March 30, 2003

Document Management Branch (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, MD 20852
USA

Re: **Comments concerning Docket #: 03D-0007**

To Whom It May Concern:

This response to these proposed guidelines is from the perspective of a central laboratory with a significant experience in anatomic pathology testing in clinical trials both nationally and globally. We perceive that the Food and Drug Administration (FDA) desires to have a greater focus on the Quality Assurance (QA) component of the efficacy evaluation. Anatomic pathology diagnoses are often difficult to manage due the complex and significantly subjective nature of the analysis and classification. Textual descriptions, constantly evolving classification systems, the natural limitation of complex pattern recognition and opinion leaders with differing perceptions on diagnostic criteria all contribute to the inter-observer and intra-observer variability to a degree that it is a significant expectation. However, despite of this challenge, QA measures can be improved and the commitment to patient safety and quality can be maintained. The measures proposed in this draft set of guidelines proposes a variety of QA measures, many we feel are warranted. However, in our opinion some of these should be clarified or modified to allow even a better outcome in both QA and in the delivery of optimal medical services to the subjects during the trial. We respectfully submit our comments from a service provider's perspective to aid in the construction of guidelines that will continue to serve the subjects interests by conducting an efficient, medically robust and scientifically sound clinical trial.

Section with lines 229-242

C. Inclusion and Exclusion Criteria

We recommend that:
All subjects have a uterus and have an evaluable screening endometrial biopsy (i.e., endometrial tissue sufficient for diagnosis). Findings indicating endometrial hyperplasia or cancer would result in exclusion from enrollment and subjects would be referred for *standard of care* clinical management.

Comment 1:

In menopausal subjects with atrophy, a significant number (up to 30%, incidence increases with duration of the menopausal state) have a biopsy sample that consists of only limited endometrial surface epithelium without intact gland or stroma. Although this is an expectation, there is no reliable way to exclude that such specimens are non-representative. When correlated with a very low TVU thickness (less than 4-5mm) this finding is an "imaging correlated" expectation. The textbook recommended for classification although written from the perspective of biopsy findings in subjects with symptoms requiring evaluation, describes this pattern as consistent with atrophy but does not establish a low threshold guideline for adequacy in asymptomatic menopausal women. At our laboratory we arbitrarily

established a guideline that requires that there are 10 or more atrophic epithelial strips on multiple sections to be classified as "surface endometrium without intact glands or stroma" which falls under the category of atrophy. Although this quantity establishes a guide for consistency there are no outcomes data to confirm this threshold is appropriate. We continue to have concern that this does not address patient safety if any such case would be a non-representative when a serious pathology was present. We suggest that the FDA consider an approach to correlate these findings since the clinical trial structure typically blinds the readers to clinical circumstances, despite the expectation that these are asymptomatic subjects. We would feel comfortable with the guideline that a correlative low TVU (possibly < 6 mm) be used to allow the sponsor to classify such minimal tissue findings as atrophy.

239• A negative screening mammogram (obtained at screening or within 3 months of study enrollment) and normal clinical breast examination be documented prior to enrollment in clinical studies for women > 40 years old. Findings indicating any suspicion of breast malignancy would result in exclusion from enrollment.

No comment

Section with lines 243-292

D. Monitoring

We recommend that:

• The endometrial tissue obtained by endometrial biopsy at screening, during the conduct of the study, and at the end-of-study be processed in the same manner by a central laboratory.

• Endometrial biopsies and not uterine ultrasounds be used for the evaluation of endometrial hyperplasia (sponsors interested in establishing a correlation between transvaginal ultrasound and endometrial biopsy results may perform transvaginal ultrasound immediately preceding endometrial biopsies).

Comment 2:

This might be modified to encourage TVU correlation of limited atrophic tissue, such as by stating:

"(Sponsors interested in establishing a correlation between transvaginal ultrasound and endometrial biopsy results ***should*** perform transvaginal ultrasound immediately preceding endometrial biopsies **and ultrasound finding consistent with atrophy can be utilized to effectively exclude a non-representative sample when tissue is scant and atrophic but devoid of intact glands)".**

• A single pathologist reader (any one of the three blinded pathologists) initially assess the slides from the endometrial biopsies obtained at screening or because of participant bleeding while on study drug (safety reading).

