN2) [Ref. 20]. HPV testing may help identify the subset of women 30 years and older with normal cytology who are at a higher risk for cervical cancer. To demonstrate that your device is capable of identifying this higher risk subset of women, you should estimate the absolute risks and the relative risk for ≥CIN2 in this population for individuals positive vs. negative by your assay as described below. Estimating absolute risks and relative risk for this intended use population can be accomplished with at least one of the following prospective clinical study designs:

1. Characterize a population of women 30 years and older with normal cytology as positive or negative by your investigative device and establish agreement with a valid comparator HPV detection device, such as an FDA-approved HPV detection device, or PCR/sequencing at baseline; then follow the women at yearly time intervals for a minimum of three years. See Section VII(A)(8) “Evaluation of HPV Detection in the Clinical Dataset” above for more details on the baseline analysis of the prospective clinical dataset with respect to HPV detection. All women who develop abnormal

DRAFT

cytology (≥ASC-US) during follow-up should be sent to colposcopy regardless of their HPV status. Unlike option two below, women with normal cytology are not sent to colposcopy with this study design, particularly if the decision to send to colposcopy is based on investigative or approved HPV test results[3] at the baseline visit or during follow-up (this is important for avoiding bias). Note that follow-up will end for women who have a ≥CIN2 colposcopy result at any time during the study and that these women are considered "disease positive." Women who go to colposcopy but are <CIN2 should continue to be followed for the remainder of the study duration. The follow-up data should demonstrate a statistically and clinically significant difference in relative risk for ≥CIN2 at least at three years for women who are positive as compared to women who are negative by your device at enrollment. In addition, the absolute risk of ≥CIN2 at three years for the women who were negative by your test at the baseline should be evaluated along with 95% confidence interval. The data should demonstrate that this absolute risk is low enough to ensure your test can be safely used for adjunctive screening in women 30 and older. In addition, present the overall risk of ≥CIN2 (regardless of the HPV status at the baseline). The data analysis should be stratified by age groups (30-39 years and 40+ years). The longitudinal follow-up portion of this study may potentially be conducted post-approval (see "Longitudinal follow-up" Section VII(B)(6)(c) below for more details).

2. Characterize a population of women 30 years and older with normal cytology as positive or negative by your device at baseline, then send a subset of those women to colposcopy. You should also establish agreement with a valid comparator HPV detection device with this second study design option, but a smaller subset of samples may be evaluated since you will have more information on risk of ≥CIN2 at baseline. It is recommended you send all HPV positive women (by investigative and/or approved tests) and a random subset of HPV negative women to colposcopy. The data should demonstrate a statistically and clinically significant difference in relative risk for ≥CIN2 for women who are positive vs. negative by your device at enrollment. Using multiple imputation, absolute risks of ≥CIN2 for the subjects positive and negative by your device should be calculated. For the HPV test for detection and differentiation, the data should also demonstrate that the absolute risk of ≥CIN2 at the baseline for some positive outcomes is high enough to demonstrate effectiveness of the test in the intended use population. Because only a random sample of HPV negative women will have been sent to colposcopic examination, the data have a verification bias and therefore, an appropriate statistical method such as multiple imputation [Ref. 21] should be used for calculation of the absolute and relative risks.

b. Enrollment of "all comers" for adjunct claim:

Please note that FDA now recommends that you enroll a population of women 30 years and older undergoing routine screening ("all comers") in your clinical study to evaluate performance

[3] An exception would be if a woman was twice cytology negative and HPV positive (at consecutive yearly visits) – in this scenario she should be sent to colposcopy per the 2012 consensus guidelines [Ref. 23]. The bias created in this situation is unavoidable as patient health is paramount.

of an adjunct claim. Women are HPV tested in parallel with cytology under this claim; therefore, women with higher grade cytology are included under the adjunct claim. FDA has historically allowed sponsors to submit only the actionable data on the NILM ≥30 women, but FDA is now considering that 1) the results of an HPV test in women with higher grade cytology (>ASC-US) is proving to be an important safety signal in terms of how well the test is working overall, 2) a study size that includes all comers ≥30 years old includes approximately 3% of women with >ASC-US cytology and these women usually have colposcopy/biopsy results according to the current medical practice, and 3) the 2012 guidelines are including more actionable combinations of HPV and cytology results that result from co-testing. In light of this, FDA will be evaluating all cytology categories when reviewing data to support the adjunct claim going forward to ensure the safety and effectiveness of HPV IVD devices. Women with ≥ASC-US cytology in the "all comers" adjunct study population should be sent immediately to colposcopy. All patients in the adjunct study population who go to colposcopy but do not have histology ≥CIN2 by the CPR panel should be invited to participate in the three year longitudinal study.

