This guidance was written prior to the February 27, 1997 implementation of FDA's Good Guidance Practices, GGP's. It does not create or confer rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both. This guidance will be updated in the next revision to include the standard elements of GGP's.

POINTS TO CONSIDER FOR CERVICAL CYTOLOGY DEVICES

Introduction

This points-to-consider document supplements existing FDA guidance for premarket submission of in vitro diagnostic devices for approval or clearance. The information requested is intended to be comprehensive but may not be all inclusive. The emphasis is on FDA's particular concerns in the review of cytology devices for gynecologic specimens. These products include:

1. Collection devices for the initial cervicovaginal specimen [spatulas, brushes, etc.] (regulated under 21CFR 884.4530, class II)

2. Devices for making cell suspensions for thin layer or monolayer slides (unclassified)

3. Staining devices (regulated under 21CFR 864.3800, exempt)

4. Computer-assisted cell-locator devices (regulated under 21CFR 864.5260, class II)

5. Semiautomated and automated computer-assisted image analyzers (unclassified)

6. Any other devices used in preparation, reading and interpretation of gynecologic cytology specimens. (These may be regulated or unclassified.)

Designers of gynecologic cytology devices have study design and method validation challenges that are not found in other in vitro diagnostic devices. Some examples are:

* Over 50 million Pap tests are performed each year and about 2 million are diagnosed as abnormal. Small differences in performance of a device for Pap tests may result in major consequences in the detection rate of the tests.

* The loss of even a few cells prior to making a final slide may be critical.

* There are no animal models or sources of cellular materials to serve as positive or negative internal controls for each patient sample.

* Repeat cervical cytology samples may be unreliable until the cervical epithelium has a sufficient time interval to regenerate, usually at least 4 to 6 weeks.

* Cytologic diagnoses are dependent on multiple pre-analytic factors including sample collection and preparation even before the demanding microscopic examination.

* Cytologic diagnoses are based on qualitative criteria for image interpretation that may be interpreted differently among individual observers and among laboratories. Some degree of subjectivity is unavoidable.

FDA encourages sponsors to consult with the Division of Clinical Laboratory Devices (DCLD) for guidance and review of protocols prior to submission of a formal application, preferably prior to commencement of clinical studies. If the studies in the submission document the safety and effectiveness of the device, the FDA clearance/approval process can proceed in the most timely manner within the limits of the queue of submissions.

Minimal data required:

I. **Device Description**

Provide the following information:

A. Intended use

A description of the clinical intended use including the clinical disorders for which the device is used, the scientific basis for the disorder(s), the clinical significance of the procedure or test results, the risk-benefit issues for use of the device, and the clinical utility.

B. Principles of the procedure

A description of the principles of the procedure or test methodology including what the device does, how it is to be used, and who will use the device. Is the device a component of a system or used as a stand-alone device? Is the device for initial diagnosis of slides, confirmation of the diagnosis, screening, or monitoring? Define who will use this device; cytotechnologists, pathologists, clinicians, gynecologists, laypersons, or a combination of users.

C. Device components

A description of the components that are provided with the device. Supply instructions for acquiring any components not included with the device.

D. Manufacturing process.

Documentation that the device is manufactured following FDA Good Manufacturing Practices including product design, quality control, consistency of manufactured product with the original submission product, stability of end product, and any other applicable feature of the device.

II. Protocols

A. General considerations for protocols

Submit the entire protocol and study plan to FDA for comment.

Provide a written protocol complete for all phases of the study that is applicable to all study sites before specimen collection/selection begins. It is the responsibility of the sponsor to ensure adherence to the protocol at all study sites.

Consult with a biostatistician during the initial planning stages of the protocol and during data analysis. *The most common reason for premarket submissions to fail is poor planning of the protocol.* Once a study is completed, it is almost impossible to make a poorly planned study acceptable.

Ensure that the study has a defined hypothesis and that the study design has the power to test this hypothesis.