Comment 3:

Since the three readers are required to be from separate institutions and the samples are to be processed centrally, in reality only the central pathologist can be expected to be available for this initial evaluation and he or she also has to provide an efficacy report to the clinician investigator. (Routine laboratory regulations in the many states involved in such studies and the physical and financial separation of the readers effectively preclude all three from participation in this initial evaluation or from generating a patient care report other than via the central laboratory.) Many states, such as New York and Maryland require biopsies and/or pap smears to be only handled in inspected laboratories by tested technologists and pathologists. Other states, such as Illinois, require licensure to even issue a legal report (additional pathologists must be a "consultant" to the primary pathologist to be involved). It seems that the FDA and sponsor are typically unaware of these encumbrances that have significant operational and legal barriers

by the nature of the clinical trial scenario and requirements. We recommend that the FDA become aware of these limitations and realize that sponsors at times are unaware of the difference between the processes for the conduct of the trial to generate FDA analytical data and the medical-legal requirements and practice standards that separately guide how the clinical reports are issued to the clinical investigator. We feel the draft could be improved by making it clear that the reporting terminology and reporting mechanisms have the flexibility to allow a central laboratory to document and report data for the clinical trial while it also issues clinically detailed and timely reports for clinical care.

These studies typically span a timeframe of several months in the recruitment stages so that a single designated pathologist will at some time be absent for scheduled events in addition to a likelihood of unscheduled absences, such as illness. Even though these subjects are asymptomatic, experience shows that the combination of tight schedules to meet drug dispensing deadlines, needs for repeat testing at screening, various other investigator needs and "patient anxiety" all result in a consistent demand for reasonable turnaround time in resulting these samples. Without several pathologists being allowed to result these screening samples, the absence of this designated pathologist creates patient, investigator and laboratory disruptions and conflict. In addition there is the inevitable perception (and at times the reality) that this results in substandard clinical care due to prolonged turnaround times. These screening samples are almost always routine material, and when they represent significant pathology, we feel immediate blind review by the other two designated pathologists would be optimal for medical legal reasons. It seems unreasonable that several experienced pathologists cannot be trusted to result such routine screening samples when there is a system to assure that this is being done in a quality way (immediate blinded reviews). The reason we favor this immediate review is that since these will likely be included as a "control" with the efficacy evaluations, a therapeutic difference in diagnostic classification generated by this blinded arbitration (such as "atypical" hyperplasia from routine hyperplasia) would be required to be acted upon and would shock and dismay both clinicians and patients. The amended diagnosis of atypical hyperplasia would typically result in a hysterectomy. This could be easily avoided and still the cases could be used as "efficacy controls" with the advantage of providing more intra-reader reproducibility information for all readers.

• For the efficacy evaluation, three independent expert pathologists, blinded to treatment group and to each other's readings, determine the diagnosis for endometrial biopsy slides during the conduct of the study.

Comment 4:

Since the three readers are required to be "experts" it seems that this needs some definition. No matter what the criteria, it would seem prudent to have some latitude in the definition for specialty training and allowances for practical experience. In every field of pathology there are individuals working in subspecialty areas that might be considered expert by such a work title, but diagnostically they are indecisive or inconsistent. It has been our experience that it is difficult to draw any firm conclusion about a reproducibility or accuracy only by "fellowship experience", stated work experience or title without an independent measure of functioning in the task being anticipated. It is clearly reasonable to use these elements in the process, but we need to be aware that the quality of the similar experience and certain innate "talents" of visual pattern recognition are primarily important, particularly when this mandates a "blinded" evaluation and the comparisons are between pathologists presumed to be of a somewhat heterogeneous training history. It seems reasonable since endometrium is a very common diagnostic specimen for even the average practicing pathologist that objective measures for these readers by blind reviews of similar material could be an avenue to validate the appropriate level of expertise. It is highly likely that a sponsor would not risk a project's success on one or more pathologists without being

confident of their abilities, but they are often unable to establish a standard and in general have a limited understanding of anatomic pathology. The topic in the following paragraph might aid in this regard.