c. Longitudinal follow-up

Given that establishing clinical sensitivity and specificity in a population of women with normal cytology involves either a very large sample size and/or long term patient follow-up, the FDA has considered options that would allow faster access to these important devices while assuring their safety and efficacy. The FDA believes that in cases where an HPV test is receiving, or has received, approval for the ASC-US triage intended use and where the test has shown a high degree of clinical sensitivity for cervical precancer/cancer (≥CIN2), there is a high degree of confidence that the test performs at a level consistent with current expectations for HPV testing [Ref. 2]. In such cases, to receive the adjunctive intended use for the same HPV test, FDA may provide for the longitudinal follow-up portion of the adjunctive study described in Option 1 of Section VII(B)(6)(a) above to be completed post-market, as long as it has been shown that HPV detection by the investigative test in the prospectively collected NILM 30 and older (NILM ≥30) dataset is comparable to HPV detection in the ASC-US population. In this scenario, the same patients from the prospectively collected NILM ≥30 dataset for whom HPV detection characteristics have been established will be followed longitudinally as part of a post-approval study to establish the cumulative three year risk of precancer/cancer in patients positive vs. negative by the investigative HPV test in this population. This approach will be considered for tests that detect HPV types that are supported for use in the NILM ≥30 population by current clinical practice guidelines [Ref 13]. Please note that post-market studies for devices of this type are only appropriate when the degree of uncertainty about certain risks or benefits is acceptable in the context of the overall benefit-risk profile of the device at the time of premarket approval. Also, see FDA's guidance entitled, "Procedures for Handling Post-Approval Studies Imposed by PMA Order" (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070974.htm). Sponsors should contact FDA to discuss their eligibility to complete their longitudinal evaluation post-market.

DRAFT

(7) Adjunct Intended Use – HPV Tests for Detection and Differentiation (HPV Genotyping Tests)

The study options described in the preceding Section VII(B)(6) "Adjunct Intended Use" to establish relative risk for ≥CIN2 in women 30 and over with normal cytology can be applied to both dual outcome HPV tests (positive or negative for HPV) and multiple outcome HPV genotyping tests (tests that not only detect, but also differentiate between the different HPV types). The more outcomes an HPV genotyping test has, the more challenging it is to demonstrate a statistically significant difference in the relative risk of each outcome.

In light of recommendations in the 2012 consensus guidelines [Ref. 23], an additional option you may wish to pursue for an HPV genotyping assay (aside from the more general adjunct screening intended use) is a specific NILM ≥30 colposcopy triage intended use for the highest risk HPV genotypes, such as HPV 16 and 18. The principles of this type of study design and evaluation would be very similar to ASC-US triage, except that you would be dealing with a different study population and test outcomes. If you wish to pursue such an intended use, please contact FDA for further assistance.

a. HPV testing in women 30 and over with >ASC-US cytology

In order to allow for parallel cytology and HPV testing, an adjunct intended use need not be limited to women with normal cytology (i.e., there is no need to wait for the cytology result to order an HPV test). For all HPV devices with an adjunct intended use, the labeling should indicate that a negative HPV result for a woman 30 years and older with >ASC-US cytology should not prevent women from going to colposcopy.

(8) Primary Screening Intended Use

An advisory meeting of the Microbiology Panel was held on March 12, 2014 which resulted in a newly approved indication for an HPV diagnostic device: Primary Cervical Cancer Screening. If you are interested in pursuing a primary screening claim, we recommend you review the information on this meeting, which can be found on the CDRH Website at http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/MicrobiologyDevicesPanel/ucm388531.htm

To evaluate an HPV primary screening indication, you should characterize a population of women 25 years and older undergoing routine screening (an all comers approach) as HPV positive or negative by your device at baseline, then send a subset of those women to colposcopy. It is recommended you send all HPV positive (by investigative and/or approved tests), cytology positive, and a random subset of HPV negative women with normal cytology to colposcopy. Women with unsatisfactory (UNSAT) cytology results should also be sent to colposcopy to assess the risk of disease in these women stratified by their HPV test results.

