Guidance for Industry and Food and Drug Administration Staff

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

Document issued on: November 28, 2011
The draft of this document was issued on September 9. 2009.

For questions regarding this document contact Kate Simon, Ph.D., at (301)796-6204 or by email at kathleen.simon@fda.hhs.gov or Marina V. Kondratovich, Ph.D., at (301)796-6036 or by email at marina.kondratovich@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Office of In Vitro Diagnostic Device Evaluation and Safety
Division of Microbiology Devices

Preface

Public Comment

You may submit written comments and suggestions at any time for Agency consideration to the Division of Dockets Management, Food and Drug Administration, 5630 Fishers Lane, rm. 1061, (HFA-305), Rockville, MD, 20852. Submit electronic comments to http://www.regulations.gov. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*. Comments may not be acted upon by the Agency until the document is next revised or updated.

Additional Copies

Additional copies are available from the Internet. You may also send an e-mail request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance or send a fax request to 301-827-8149 to receive a hard copy. Please use the document number 1740 to identify the guidance you are requesting.

Table of Contents

Guidance for Industry and Food and Drug Administration Staff

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

> *This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.*

I.   Introduction

FDA is issuing this guidance to provide industry and agency staff with recommendations for studies to establish the performance characteristics of in vitro diagnostic devices (IVDs) intended for the detection, or detection and differentiation, of human papillomaviruses. These devices are used in conjunction with cervical cytology to aid in screening for cervical cancer. They include devices that detect a group of human papillomavirus (HPV) genotypes, particularly high risk human papillomaviruses, as well as devices that detect more than one genotype of HPV and further differentiate among them to indicate which genotype(s) of HPV is (are) 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 fifteen HPV genotypes are considered carcinogenic or "high risk" [Ref. 3 and 20]. "High risk HPV test" refers to HPV IVD devices that detect, but do not differentiate between different types of HPV,

while "HPV genotyping test" refers to HPV IVD devices that detect and further differentiate between different HPV types. In this document "HPV" means "high risk HPV," except where otherwise noted.

This guidance provides detailed information on the types of studies the Food and Drug Administration (FDA) recommends to support premarket applications (PMAs) for these devices. We recommend that you contact OIVD prior to beginning your studies to discuss specific study proposals and performance goals for your device.

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. It does not address HPV devices that are intended to be used independent of a cervical cytology result. This 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 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 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 general controls 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 the guidance entitled "In Vitro Diagnostic (IVD) Device Studies – Frequently Asked Questions" at http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/Guidanc eDocuments/ucm071230.pdf, and the guidance entitled "Guidance on Informed Consent

for *In Vitro* Diagnostic Device Studies Using Leftover Human Specimens that are Not Individually Identifiable" at 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 detection, or detection and differentiation, of human papillomaviruses. 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. It does not address HPV devices that are intended to be used independently of, or prior to, a cervical cytology result. It does not address HPV testing from non-cervical specimens, such as vaginal or penile 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, a single product code has been established for HPV DNA detection devices: MAQ, class III. The recommendations in this guidance apply to HPV diagnostic devices that detect HPV nucleic acid (not only HPV DNA, but HPV RNA, as well). 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 one listed.

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 could 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 the basis for determining the safety and effectiveness of these devices.

V.   Establishing Performance Characteristics

A. Analytical Studies

Limit of Detection
We recommend 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 and/or RNA transcripts 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 centrifuging cervical cells and removing LBC collection media prior to processing, you should perform your LoD studies in whatever matrix or buffer the cells are re-suspended after the centrifugation step. If you use LBC mock-samples containing HPV-infected cell lines in any of your analytical studies (as recommended under precision, below) then you should also perform LoD studies with these types of samples. A human epithelial cell line such as Jurkat 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).

We recommend that you first define a cutoff for the signal (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). 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.

We suggest that you refer to the Clinical and Laboratory Standards Institute (CLSI) document EP17-A [Ref. 4] for the basic concepts, design and statistical analysis of your LoD studies. You may use the approach described in CLSI EP17-A, and by Linnet and Kondratovich [Ref. 5] to estimate the LoD using the standard deviation of samples with very low concentrations. Alternatively, the LoD can be estimated using hit rates (percent of virus detected) as long as the determined hit rates cover a large part of the range of detection (0%-100%). The LoD study should include serial dilutions of each targeted HPV genotype, cell line or specimen type, and Probit analysis can be used for the statistical analysis [Ref. 6]. 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. 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.