We feel that if the FDA desires better QA of the diagnostic results, then one key to this process is using robust control (or calibration) cases. In fact we are surprised that this draft proposal does not mention the approach of enriching the efficacy pool with additional cases of hyperplasia. We have heard this requirement mentioned and it was reiterated in a recent conference call with the FDA staff concerning a specific study. We would suggest that the FDA define a method that would be considered as a way to both show appropriate diagnostic skills in classifying endometrial samples in addition to selecting hyperplasia cases for "seeding" the efficacy evaluations. We would then suggest that "experienced by documented work history" or " validated by successful performance on blind evaluation of a large validated sample set" be considered as valid as an "expert" when defined by fellowship training.

We would suggest that a validation might be provided by putting together a set of 200-300 endometrial biopsy cases with about 40-50% being likely to be diagnosed as somewhere in the range of hyperplasia to hyperplasia with atypia, about 10-15% being cancer and 10-15% being disordered proliferation and the remaining 20-30% being a mixture of diagnoses between insufficient to normal and some other pathologies (endometritis and polyps). The blinded readings of these cases can establish all hyperplasia cases used to re-cut as "controls" to insert blindly and randomly with the study's efficacy evaluations. In addition, the overall statistics for each reader would be available to allow the FDA to consider whether an individual reader might be consider as appropriate to serve in a study. This need not be viewed as a mandated process, but a scientifically valid option for those interested in formally using variance studies in anatomic pathology QA in these clinical trials efforts.

• Curricula vitae for participating pathologists be provided to the FDA and document expertise in gynecologic pathology.

Comment 5:

See the comment above. To be truly fair, documenting expertise in this area would seem to include any pathologist with a quality clinical experience maintained with appropriate postgraduate education efforts. Thus we would suggest that "expertise in gynecologic pathology" be replaced by a statement such as "an appropriate experience in gynecologic pathology (In some cases supplemental documentation indicating validation of the pathologist for this role may be indicated)".

• Participating study pathologists be from different institutions with independent fiduciary and organizational reporting, and these pathologists not meet to review slides before or during the conduct of the clinical trial.

Comment 6:

Although we can understand the logic in not having active slide reviews during the trials, this restriction before the trials assumes that experienced pathologists are utilizing criteria in a standard fashion. We expect that the FDA is aware that in clinical practice simple hyperplasia is commonly over-diagnosed, so there should be concern that this diagnostic threshold is being consistently applied. Since validation processes should precede a process that is subsequently followed by QA measures, we would suggest that a "diagnostic bias" in applying the criteria for the lower threshold of hyperplasia as illustrated in the guidance textbook should be the goal. The anticipated "reader bias" that comes from consensus reviews at a multi-view microscopic session are effectively eliminated by "blinded" evaluations. We also understand that even subtle processes, such as marking slides can produce an unintended bias, so these

practices should be clearly defined as violating the blinding process. However, it is clear that even if readers follow a reference text, the illustrations and text cannot definitively address every pattern or situation. In addition, many "experts" hold post-graduate seminars for others, and both might wish to participate in a trial. It seems that better definition and/or some relaxation in the statement "these pathologists not meet to review slides before" ***so that it is clear what is being restricted. It is only logical that no study slides be reviewed, but it seems to us that QA data could be equally valid if it is possible to discuss any questions related to the reference text or to use the concept discussed above with "validation slide sets" for blind reviews in a way to better define the criteria used for the critical threshold of hyperplasia.***

Statistical studies have in many instances suggested in cytologic and histologic evaluations that experts show a poorer inter-reader variance than many would anticipate. In some studies, non-experts have at times performed as statistical equals in grading routine types of cases (ordinal data). In addition in such ordinal evaluations when multiple blind evaluations are averaged, the subsequent classification consistency of these repetitive evaluations were more "accurate" than a single evaluation. The approach proposed in the draft guidelines seems to adhere to those findings and offers the opportunity for the FDA to move to a more statistically valid efficacy measure than depending on the "tertiary expert" arbitration of the prior guidelines.