The data should demonstrate that primary HPV screening with your device shows acceptable clinical performance in detecting ≥CIN2 and ≥CIN3 (sensitivity, specificity, PPV, NPV, absolute

DRAFT

risk and likelihood ratios) when compared to accepted cervical cancer screening methods [Refs. 13, 23, 24, and 25]. For tests intended for detection and differentiation, the data should also demonstrate that the absolute risk of ≥CIN2 at the baseline for positive outcomes is high enough and the absolute risk of ≥CIN2 at the baseline for negative outcomes is low enough to demonstrate effectiveness of the test in the intended use population. Because only a random sample of HPV negative women will have been sent to colposcopic examination, the data will have a verification bias, and therefore, an appropriate statistical method such as multiple imputation should be used for calculation of the absolute and relative risks [Ref. 21].

a. Longitudinal follow-up

All patients in the primary screening study population who underwent colposcopic examination and did not have histology ≥CIN2 by CPR should be invited to participate in a three year longitudinal study. Subjects in the follow-up study should undergo annual visits for cervical sampling for cytology, and all subjects with ≥ASC-US should be invited to proceed to colposcopy. Colposcopy and biopsies should be performed in a standardized manner as described in Section VII(B)(3)(b) "Histology review" above. All cervical biopsies should be examined by the CPR panel. All subjects with ≥CIN2 by CPR should exit the study and those with <CIN2 by CPR should be invited to proceed to the next year's follow-up visit. In order to maximize disease ascertainment, it is recommended that an exit colposcopy and endocervical curettage (ECC) be considered for all follow-up subjects in Year 3.

You should consider enrolling a random subset of women who were cytology and HPV negative (by the investigative device and at least one FDA approved device) and were not selected for colposcopy at baseline in a parallel longitudinal study where these women are sent to colposcopy after a 3 year screening interval. This will allow a more accurate assessment of ≥CIN3 risk over a three year screening interval for these women. Three year risk estimates for ≥CIN3 obtained in this manner will not be affected by detection and treatment of CIN2 lesions at baseline.

b. Blinding issues for primary screening study

Investigators, patients, and clinicians (including those conducting colposcopy and histology) should be blinded not only to a patient's HPV status, but their cytology status as well, until colposcopy/histology is completed to avoid bias in the study. Patients should be flagged as requiring colposcopy/histology without specifying the test results associated with the referral.

Cytologists are intentionally blinded to all other patient test results for a primary HPV screening study to avoid biasing their assessment of the cytology slides based on the knowledge of other test results (otherwise performance of cytology alone as a comparator algorithm could be potentially biased). However, cytology performance could be different in a real-life setting in the context of using an HPV test as a primary screening device when cytologists know the HPV status of the specimens they are screening. To assess how different the performance of HPV primary screening with the investigative device could be in this real-life setting, a subset of cytology slides should be re-read at the testing sites with knowledge of the HPV status available

DRAFT

at the time of the repeat reading. The effect of this un-blinding on the performance of your device for HPV primary screening should be determined.

c. Benefit-risk analysis

You should evaluate the benefit-risk of your HPV primary screening device for detection of high-grade cervical disease (CIN2, ≥CIN3) relative to current screening methods. To permit an evaluation of benefit-risk, you should present estimates of the expected number of tests and procedures (number of cytology, HPV tests, and colposcopy procedures), and the expected number of true positives, false positives, and false negatives per 10,000 screened women for both your device and current screening method(s). This analysis should be performed considering both blinded and un-blinded performance estimates for HPV primary screening with regard to the cytologists' knowledge of HPV results. For more information, please see FDA's guidance entitled "Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and *De Novo* Classifications" at (http://http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379.pdf).

d. Performance in women with cancer

Sponsors seeking a primary screening claim should ensure that their test is highly sensitive not only for cervical pre-cancer, such as CIN2 and CIN3, but for cervical cancer itself. Therefore, additional testing with the investigational device should be conducted on cytology samples collected from females who have subsequently been diagnosed with cervical cancer, using banked samples from the United States. We recommend submitting a separate protocol for this study to FDA for review. In your protocol, please provide information on the collection of these samples and a detailed description of how you plan to blind these samples.

e. General considerations

Although the basic elements of an HPV primary screening study are provided above, this type of study is complex and therefore, it is strongly recommended that any sponsor seeking an HPV primary screening indication should submit their detailed clinical protocol to FDA for review via the Pre-Submission Program.

