STATISTICAL ANALYSIS PLAN

Study Title:	Clinical Validation Protocol for Feature Detection Precision by WSI using Hamamatsu NanoZoomer S360 MD Digital Slide Scanner System
Study Number:	HCT-P002
Protocol Version and Date:	5.0, 11 May 2020
Sponsor:	Hamamatsu Photonics K.K.
Plan Prepared by:	Ping-Yu Liu Liu Associates Consulting, LLC
Version and Date:	Version 1.0: Date (08 OCT 2021)
Clinicaltrials.gov	NCT#: 03992768

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
CAP	College of American Pathologists
CI	Confidence Interval
CRF	Case Report Form
FDA	Food and Drug Administration
FFPE	Formalin-Fixed Paraffin Embedded
FOV	Field of View
NPA	Percent Negative Agreement
PIPS	Philips IntelliSite Pathology Solution
PPA	Percent Positive Agreement
RP	Reader Pathologist
SAP	Statistical Analysis Plan
TIFF	Tag Image File Format
US	United States
WSI	Whole Slide Imaging

1. BACKGROUND AND RATIONALE

Whole Slide Imaging (WSI) has emerged as an alternative way to view pathology slides instead of the conventional method of viewing slides of tissue under a traditional light microscope used for primary diagnosis. There is growing adoption of the use of WSI for primary diagnosis in many countries. WSI is particularly advantageous for remote consultation where pathologists are not available over vast geographical regions, and for easier archiving of images of slide material that may have longer shelf life in digital form. In the United States (US), WSI is increasingly used for teaching, archiving, consultation, and research. Furthermore, the College of American Pathologists (CAP) has published recommendations to pathologists who wish to validate WSI in their clinical practices (1)

Quite recently, in a *de novo* authorization letter (2) and device summary (3), the Food and Drug Administration (FDA) has announced the authorization of the Philips IntelliSite Pathology Solution (PIPS) for use of WSI for primary diagnosis, specifically permitting WSI for *in-vitro* diagnostic use as an aid to the pathologist to review and interpret digital images of surgical pathology slides prepared from formalin-fixed paraffin embedded (FFPE) tissue, but not intended for use with frozen section, cytology, or non-FFPE hematopathology specimens.

Thus, the authorization of the PIPS device for primary diagnosis allows the PIPS system to properly serve as a predicate device for future 510(k) submissions for WSI systems seeking clearance for intended use of WSI. The FDA has indicated that a clinical study to prove repeatability and reproducibility of histological feature detection by viewing WSI images is a "special control" required to acquire that clearance.

Presently, Hamamatsu is developing a digital slide scanner system, the NanoZoomer S360MD Digital Slide Scanner System (NanoZoomer), for the same intended use as the PIPS system. Thus, Hamamatsu will make the submission in the form of a 510(k) premarket notification with the PIPS device as the predicate device and will accordingly adhere to the special controls that were established. Hamamatsu will follow similar study designs to test the NanoZoomer system, and intends to use the data from this study for the 510(k) submission to clear its NanoZoomer system for the same intended use. Additional study design input became available from the FDA presubmission review of the study protocol. (4; 5; 6; 7; 8; 9)

In order for pathologists to make a primary diagnosis using a WSI image they need to be able to both detect and interpret various microscopic morphological features in the pathology digital image. For example, pathologists frequently rely on finding mitotic figures and/or necrosis in tumors to help them make a diagnosis of malignancy. This ability may vary among pathologists (e.g., skill increases with experience), between cases (e.g., simple versus complicated biological entities), and with diagnostic image quality (e.g., may be easier with higher resolution). In general, the reproducibility of feature recognition and related accuracy of diagnosis in the published literature is poor (10); (11); (12); (13); (14); (15); (16); (17); (18)

This study is being conducted in order to meet the regulatory requirements for feature detection of the Hamamatsu NanoZoomer S360 MD system. In particular, the study will determine how well a pathologist can repeatedly identify key histologic features present in a "field of view" (FOV) of a region of the histological slide at 3 different magnification levels. The reader pathologist will read multiple tag image file format (TIFF) images of these FOVs from multiple scans across the study.

There are three sub-studies evaluating the within and between scanner precision as well as the between site precision, namely:

- Intra-Scanner Precision Sub-study
- Inter-Scanner Precision Sub-study
- Inter-Site Precision Sub-study

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary and analysis of data collected within the scope of the Feature Detection study. The SAP should be read in conjunction with the study protocol and the Case Report Forms (CRFs).