We feel a "pre-study" validation process will minimize any misunderstanding of analytical criteria, but will not eliminate the inherent variability. We further feel that the multiple blinded reads will serve their purpose best if there are adequate numbers of evaluations around this critical threshold that in such studies is typically an uncommon event. If encouraged (or at least allowed) prior to any study activity, such a validation could proactively establish expectations for inter- and intra-reader variance. It has been shown that such variance is greater in conditions where morphologic patterns overlap and that typically only the most aberrant conditions have high levels of reproducibility. Thus we anticipate that simple hyperplasia will have more variance than complex hyperplasia and carcinoma, similar to what is seen in Low grade SIL, High grade SIL and invasive carcinoma in cytology and histology. In this situation the introduction of a statistically significant number of "efficacy controls" at this critical threshold would seem to make any dataset more robust and in the final analysis a "pre-study validation" process to establish benchmark cases (agree upon by all readers) could both provide this "control" resource and define "expertise".

• Standardized criteria as provided in Blaustein's pathology text (Pathology of the Female Genital Tract) be used for the diagnosis of endometrial hyperplasia (see Appendix for recommended histologic characteristics of the endometrium).

Comment 7:

This is an excellent text and we routinely use it as a guide, however there are areas in this and all textbooks that do not address nearly as completely samples that are not fairly typical "benchmark" cases. In these circumstances a variety of circumstances arise that become issues in clinical trials. Also the key in these studies relates to the issues such as: sample adequacy criteria, reproducible criteria for the threshold of hyperplasia (focality/gland density and "atypia") and classification for unusual diagnostic findings. The limited photographs in all textbooks fall short of illustrating the many patterns of hyperplasia that present in blind outpatient samples that contrast with the "benchmark" abundant samples typically illustrated in textbooks. This is why is seems quite practical to have more controls in the casework and to formally define options for any "pre-study validation" of the readers to get

reproducibility data without "criteria shift" and to understand which pathologists show expertise in actual practice (see prior comment). During the efficacy evaluations of a study, the seeding of some of these cases into the case mix could serve to "calibrate" this threshold, since in reality almost all of these studies have very few cases of hyperplasia.

• Endometrial polyps be fully characterized as to the glandular proliferation and atypia (see Appendix for additional histologic characteristics of the specimen).
See comment below in Histologic Characteristics, comment 14.

• Subjects found to have endometrial hyperplasia or adenocarcinoma of the endometrium be excluded from further drug treatment (if discovered during study drug treatment period) and referred for *standard of care* clinical management and followed to complete resolution, and the report of any medical or surgical procedures and the resultant pathology be provided to the FDA.

Comment 8:

Since the three readers will subsequently see these cases (next guideline) it seems that some method for these cases to have at these the additional evaluations immediately will avoid any consternation should a diagnosis be altered later on in the multi-read process. It would be medico legally risky for the laboratory and sponsor and very disappointing for a patient to have to be told as long as 6-12 months later that a sample diagnosed as "hyperplasia without atypia" is now considered to harbor atypia, which would likely result in hysterectomy rather than chemical or endometrial curettage. Again, timeliness and finality of these results are both an expectation and arguably a right for the patient. This process would not preclude that these samples still be used as "control cases" the efficacy evaluations, and in fact would serve to allow reproducibility data on more that the central reader of the safety/screening samples.

• If hyperplasia is diagnosed by the single safety reader for a subject who has bled while on study drug, this diagnosis be maintained for the efficacy evaluation and the slides become part of the slide set given to the two other pathologists for reading.

Comment 9:

See the comment above. We would ask that it be made clear that these cases will be read by all three readers in the efficacy slide sets.

• For the efficacy evaluation, the concurrence of two of the three pathologists be accepted as the final diagnosis. If there is no agreement among the three pathologists, the most severe pathologic diagnosis (i.e., atypical hyperplasia > complex hyperplasia > simple hyperplasia > benign endometrium) would be used as the final diagnosis.