(9) Study Design to Cover All Three HPV Testing Claims (ASC-US Triage, Adjunct and HPV Primary Screening)

Please note that the study design described above to evaluate the HPV Primary Screening intended use can also be used to evaluate the ASC-US Triage and Adjunct intended uses, noting that some additional women will need to be enrolled who are 21-24 years old with ASC-US cytology. These young women with ASC-US can be invited into the study after their cytology result is known, and therefore, the number of additional women needed to evaluate all three HPV testing claims in a single study will be minimal (i.e. beyond the number needed for a primary

screening claim). Please see recommendations under Section VII(B)(4) "ASC-US Triage Intended Use" above for considerations when evaluating ASC-US Triage performance from a population of women enrolled based on their cytology result (in particular recommendations on the source of these specimens).

C. Controls

When conducting the performance studies described above, we recommend that you run appropriate external controls every day of testing for the duration of the analytical and clinical studies. Since HPV cannot be readily cultured, appropriate external controls include HPV genomic DNA contained within plasmids or synthetic HPV RNA transcripts (depending on whether your test targets HPV DNA or RNA) in a matrix that mimics clinical samples as closely as possible. The HPV genotype(s) selected for use in your controls should be among the most clinically relevant HPV genotypes (e.g., HPV 16). As the clinical significance of HPV strains shift due to vaccination programs, appropriate control sequences may need to be re-assessed.

We recommend that you consult with FDA when designing specific controls for your device. If your device is based on nucleic acid technology, we recommend that you include the following types of controls:

(1) External Controls

a. Negative control

The negative external control contains an appropriate buffer or sample transport media and is run through the entire assay process in the same manner as a clinical specimen. This control is used to rule out contamination with target nucleic acid or increased background in the amplification and/or detection reaction.

b. Positive control

The positive external control contains target nucleic acids at levels approximately two-fold above the C_{95} concentration of the assay in an appropriate buffer or sample transport media, and is run through the entire assay process in the same manner as a clinical specimen. For a test that targets HPV DNA, the cloned HPV 16 genome in carrier plasmid DNA suspended in sample transport media would be an appropriate control. The complete targeted conserved region of the HPV 16 genome, such as the L1 region, can also be utilized in lieu of a full-length genomic clone. For a test that targets HPV RNA transcripts, synthetic full-length transcripts of the targeted genes suspended in sample transport media would be an appropriate control. For controls with analyte levels that do not adequately challenge medical decision points, as part of ensuring compliance with 21 CFR 809.10(b)(8)(vi), the following warning should be included in the labeling:

"The Positive and Negative Controls are intended to monitor for substantial reagent failure. The Positive Control should not be used as an indicator for cut-off precision and only ensures reagent

DRAFT

functionality. Quality control requirements must be performed in conformance with local, state and/or federal regulations or accreditation requirements and your laboratory's standard Quality Control procedures."

For manufacturers that do not wish to provide external controls, instructions should be included in the package insert to instruct end-users how to make their own external controls. This option is only appropriate for devices that include an internal positive control for all samples and that do not have a primary screening claim.

(2) Internal Control

The internal control is a non-target nucleic acid sequence that is co-processed (i.e., extracted and amplified) with the target nucleic acid. It controls for integrity of the reagents (polymerase, primers, etc.), equipment function (thermal cycler), and the presence of inhibitors in the samples. Examples of acceptable internal control materials include human nucleic acid co-processed with the HPV and primers amplifying human housekeeping genes (e.g., RNaseP, β-actin). An internal control for a human "housekeeping" gene may also help ensure adequate cellular sampling of the aliquot material. This type of control is needed for all HPV devices with a primary screening claim to help reduce the likelihood of a false negative result; otherwise, the need for this control should be determined on a device case-by-case basis [Ref. 22].