Once the study begins, justify and document any and all changes to the protocol. All changes must be written into the protocol. Inform FDA of proposed protocol changes and what effects these changes will have on data interpretation. Discuss these points with the biostatistician prior to implementation.

Collect data for each intended specimen type or target population.

Collect performance data using a final production model device and not a prototype.

Appoint a study coordinator to oversee all aspects of the protocol study. This person must have knowledge of all details of the study and serves to protect the integrity of the data.

Most study designs will compare a new cytology device to the conventional manually prepared and manual microscopy-read method. In some cases, the new cytology devices may yield test results that were not detected by the conventional PAP test. All abnormal diagnoses on slides that were previously diagnosed as negative must be reported to the laboratory. This includes ASCUS and above, AGCUS and above, etc. The sponsor has the responsibility of monitoring and assuring appropriate patient follow-up. A confirmatory test must be performed according to the Bethesda System (TBS) guidelines.

B. Pre-Clinical Studies

Define and test for the following performance characteristics:

Accuracy (lack of bias)

Comparison to a gold standard

Precision (lack of random error)

Reproducibility (inter-observer, -laboratory, -instrument, intra-observer,-instrument)

Percent agreement, adjusted

Morphologic criteria

Provide data to demonstrate the ability of the device to preserve morphological structure of the cells in comparison to the conventional smear; to capture adequate representation of endocervical components; and to indicate the presence of inflammation and infectious agents.

FDA recommends adherence to usage of The Bethesda System (TBS) criteria.

Stability

Provide real-time studies from three different manufacturing lots which include data to demonstrate the stability of any reagents and\or buffers under expected shipping, handling, and storage conditions including various extremes of temperature, humidity, and light. This includes pre-collection buffer solutions and post-collection preservation of cellular morphology.

C. Clinical Studies

1. Study Design

Provide the following study design information:

a. Patient inclusion and exclusion criteria.

Account for all patients and provide explanation for all patients who are excluded from study.

b. Specific patient selection sampling plan.

For methods not limited to the conventional Pap smear:

* Prospective patient sampling is preferred so that within a defined time frame all patients are consecutively entered into the study. If not, provide the detailed statistical sampling plan used. Frequently sponsors fail to use <u>all</u> consecutive patients during the prespecified study time period, therefore a clear description of the sampling method must be provided. The method of sampling should be selected to ensure the representativeness, completeness, and generalizability of the sample's data to the target population.

For methods limited to the conventional Pap smear:

* Retrospective studies may be permitted for methods that are solely based on the conventional Pap smear, e.g., computer-assisted Pap test reading versus manual Pap test reading. However, these will be considered a feasibility study only.

  If negative archival slides are used, there must be an IRB-approved written protocol detailing what will be done with the study information from the new device. If the new method yields abnormal diagnoses (including ASCUS and above, AGCUS and above, etc.) on slides that were previously diagnosed as negative, the laboratory that diagnosed the conventional PAP slide <u>must</u> be notified. The sponsor has the responsibility of

monitoring and assuring appropriate patient follow-up. A confirmatory test must be performed according to TBS guidelines.

For all devices:

* Select the endpoint of the clinical studies that will support the clinical intended use.

  Novel and unproven devices may require the most stringent data with endpoints that may include clinical outcomes, i.e., effect of the device on morbidity and/or mortality, colposcopy results, and directed biopsy results as clinically indicated. Surrogate endpoints include colposcopy with directed biopsy, conventional Pap test versus thin-layer or monolayer Pap test, manually read and interpreted Pap test vs. computer-assisted or automated Pap test.

* Define study period from which the patients will be selected.

* Provide the type and number of study centers.

  Carefully select clinical study sites to represent the spectrum of demographic features appropriate for the study of cervical pathology and include low prevalence and high prevalence populations for the diseases for which the device is to be used.

* Provide the number of patients per center.

  Study adequate numbers of patients with confirmed positive and confirmed negative results from each site to allow for site stand-alone analysis and to support the difference detected in clinical sensitivity and clinical specificity at the pre-specified statistical significance level and power.