Comment 10:

It should be made clear that this is guideline is for the reporting for the dataset to the FDA for use in analysis. If taken at "face value", this suggests that reports might come directly from a reader external to the central laboratory to be issued to clinicians caring for subjects in a variety of states or even countries. In many situations this is not a legal option. The only legal option in a national/global trial is to have the licensed pathologist at the central laboratory issue all reports with the outside readers serving as "consultants", and even this is getting to be very difficult to maintain. New state laws are being added in some states (in response to "telemedicine") to restrict case analysis to pathologists with a state license for the same state as the subject. Other states already require special state inspections if any specimen comes from that state (e.g. New York, Maryland). This is going to create real functional issues on top of the already restrictive format in the draft guidelines.

• The slide set distributed to each of the three pathologists for the end-of-study pathology review incorporate control sides representing a randomly selected 10 percent of the screening normal slides and all slides from subjects excluded for the diagnosis of hyperplasia or cancer to insure quality control.

Comment 11:

Since the three readers are stated to not have any opportunity review slides before the study, there seems to be a potential for inequality in the use of categorization criteria. Quality is generally defined in terms of reproducibility and accuracy. To fairly measure consistency would seem require some opportunity to fully understand definitions for categorizations and some expectation of what is an acceptable range of variation. Without a pre-study agreement on benchmark examples of the hyperplasia threshold, there would seem to be a lack of definition for this anticipated range. Statistically it seems that the rate of hyperplasia is low enough that statistical significance may not even be reached without increasing the number of threshold cases (only 2-3% of many cases subject will likely show hyperplasia on biopsy and that is only 10-20 cases in a study of 1000-2000 patients). It will unlikely validate the threshold of hyperplasia with atypia, which is the diagnosis that is presently the more accepted risk for carcinoma.

The FDA should be aware of the operational difficulty in "seeding" cases from this screening pool or from outside sources. To truly blind the pathologists, these cases have to be relabeled (altering "source" documents) including a total patient identity in central laboratories that use more complex tracking and data management systems.

• Digital recording of diagnostic areas of the slides be maintained by the central laboratory and be made available upon FDA request.

Comment 12:

This requirement poses many problems:

a. **What areas are considered diagnostic is a matter of opinion and these disrupted outpatient samples are very difficult to illustrate without multiple images (various areas and various powers of magnification). These guidelines seem to suggest that only the central reader would be required to perform this function and that all of the cases would be imaged.**
b. **The ability to "prove" and image is from a certain case is impossible without reviewing the original slides imaged and the vast majority of cases will have extremely routine findings. These images of the cases would by their nature have the "bias" of the reader and the original slides would be the source data, which have much more "granularity" of information and are "unbiased" if this imaging is intended as tool to arbitrate or investigate questionable data.**
c. **The only reason these images could be of any value in this process would be validate and this is a process that would really require the review of the source data, which is the original glass slides. These slide have to be stored for at least 15 years and since all cases are processed at the central laboratory, would be in one location. The additional burden of imaging seems to be of no known value and is without any guidelines.**

• Any new findings noted during the conduct of the study and on end-of-study physical examination (including findings related to the breast) receive careful and appropriate evaluation and be monitored until there is complete clinical resolution of any diagnosed condition.

No comment

E. Primary Endpoints 297-300
For protection of the endometrium, we recommend the evaluation of the incidence rate of endometrial hyperplasia at 12 months.
No comment

F. Study Analysis 302
See Section III.F. For analysis of primary endpoints for treatment of moderate or severe vasomotor symptoms or moderate to severe symptoms of vulvar and vaginal atrophy associated with the menopause. The objective of the clinical trial is to demonstrate the lowest effective dose of the progestin drug that reduces the estimated risk of endometrial hyperplasia after 1 year of estrogen/progestin treatment. The reported 1-year background incidence rate for endometrial hyperplasia in postmenopausal women and in postmenopausal women treated with currently marketed combination estrogen/progestin drugs is approximately 0-1 percent. We recommend that the results from the clinical trial demonstrate a hyperplasia rate that is = 1 percent with an upper bound of the one-sided 95 percent confidence interval for that rate that does not exceed 4 percent. The frequency of atypical hyperplasia and cancer are important additional factors to be considered in determining approvability of the drug product. The incidence of hyperplastic polyps and associated atypia would be considered in the safety review.