VIII. References

1. de Villiers EM, Fauquet C, Broker, TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology. 2004; 324(1):17-27.

2. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. Journal of Pathology. 1999; 189(1):12-19.

3. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ, International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. New England Journal of Medicine. 2003; 348(6):518–527.

4. Clinical and Laboratory Standards Institute. 2012. Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline – Second Edition. EP17-A2.

5. Linnet, K. and Kondratovich, M. Partly nonparametric approach for determining the limit of detection. Clinical Chemistry. 2004; 50(4), 732-740.

6. Weusten J.J. A statistical approach to assess analytical sensitivity of diagnostic assays in the presence of false-positive results. Journal of Virological Methods; 151(2):301-10.

DRAFT

7. Clinical and Laboratory Standards Institute. 2014. Evaluation of Precision of Quantitative Measurement Procedures; Approved Guideline. EP05-A3. Clinical and Laboratory Standards Institute, Wayne PA.

8. Clinical and Laboratory Standards Institute. 2002. User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline. EP12-A2. Clinical and Laboratory Standards Institute, Wayne PA.

9. Clinical and Laboratory Standards Institute. 2014. User Verification of Precision and Estimation of Bias; Approved Guideline. EP15-A3. Clinical and Laboratory Standards Institute, Wayne PA.

10. Clinical and Laboratory Standards Institute. 2005. Interference Testing in Clinical Chemistry; Approved Guideline. EP07-A2. Clinical and Laboratory Standards Institute, Wayne PA.

11. Haeckel R. Proposals for the description and measurement of carry-over effects in clinical chemistry. Pure Applied Chemistry 1991; 63:301-306.

12. Clinical and Laboratory Standards Institute. 2009. Evaluation of Stability of In Vitro Diagnostic Reagents; Approved Guideline. EP25-A. Clinical and Laboratory Standards Institute, Wayne PA.

13. Wright T, Massad L, Dunton C, Dunton J, Spitzer M, Wilkinson E, et al. 2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests. American Journal of Obstetrics and Gynecology. 2007; 197(4):346-55.

14. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, et al. Forum Group Members; Bethesda 2001 Workshop. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA. 2002; 287(16):2114-9.

15. Schiffman M, Clifford G, Buonaguro F. Classification of weakly carcinogenic human papillomavirus types; addressing the limits of epidemiology at the borderline. Infectious Agents and Cancer. 2009; 4:1-8.

16. Clinical and Laboratory Standards Institute. 2008. Verification and Validation of Multiplex Nucleic Acid Assays: Approved Guideline. MM17-A. Clinical and Laboratory Standards Institute, Wayne PA.

17. Schiffman M, Rodriguez AC. Heterogeneity in CIN3 diagnosis. Lancet Oncology. 2008; 9(5):404-6.

18. Linnet K., Brandt E. Assessing diagnostic tests once an optimal cutoff point has been selected. Clinical Chemistry. 1986; 32(7): 1341-6.

DRAFT

19. Kondratovich M, Yousef WA. Evaluation of accuracy and 'optimal' cutoff of diagnostic devices in the same study. Joint Statistical Meeting. 2005. ASA Section on Statistics in Epidemiology, p.2547-2551.

20. Khan MJ, Castle PE, Lorincz AT, Wacholder S, Sherman M, Scott DR, Rush BB, Glass, AG, Schiffman M. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. Journal of the National Cancer Institute. 2005; 97(14):1072-79.

21. Kondratovich M. Comparing two medical tests when results of reference standard are unavailable for those negative via both tests. Journal of Biopharmaceutical Statistics. 2008; 18(1): 145-66.

22. Clinical and Laboratory Standards Institute. 2006 Molecular Diagnostic Methods for Infectious Disease; Proposed Guideline. MM3-A2. Clinical and Laboratory Standards Institute, Wayne PA.

23. Massad LS, Einstein MH, Huh WK, et al. 2012 Updated Consensus Guidelines for the Management of Abnormal Cervical Cancer Screening Tests and Cancer Precursors. Obstet Gynecol. 2013; 121(4): 829-46.

24. Saslow D, Solomon D, Lawson H, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. Am J Clin Pathol. 2012; 137(4): 516-542.