* Justify statistically all pooling of data from multi-center studies. Collect sufficiently large numbers of specimens

from each study site to allow for stand-alone analysis at each site.

c. Hypothesis testing

Null and alternative hypotheses for 3 by 3 categorical classification of ordinal data.

The Bethesda System (TBS) cervical cytology diagnoses represent qualitative or categorical data, e.g., negative, atypical squamous cells of undetermined significance (ASCUS), low grade squamous intraepithelial lesion (LGSIL), high grade squamous intraepithelial lesion (HGSIL), etc. TBS diagnoses can be grouped into three treatment categories: negative, ASCUS, and LGSIL and above. These data are ordinally scaled and may be formed into two (new device and reference) stochastically ordered distributions.

Clearly define the null and alternative hypotheses to be tested for the 3 by 3 ordinal classification tables as shown above. The null hypothesis should be reasonably and consistently constructed from the data analyses of the pilot studies for the reference and the new device. The alternative hypothesis is the hypothesis designed to include all possibilities not included in the null hypothesis.

For example:

Null hypothesis = The two marginal distributions of clinical outcomes (negative, ASCUS, LGSIL and higher) are <u>equal</u> between the new and reference devices.

Alternative hypothesis = The two marginal distributions of clinical outcomes are <u>not</u> <u>equal</u> between the new and the reference devices (two-sided). This hypothesis considers the possibility that the device can be either better or worse than the reference device.

Perform appropriate statistical significance testing for the 3 by 3 ordinal classification data.

If the null hypothesis is rejected in the above test for the 3 by 3 classification data, perform statistical significance tests for two 2 by 2 classification tables by combining ASCUS into the negative or positive (LGSIL and above) groups.

For each of the 2 by 2 tables, if the patient true disease status is known, calculate the clinical sensitivity values for the disease positive group and the clinical specificity values for the disease negative group for both the reference and new device. If the patient true disease status is not known, calculate the proportions of disease (LGSIL and higher) for the reference and new device respectively.

When the patient's true disease status is known, perform statistical significance testing for comparing two true clinical sensitivities and two true clinical specificities (or for two true proportions of disease if patient disease is not known) between the reference and new device. The appropriate procedure for the multiple comparison problem needs to be considered since each of the above 2 by 2 tables were reconstructed from the SAME 3 by 3 table when the 3 by 3 ordinal data leads to the rejection of the null hypothesis.

d. Confidence Intervals

For the above 2 x 2 tables provide the 95 percent confidence intervals for true clinical sensitivity based on the disease-positive confirmed group and clinical specificity based on the disease-free confirmed group for the new device if the patient disease status is known.

Also, provide the 95 percent confidence intervals for the differences of the two true clinical sensitivities and two true clinical specificities between the new and reference devices. This calculation can be useful in making clinical decisions.

When the patient's true disease state is not known, provide the true proportions of disease states, LGSIL or higher, along with the 95 percent confidence intervals for the difference between the two true proportions.

e. Justification of Sample Size

The following must be clinically and statistically pre-specified for estimating the required sample size:

* Type I error (probability of rejecting a true null hypothesis)

* Power (probability of rejecting a false null hypothesis)

* Hypothetical clinical sensitivities and specificities (true disease status is known)

* Hypothetical proportions of LGSIL, HGSIL, and carcinomas (true disease status is not known)

* Disease prevalence

Collect a sufficient number of disease positive and negative samples to test the difference to be detected in clinical performance characteristics between the new device and the standard reference device at the prespecified statistical significance level and power.

f. Specify Statistical Analysis for ORDINAL Data for the following parameters:

Clinical Sensitivity

Clinical Specificity

Positive and Negative Predictive Values

Inter- and Intra-observer agreement (masked)

g. The study population must simulate the intended population for the device.