Comment 13:
These new operational guidelines seem to be focused on more quality assurance of this threshold of hyperplasia. It is common knowledge by the most of pathologists that the lower threshold of this diagnosis is poorly maintained in many practices, including medical centers. We have studied both experiences and so called experts in our work with thousands of samples in truly blinded evaluations in addition to retrospective evaluations and it seems that simple hyperplasia is over-diagnosed in about 30-60% of cases in almost all of these environments. It is expected in small Pipelle (TM) samples to have more cases the disruption or scant volume that would result in an increase is "over-calls", but this is overshadowed by a much greater tendency for the diagnosis of "simple hyperplasia without atypia" than is warranted on retrospective review.

Section with lines 317-395
APPENDIX: HISTOLOGIC DESCRIPTIONS RECOMMENDED FOR USE WHEN READING ENDOMETRIAL BIOPSY SLIDES
Histologic Characteristics of the Endometrium
0. No tissue
1. Tissue insufficient for diagnosis
2. Atrophic
3. Inactive
4. Proliferative
   a. Weakly proliferative
   b. Active proliferative
   c. Disordered proliferative
5. Secretory
   a. Cyclic type
   b Progestational type (including stromal decidualization)
6. Menstrual type
7. Simple hyperplasia without atypia
8. Simple hyperplasia with atypia

9. Complex hyperplasia without atypia
10. Complex hyperplasia with atypia
11. Carcinoma (specify type)

Additional Histologic Characteristics
If there are any polyps, please specify the type or types.
Functional
Atrophic
Hyperplastic without atypia
Hyperplastic with atypia
Carcinomatous

Comment 14:

It is assumed that the list above for reporting data to the FDA for study analysis and it is not meant to be entirely comprehensive (inflammatory conditions absent and some classifications above are treated somewhat differently in the reference text). Additionally, we have seen many polyps suggesting origin in the lower uterine segment with a surface of endocervical type mucinous epithelium and a mixture of "low density" inactive to functional endometrial glands (we call these "mixed" type and indicate the epithelial patterns present). Also we expect the FDA is sensitive to the fact that the reporting of findings to the clinical investigator will and should have the flexibility to be as inclusive and fully diagnostic as needed to meet the "standard of care" for routine practice. In that regard, the more general categories stated in the classification above do not have to be strictly applied in these investigator reports as long as the semantics used are consistently and directly "mapped or linked" to the above categories for statistical analysis by the FDA (we use relational databases to allow multiple statistical categorizations to textual diagnostic statements).

If there is any stromal tissue, please specify the type or types.
Smooth muscle tissue, normal
Features suggestive of adenomyoma
Features suggestive of stromal nodule
Sarcoma (specify type)

Comment 15:

None other than for comment 14.

If there is any metaplasia, please specify the type or types.
Squamous
Papillary
Eosinophilic
Ciliated
Mucinous
Syncytial
Other type (specify type)

Comment 16:

Metaplasia is very common, but typically affects less than 20% of the glands. We would suggest that if metaplasia is felt to be important to monitor, that a threshold for reporting and any stratification into degree of gland involvement be defined.

If there is any cervical tissue, please specify the type or types.
Fragments of negative cervical epithelium
Endocervical polyp
Atypical endocervical glandular epithelium
Atypical squamous metaplasia
Squamous dysplasia
Cervical carcinoma
No comment

We hope these observations will help shed light on the many complexities that the central laboratory will face with these proposed guidelines and help the FDA shape a truly functional document that will accomplish the degree of integrity and standardization required to reach reliable and responsible conclusions. We would be happy to clarify any comment in this document and hope that the FDA receives these comments understanding that we desire to facilitate a scientifically valid approach that does not preclude providing the appropriate standard of care and respecting the full spectrum of legal requirements that confront a central laboratory managing such diverse studies.

Sincerely,

Michael D. Glant, MD
Chief Medical Officer
DCL Medical Laboratories, Inc.
(Teaming partner with Covance Central Services, Inc.)
9550 Zionsville Road, Suite 200
Indianapolis, IN 46268
(800) 837-3254
mick@dcla.com