Within-Laboratory Precision/Repeatability

We recommend you conduct the following within-laboratory precision/repeatability study in-house when minimal days of testing are desired at your external sites. Another option is to combine the within-laboratory precision study and the reproducibility study into a single study when this study is conducted over 12 days both internally and at your external testing sites.

Specimens for Precision 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). Sufficient volume is of concern for the following reasons: 1) HPV cannot be cultured, 2) access to clinical specimens is limited, and 3) full characterization of residual specimens can use up specimens available for testing. One option to address this concern is to use HPV-infected cell lines (mixed with a non-HPV infected "background" cell line such as Jurkat) to create panel members that mimic clinical specimens. 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 your clinical cutoff, plus at least one clinical sample negative for HPV in your precision study panels. Real clinical samples are needed 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 concentrations of viral levels. Viral load cannot be defined for clinical specimens, but you should challenge medical decision points [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, processed 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's Office of In Vitro Diagnostic Device Evaluation and Safety (OIVD) to discuss recommendations for additional characterization of such specimens.

We recommend that you conduct within-laboratory precision studies for devices that include 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. These test days should span at least two instrument calibration cycles (if applicable). You should assess precision between three different instruments and three reagent lots in your precision study.

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:

- 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): a sample 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: a 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).

Chart 1
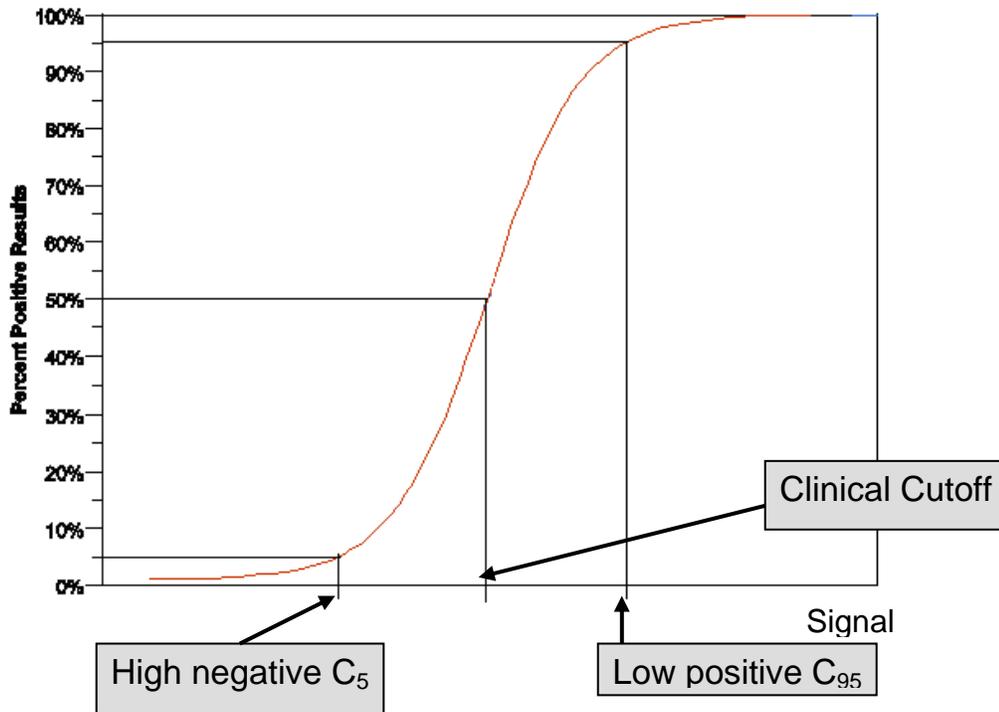
When the LoB is used as a clinical cutoff, then the concentration C_{95} is the same as the limit of detection (LoD) and the zero concentration (no analyte present in sample) is C_5 [Ref. 4]. CLSI documents EP05-A2 [Ref. 7] and EP12-A2 [Ref. 8] contain further information about designing and performing precision studies.

For within-laboratory 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.