25. Darragh TM, Colgan TJ, Cox JT, et al. The Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions: Background and Consensus Recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. J Low Genit Tract Dis. 2012;16(3): 205-42.

DRAFT

IX. Appendix – Statistical Analysis

Calculating Score Confidence Intervals for Percentages and Proportions

The following are additional recommendations for performing statistical analyses of percentages or proportions. There are several different methods available. We suggest that either a score method described by Altman, et al. (Altman D.A., Machin D., Bryant T.N., Gardner M.J. eds. *Statistics with Confidence.* 2nd ed. British Medical Journal; 2000) or a Clopper-Pearson Method (Clopper CJ, Pearson E. The use of confidence or fiducial limits illustrated in the case of binomial. *Biometrika* 1934; 26:404-413) be used. The advantages with the score method are that it has better statistical properties and it can be calculated directly. Score confidence limits tend to yield narrower confidence intervals than Clopper-Pearson confidence intervals, resulting in a larger lower confidence limit. Thus when n=70 samples and 65/70=92.9%, the score lower limit of two-sided 95% confidence interval is 84.3%. In contrast, the Clopper-Pearson lower confidence limit is 84.1%. In this document, we have illustrated the reporting of confidence intervals using the score approach. For convenience, we have provided the formulas for the score confidence interval for a percentage.

A two-sided 95% score confidence interval for the proportion of A/B is calculated as: $\left[100\%\left(Q_1 - Q_2\right)/Q_3, 100\%\left(Q_1 + Q_2\right)/Q_3\right]$, where the quantities Q_1, Q_2, and Q_3 are computed from the data using the formulas below. For the proportion of A/B:

$$Q_1 = 2 \cdot A + 1.96^2 = 2 \cdot A + 3.84$$

$$Q_2 = 1.96\sqrt{1.96^2 + 4 \cdot A \cdot (B - A)/B} = 1.96\sqrt{3.84 + 4 \cdot A \cdot (B - A)/B}$$

$$Q_3 = 2 \cdot (B + 1.96^2) = 2 \cdot B + 7.68$$

In the formulas above, 1.96 is the quantile from the standard normal distribution that corresponds to 95% confidence.

For an example of proportion if (65/70), Q_1=133.84, Q_2=9.28, and Q_3=147.68, then the two-sided 95% score confidence interval is 84.3% to 96.9%

Calculation of Confidence Intervals for Positive Predictive Value (PPV) and Negative Predictive Value (NPV) based on Confidence Intervals for Likelihood Ratios (Prevalence is Constant)

PPV is $(1+PLR^{-1}*(1-\pi)/\pi)^{-1}$, where PLR is positive likelihood ratio (PLR=se/(1-sp)); NPV is $(1+NLR*\pi/(1-\pi))^{-1}$, where NLR is negative likelihood ratio (NLR=(1-se)/sp)) and π is prevalence. For the calculation of 95% confidence intervals for the likelihood ratios, use calculation of confidence intervals for the ratio of two independent proportions (the estimate of Se and the estimate of (1-Sp) for PLR and the estimate of (1-Se) and the estimate of Sp for NLR). There are several different methods available for calculation of the confidence intervals for the likelihood ratios (see Altman D.A., Machin D., Bryant T.N., Gardner M.J. eds. *Statistics with Confidence.* 2nd ed. British Medical Journal; 2000, pages 18-110). We suggest that a score

DRAFT

method described in paper by Nam (Nam J. Confidence limits for the ratio of two binomial proportions based on likelihood scores: non-iterative method. Biom J 1995; 37:375-9) be used. Using the 95% confidence interval for the corresponding likelihood ratio, it is easy to calculate the 95% CI for the corresponding predictive value where π (prevalence) is a constant.

Note:
Suppose that [L, U] is a 1-r level confidence interval for b and suppose that G is a function defined on the parameter space.
If G is increasing, then [G(L), G(U)] is 1-r level confidence interval for G(b).
If G is decreasing, then [G(U), G(L)] is 1-r level confidence interval for G(b).
(Functions $(1+x^{-1}*(1-\pi)/\pi)^{-1}$ and$(1+x*\pi/(1-\pi))^{-1}$ are monotonic functions when π is a constant.)

DRAFT