* Sample and perform testing at 3-5 different geographical sites. Selection of U.S. sites is strongly encouraged. Only one testing site may be the manufacturer's own test

site. Simulate the testing and the use of the device for the site(s) for which the device is intended.

* Collect specimens from populations with a range of low and high prevalences for the disease(s) to be detected by the device according to its intended use.

* Sample examples of all the cervical diseases and conditions covered by the Bethesda System (TBS). Clearly list in the intended use statement any *exclusions* of diagnostic categories, e.g., "this device is not intended to process, detect, interpret inflammatory cells, etc." or "no examples of X condition were detected or analyzed by this device during the premarketing studies. The performance of this device is not known for X condition." (X is a rare condition not readily detected in prospective clinical studies of a size with sufficient power for most conditions.)

Design preclinical tests to support, at least theoretically, that the device is safe and effective for all intended conditions claimed, even the most rare ones.

Design post-market studies to validate the safety and effectiveness of the device for rare conditions using sufficient numbers of clinical samples.

* Plan for any other sampling requirement for other conditions claimed in the intended use statement, e.g., processing of the samples for viral culture, nucleotide studies, etc.

2. End-point: Clinical accuracy.

A validation of an in vitro diagnostic method requires an independent reference endpoint as a measure of accuracy (truth). Also, once a reference method is chosen, <u>all</u> results must be compared with this method.

Measure how well the device represents the true clinical condition of the patient from whom the sample was taken. Collect and analyze all the results from the screening cytotechnologist for all specimens and

separately analyze the results for all the specimens referred to the pathologist. Ensure that the protocol's criteria for referral from the cytotechnologist to the pathologist are followed for *all* cases, both the new device and the reference method. Mask all comparisons of the new device to the reference method.

The following endpoints are listed in descending order of their power as scientific evidence for validation of a new device, beginning with the gold standard method:

a. Clinical outcome

This is the gold standard reference for accuracy and clinical utility but has the disadvantage that long term follow-up intervals are required to evaluate gynecology cytology.

Longer clinical studies may be required for devices that are not based on conventional Pap smears because it may not be possible to use archival cases for longitudinal studies.

b. Cervical biopsy

This method is less than a gold standard because of difficulty in biopsy sampling. Biopsies may yield false negative results from sampling error. Some of these patients may have a negative cytology but actually have a pre-neoplastic or neoplastic condition. Also, there are additional problems in that it is not practical to biopsy a large number of patients with negative or ASCUS or AGCUS cytology.

c. Colposcopy examination and directed cytology and/or biopsy sampling

If a colposcopy cervical cytology sample is to be used for confirmation, ensure that a sufficient time has elapsed from any previous cytology sampling to allow for cervical epithelium to regenerate (4 to 6 weeks).

3. Surrogate end-point: Specimen accuracy.

A conventional Pap smear with refereed manual microscopic

reading may be used as a data subset to resolve discrepancies and to assess specimen accuracy.

Masked manual microscopy is an acceptable surrogate reference method (a less-than-gold standard) for devices using a conventional Pap smear and as a comparison method for thin-layer or monolayer slide preparations. Record the masked refereed diagnoses before performing any unmasked comparison or "consensus" readings of slides or images.

Mask each cytotechnologist and pathologist to the results of the comparison method and each other's results for the conventional Pap smear and for the new device, e.g., an automated or semi-automated image analysis or computer-assisted cell locator device.

Document that the masking code will not be broken until all the data is gathered and recorded. It is important to maintain the masking code until the <u>final</u> data analysis.

3. Sample collection

a. Collection device for cervical specimens.

Limit cervical sampling collection devices to those that are FDA cleared. The device must sample the endocervix and the exocervix. Data should be collected to ensure TBS criteria for adequate cellularity, endocervical component, etc. The Ayres spatula used alone is not sufficient for endocervical sampling. The Brush used alone is not sufficient for exocervical sampling. A combination of collection devices or a combination device, e.g., cervical broom provide sufficient sampling of the endocervix and exocervix.