Reproducibility

The protocol for the reproducibility study may vary slightly depending on the assay format. Note that if a larger 12-day study analogous to the in-house precision study above includes 2 external sites in addition to your in-house testing, then a separate reproducibility study is not necessary. When fewer days of testing are desired at your external sites, we recommend the following protocol:

- Evaluate the reproducibility of your test at three testing sites (for example, 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.
- 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 should be processed for each run starting from the nucleic acid extraction step with an independent extraction for each run.

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 document, EP05-A2 [Ref. 7] and EP15-A2 [Ref. 9], for additional information on reproducibility study design and statistical analysis.

Cross-Reactivity

We recommend that you test 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

---

[1] If the standard deviations (SD) in the precision studies for concentrations around the cutoff value 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 $C_5 = C_{50} / (1 + 1.645 \times CV)$.

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. Specific species are recommended according to prevalence and/or clinical relevance, 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 human papillomavirus (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*

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.

Potentially interfering substances include, but are not limited to, the following: whole blood (human), leukocytes, contraceptive and feminine hygiene products. The active ingredients and/or brand names of selected products and tested concentrations should be provided in your labeling. Examples of potentially interfering substances are listed 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 ("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 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 (1×10^6 cells/ml)
Contraceptive jelly
Douche
Anti-fungal cream
Spermicide
Vaginal lubricant
Feminine spray
Intravaginal hormones
Mucus

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 details, see Haeckel [Ref. 11]. For devices that are indicated for testing residual cytology samples, an analysis of the carryover effects of any upstream automated cytology processing system(s) should be provided.

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 the specimen types claimed in your intended use and analyte levels that challenge the medical decision point(s) of your assay.

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.

B. Clinical Performance Studies

Consideration of 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], which are the most current consensus guidelines available on cervical cancer screening.

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.

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 this type of device:

> 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 with ASC-US (atypical squamous cells of undetermined significance) cervical cytology results to determine the need for referral to colposcopy.
>
> 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.

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

General study design considerations common to ASC-US triage and adjunct intended uses (and likely any other intended uses):

Use of study sites outside the United States (21 CFR 814.15)
If you rely on foreign clinical data to support your PMA, FDA must be satisfied 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 be scientifically valid, your data should be applicable to the intended population and United States medical practice. We encourage you to meet with us in a presubmission meeting if you intend to seek approval based on foreign data, thus reducing the risk that the foreign study will not support your intended uses. Some areas of concern are 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, and ethnicity.

Histology review
FDA considers results of colposcopy and biopsy (if necessary) to be the clinical reference standard (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 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). We recommend that the panel establish the clinical reference standard (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.

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.

HPV genotypes
For an assay to detect "high risk" HPV, 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 high risk genotypes, you should explain why. Additional genotypes, such as those deemed "probably carcinogenic" or "possibly carcinogenic" by IARC (i.e. types 66, 68 and 73) may also be considered for inclusion, but this should be discussed with FDA prior to beginning your studies.

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.

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 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.

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 (i.e., 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 OIVD prior to beginning your studies. Note that biopsies taken from lesioned and non-lesioned areas should be denoted differently on your case report forms.

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 off-label processing of their cytology specimens.
Alternatively, sponsors who work from residual cytology samples will need to analytically assess the effects of carryover during cytology slide processing (see Carry-Over and Cross-Contamination Studies in the Analytical Study Section). 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.

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 may be reported in groups, instead of individually, where appropriate.

HPV vaccination and study populations

Due to the likelihood of an increase in the number of HPV vaccinated individuals in the United States over the coming years, the most clinically relevant HPV genotypes are expected to shift over time. For this reason, you should consider including in your analytical evaluations (precision, carryover, stability, etc.), one of the most clinically relevant non-vaccine targeted HPV genotypes (such as HPV 45 or 31) in addition to HPV genotypes targeted by any current FDA licensed vaccine(s).

When making sample size estimations, you should consider that increasing numbers of HPV-vaccinated individuals will decrease 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 would also need to be evaluated. Sponsors considering this type of design should discuss this option with FDA before beginning their studies.

Evaluation of HPV detection in 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 a composite HPV comparator that incorporates both FDA-approved HPV test(s) and/or PCR/Sequencing is also an option. For PCR/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).
- Sequence matches the reference or consensus sequence with an Expected Value (E-Value) < 10-30 for the specific target (for a BLAST search in GenBank, http://www.ncbi.nlm.nih.gov/Genbank/).

A comparison against PCR followed by sequencing of the amplicon is especially important for HPV genotyping assays to establish that the correct HPV genotype has been identified by your device. Note that any extra testing of discordant specimens is not necessary and CANNOT be utilized to alter estimates of agreement. However, the results of additional testing on discordant specimens can be footnoted in performance tables.

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: a) 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 below = "LoB<signal<clinical cutoff") or b) the HPV test did not detect the analyte of interest

("Not Detected" in table below = signal≤LoB). For the comparison of the HPV test and PCR/Sequencing (or other valid comparator, as discussed above), please present 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 and NILM (Negative for Intraepithelial Lesion or Malignancy) ≥30 populations separately in tables. In a request by industry, we were asked how they might provide this information. Table 3 is a format that you can use as an aide.

Table 3.

		PCR/Sequencing			Total
		High Risk Positive	High Risk Negative	Indeterminate	
HPV Positive					
HPV Negative	Detected				
	Not Detected				
Other (Invalid/Indeterminate)					
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 PCR/Sequencing vs. all outputs of the HPV test in a table separately for the ASC-US and NILM ≥30 populations. For details, please see Section 9 of CLSI MM17-A [Ref. 16].

In a request by industry, we were asked how they might provide this information for an HPV genotyping test with five possible outcomes: HPV16 Positive, HPV18 Positive, HPV16 & HPV18 Positive, Negative, and Invalid (Indeterminate): Table 4 is a format that you can use as an aide

Table 4

	PCR/Sequencing								
	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									

ASC-US Triage Intended Use – High Risk HPV Tests

You should conduct prospective clinical studies using specimens representing the intended use population, i.e. patients with ASC-US 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. 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)] 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 at a minimum of three study sites that are representative of clinical sites in the United States.

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 HPV triage indication, 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 intended use is to be pursued in a separate study, is to enroll only patients with ASC-US 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.

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, CIN1, CIN2, CIN3 and Cancer.

Clinical Performance Evaluation
The clinical performance of a test for the detection of HPV (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. The clinical performance of a test for the detection and differentiation of HPV genotypes (test with multiple outcomes) 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 .

In a request by industry, we were asked how they might provide this information. Table 5 is a format that you can use as an aide for a qualitative test with two outcomes (Positive and Negative):

Table 5

		Central Histology				
	Neg Colpo	Neg	CIN1	CIN2	$\geq$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" ($\geq$CIN2) should be evaluated as follows:
Sensitivity = $(A_4+A_5)/(A_4+A_5+B_4+B_5)$;
Specificity = $(B_1+B_2+B_3)/(A_1+A_2+A_3+B_1+B_2+B_3)$
PPV=$(A_4+A_5)/(A_1+A_2+A_3+A_4+A_5)$
NPV=$(B_1+B_2+B_3)/(B_1+B_2+B_3+B_4+B_5)$

Prevalence of $\geq$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" ($\geq$CIN3) should 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 $\geq$CIN3=$(A_5+B_5)/N$

The estimation of sensitivity and specificity, 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 Statistical Analysis Appendix 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 Statistical Analysis Appendix).

The clinical performance for the target condition $\geq$CIN2 should be stratified by age. For age groups <21, 21-30, 30-39, and 39+, present the prevalence of $\geq$CIN2, sensitivity, specificity, PPV and NPV along with 95% CI.

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 ($\geq$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 and subsequent negative predictive value(s) of your device fall short of current expectations for HPV testing [Ref. 13], panel review of your performance data may be necessary to allow assessment of the clinical effectiveness of your test.

Selection of clinical cutoff
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 (for example, 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, 19].

ASC-US Population – HPV Genotyping Tests:

A test for the detection and differentiation of HPV genotypes usually has multiple outcomes. For example: HPV16+, HPV18+, HPV16/18+, etc. The study principles described in the preceding section (ASC-US Triage Intended Use – High Risk HPV Tests) to establish clinical sensitivity and specificity for ≥CIN2 in women with ASC-US cytology apply to both dual outcome high-risk HPV tests (positive or negative for high risk HPV), and multiple outcome HPV genotyping tests. 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.