> One disadvantage is that some brushes may cause excessive bleeding and may not be used during pregnancy.

b. Devices that are not based on collection, preparation, and observation of conventional Pap smear slides, e.g., thin or monolayer slides.

An Institutional Review Board (IRB) approval and informed consent is required if a patient is to be screened by a method that does not allow for a conventional Pap test (conventional Pap smear read manually by a cytotechnologist with potential for referral to a pathologist).

Specimen collection may be handled in one of the following two ways:

If a Pap smear is to be made, collect the cervical sample and prepare a conventional Pap smear in the usual fashion before performing any of the steps for the non-Pap smear slide. The sample for the new methodology, e.g., thin layer or monolayer suspension may be made from the residual material on the collection spatula, brush or broom.

If no conventional Pap smear is made before sampling for the thin layer or monolayer preparation, randomize the study subjects into two groups: One group will have a conventional Pap smear and the other group will be sampled for the non-conventional method, e.g., thin layer or monolayer slides. Appropriate statistical design is needed to ensure equal or nearly equal numbers of samples in each of the two groups. This method may require an Investigational Device Exemption (IDE) and signed patient consent forms as well as an IRB approval if the study is a PMA.

The end-point for these studies is a comparison of the proportion of TBS diagnoses found in each of the two study groups.

See attached addendum for additional sampling considerations.

4. Sample processing

Consideration must be given to the method for processing the cervical sample. The information and data submitted for a smear would vary from that of a suspension or a differential separation of the cervical sample. The differential separation of cervical cells from mucus, red blood cells, leukocytes, and acellular material would require proof that diagnostic and contextual cells were not lost in the processing of the sample. The method of separation, filtration or centrifugation, must be addressed as well as the type of fixation.

5. Slide preparation

Issues involving the transfer of the cervical specimen to the slide will vary according to the method used. The transfer of cells from a suspension by filtration, centrifugation, or sedimentation is vastly different from a direct smear.

6. Cell finding-locating device

Issues to be addressed for this type of device include the theory of operation, the sensitivity and specificity of the image processing for the selection of cells. Does the device provide for marking of suspicious cells? What is the operator/instrument interaction? What type of long-term record is provided? Documentation for software that was designed as part of the cell-locating device should be provided.

7. Image interpretation

For devices that provide an image interpretation, issues to be addressed include the theory of operation, the comparison with conventional microscopy, operator\instrument interaction, long-term record keeping, and a hazard analysis and software documentation conforming to the criteria outlined in *Reviewer Guidance For Computer Controlled Medical Devices Undergoing 510(k) Review.*

III. Data

A. Record keeping

Stratify data and results by TBS diagnoses for claimed intended use by screening cytotechnologist and pathologist. The following terminology should be used: high grade squamous intraepithelial lesion (HGSIL), low grade squamous intraepithelial lesion (LGSIL), atypical squamous cells of undetermined significance (ASCUS), atypical glandular cellular of undetermined significance (AGCUS), etc. The definition of ASCUS should be confined to the TBS category. The descriptive diagnoses pertaining to presence of inflammation and infection as well as reactive changes and specimen adequacy should also be noted.

B. Data integrity

1. Retain all original work sheets and make them available for planned or

unannounced FDA inspections. In cases where the original form does not remain with the study, all transcribed data should show the initials of the recorder and the date of transcription. If original work sheets may not be available at a later date, it is important to provide a verification for the transcribed data which should include the initials of the individual verifying the transcription and the date of verification.

Document in writing any correction or modification of original work sheets. Corrections or modifications should follow standard good laboratory practices in that a single line is drawn through the notation that is changed. The correction is written above the line. Record the date and the initials of individual making any changes.

Ensure the integrity of data on computerized work sheets (databases) so that there is permanent recording of date of input of all data and all revisions.

Mask reading, interpretation, and recording of data. Mask all data analysis and interpretation.