In a request by industry, we were asked how they might provide the data of the HPV genotyping clinical study in the ASC-US population. Table 6 is a format that you can use as an aide (Example shown has outcomes: HPV16+. HPV18+, HPV16/18+ etc.):

Table 6

	Neg Colpo	Central Histology				
		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 for the test outcome X summarizes how many times more (or less) likely subjects with the disease (≥CIN2) are to have that particular result X than subjects without the disease: $LR(T=X) = Pr(T=X|D+)/Pr(T=X|D-)$.

- The likelihood ratios for each outcome should be calculated along with 95% confidence intervals.
- In addition to the likelihood ratios, probabilities that a patient has ≥CIN2 for each outcome of the test should be calculated along with 95% confidence intervals. The probabilities for each $Outcome_K$ are evaluated as:
  $Probability(≥CIN2|Outcome_K)=(A_{K4}+A_{K5})/(A_{K1}+A_{K2}+A_{K3}+A_{K4}+A_{K5})$ for each K.
- Also present the percent of each outcome in the clinical data set as:
  $Probability(Outcome_K)= (A_{K1}+A_{K2}+A_{K3}+A_{K4}+A_{K5})/N$ for each K.

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.

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.

Adjunct Intended Use – High Risk HPV Tests

General study design options
Per the 2006 consensus guidelines, 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 undetected 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 two 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 entitled "Evaluation of HPV detection in 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 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[2] 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 vs. 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% CI. The data should demonstrate that this absolute risk is low enough

---

[2] 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 2006 consensus guidelines [Ref 13]. The bias created in this situation is unavoidable as patient health is paramount.

      to ensure your test can be safely used for adjunctive screening in women 30 and older. In addition, present 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" 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.

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 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 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 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 CDRH's Office of Surveillance and Biometrics has issued a post-approval studies guidance, "Guidance for Industry and FDA Staff - Procedures for Handling Post-Approval Studies

Imposed by PMA Order."[3] Sponsors with novel HPV targets should contact FDA to discuss their eligibility to complete their longitudinal evaluation post-market.

Adjunct Intended Use – HPV Genotyping Tests
The study options described in the preceding section (Adjunct Intended Use – High Risk HPV Tests) to establish relative risk for ≥CIN2 in women 30 and over with normal cytology can be applied to both dual outcome high risk HPV tests (positive or negative for high risk HPV) and multiple outcome HPV genotyping tests (tests that not only detect, but differentiate between the different high-risk 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 2006 consensus guidelines, an additional option a company 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 OIVD for further assistance.

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.

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

---

[3]http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070974.htm

External Controls

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.

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 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."

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 human papillomavirus 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. The need for this control should be determined on a device case-by-case basis [Ref. 22].

VI. References

1. De Villiers EM. Taxonomic classification of papillomaviruses. Virology. 2004; 324(1):17-27.

2. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. Journal of Pathology. 1999; 189:12-19.

3. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ. Epidemiologic Classification of Human Papillomavirus Types Associated with Cervical Cancer. 2003. New England Journal of Medicine 348:518–527

4. Clinical and Laboratory Standards Institute. 2004. Protocol for Determination of Limits of Detection and Limits of Quantitation; Approved Guideline. EP17-A.

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

6. Weusten J.J.A.M (2008). A statistical approach to assess analytical sensitivity of diagnostic assays in the presence of false-positive results. Journal of Virological Methods (2008), doi:10.1016/j.jviromet.2008.05.013

7. Clinical and Laboratory Standards Institute. 2004. Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline. EP05-A2. 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. 2006. User Verification of Performance for Precision and Trueness; Approved Guideline. EP15-A2. 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; 63:302-306.

12. Clinical and Laboratory Standards Institute. 2009. Evaluation of Stability of In Vitro Diagnostic method products; 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. Available at: http://download.journals.elsevierhealth.com/pdfs/journals/0002-9378/PIIS0002937807009301.pdf

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; 287:2114-9.

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; 9:404-6.

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

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; 18:145-66.

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

VII. Statistical Analysis Appendix
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 provide 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: $[100\%(Q_1 - Q_2)/Q_3,\ 100\%(Q_1 + Q_2)/Q_3]$, 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 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.)