2. At the intended site of use, conventional manually prepared Pap slides and slides prepared with a new device or read by a new device, e.g., computer-assisted reader or cell locator, may result in screening diagnoses that fall between TBS categories. It is most likely that these slides will be treated like the diagnosis of the higher rather than the lower TBS diagnosis bracketing the provisional diagnosis. The slide would be referred by the cytotechnologist to the next higher trained observer.

If the pathologist's diagnosis falls equivocally between two TBS diagnostic categories, it would be expected that the pathologist would refer the slide to another pathologist or would note the equivocal diagnosis in the report to the clinician for appropriate follow-up.

Record all diagnoses on the work sheet that are not definite TBS diagnoses, e.g., hedging diagnoses such as tentative, provisional, presumptive, rule out, R/O, or question mark ?, as the next higher grade. Establish a written algorithm in the protocol for all possible diagnostic variations before the study begins.

Ensure that data entry personnel are aware of TBS diagnostic criteria. Any equivocal or hedging diagnoses should be referred to the next level

reviewer in the cytology protocol, i.e., all slides that are not unequivocally negative.

3. If the device is computer-assisted, verify abnormal cells by manual microscopy to check and re-check cells identified as abnormal.

   If it is not possible to check abnormal cells by manual microscopy, justify the safety and effectiveness of the new methodology.

4. If the device displays abnormal cells on a computer monitor, document how these images are stored for additional observers to re-analyze.

5. Provide instructions in the study protocol to ensure adherence to the manufacturer's protocol by the cytotechnologists and pathologists.

6. Devices that allow for little or no opportunity for human intervention, in particular a software-controlled "black box-type" devices, may require software validation studies to document that the device can be expected to maintain reasonable performance for its intended purpose with the full range of expected specimens.

IV. Training Requirements

State the kind and amount of training requirements for any step and interpretation that differs from the conventional Pap test (manually prepared smear, manually processed, and read and interpreted without computer assistance).

Recommend the appropriate procedures for the laboratorian and laboratory to convert from conventional Pap test methodology to that of the new device.

Some training issues to consider are:

Will the new device affect the appearance of the finished slide?

Will the slide be a conventional Pap smear or a thin layer or monolayer slide?

Will the new slide methodology contain all of the components of the conventional slide?

If not, what is the possibility of misdiagnoses from loss of that (those) components?

How will these issues concern the clinician during sample collection, sample preparation, slide preparation, and final interpretation of result?

How will these issues affect the cytotechnologist during sample processing, slide preparation, identification of abnormal cells, and interpretation of cells from the preparation?

Consider how these issues affect the pathologist in identification of abnormal cells and interpretation of abnormal cells?

V. Workload Limit

All gynecologic cytology devices submitted for FDA approval or clearance must have an evaluation of their workload limit. This applies to devices for making cell suspensions for thin layer slides, computer-assisted cell locator devices, semi-automated and automated computer-assisted image analyzers, and any other device used in the preparation, reading, and interpretation of gynecologic cytology specimens. Data must be provided to assess and evaluate the fatigue factor in order to establish appropriate workload limits.

VI. References

National Cancer Institute Workshop. The 1988 Bethesda System for reporting cervical/vaginal cytological diagnosis. *JAMA*. 1989;262:931-934.

Broder S. Rapid communication--the Bethesda system for reporting cervical/vaginal cytologic diagnosis--report of the 1991 Bethesda Workshop. *JAMA*. 1992;267:1892.

Kurman RJ, Henson DE, Herbst AL, et al. Interim guidelines for Management of Abnormal Cervical Cytology. *JAMA*. 1994;271:1866-1869.

Kurman RJ, Solomon D. *The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Definitions, Criteria, and Explanatory Notes for Terminology and Specimen Adequacy*. New York, NY:Springer-Verlag NY Inc;1994.

Center For Devices and Radiological Health, Food and Drug Administration. *Reviewer Guidance For Computer Controlled Medical Devices Undergoing 510(k) Review*. August 29, 1991. This document is available from the Division of Small Manufacturers Association (DSMA), 1-800-638-2041.