2. STUDY OBJECTIVES

2.1 Primary Objective

To evaluate the repeatability and reproducibility of detection of histological features when using WSI under following sub-studies:

1. Scans within scanner (Intra-scanner Precision),
2. Scanners within site (Inter-scanner Precision), and
3. Scans between Sites (Inter-site Precision).

2.2 Primary Endpoint

The endpoints are specific to each sub-study:

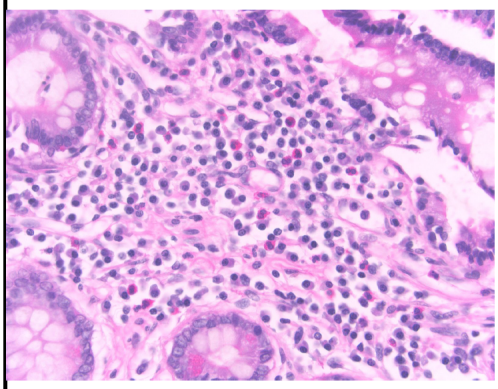
1. Overall percent positive agreement between scans within a given scanner (Intra-scanner precision)
2. Overall percent positive agreement between scanners (Inter-scanner precision)
3. Overall percent positive agreement for scans between sites (Inter-Site precision)

3. STUDY DESIGN

The features study is designed to assess the precision of repeatability and reproducibility of the NanoZoomer image system. Readings of glass slides will not be included in this study. The design of each sub-study will be described in more detail in each sub-study section.

Precision of the NanoZoomer will be assessed in three sub-studies: In all three sub-studies, pathologists will review several slides consisting of different features (one per slide) from multiple organs and from different magnifications. These slides will be chosen to meet pre-specified inclusion and exclusion criteria. The pathologist will not be presented with the entire WSI image but with a very limited FOV in the form of a TIFF image that contains a known particular feature. The TIFF images with the FOVs will be created from repeated 40x WSI scans from different scans, scanners, and sites, depending on the requirements of the sub-study.

The size of the TIFF image shall be standardized and comparable to the amount of surface area seen at that magnification (10x, 20x or 40x). The pathologist will be asked to choose all the features detected in the FOV from the multiple-choice list. The multiple-choice list shall include 10 choices, the 7 study features and the 3 sham features for that magnification, such as in the following example (see Figure 1).



List of Features		
Feature #	Choice	40 x Magnification Group
1	Y or N	Cilia
2	Y or N	Eosinophil granules
3	Y or N	Mitotic Figure
4	Y or N	Nuclear Membrane
5	Y or N	Nucleolus
6	Y or N	Prickles Desmosomes
7	Y or N	Hemosiderin
8	Y or N	Intranuclear inclusion
9	Y or N	Melanin pigment
10	Y or N	Crystals

Legend: Eosinophil granules is the enrolled feature for this FOV. The magnification is 40x. The Reader Pathologist, however, may select any and all of the 10 multiple choices features listed, if detected in the FOV. The primary feature is present in the FOV, but other secondary feaures may also be in the same FOV.

Figure 1: Example CRF for Feature Detection at 40x Magnification

3.1 Required Number of Slides

The required number of slides is based upon the previously conducted study for the PIPS system. In all three sub-studies, the precision will be based on three reading pathologists' assessments and identification of specific histologic features that are observed in TIFF images of FOVs containing a primary feature.

In two PowerPoint presentations (19), (20) the FDA has stated that manufacturers seeking clearance of their WSI devices for primary diagnosis in surgical pathology must demonstrate that a pathologist using WSI technology can achieve reproducible detection of histologic features. Those presentations provide only some general guidelines regarding the number of features, organ systems, magnification at which feature is detected, and number of scanners and readers. The additional information from the clearance letter and device summary provides deeper insight into the approval process for which substantial equivalence must be demonstrated.

In the FDA information, three sub-studies for intra-scanner, inter-scanner, and inter-site repeatability and reproducibility are described (21) and further information is also provided related to the number of sites, scanners, pathologists, features, organs, magnifications, and examination of a limited FOV for detection of features. In addition, the precision of feature detection is mentioned as a special control for WSI clearance. (2)

Outside of the FDA released information, further information is available from ClinicalTrials.gov study registration website (22) as well as presentations made by Philips, (23) in May 2017, at the Pathology Informatics Summit in Pittsburgh, PA. (24) Philips presented their features study with the features selected, the different magnifications, multiple choice selections of the expected feature, inclusion and exclusion criteria for slides with the desired feature, the limited FOV used for feature detection, and the acceptance criterion. According to the presentation, the acceptance criterion was established as the lower limit of the 95% confidence interval greater than 85% for overall agreement. Additional discussions with FDA clarified that the acceptance criterion for the current study was the lower limit of the 95% confidence interval for Positive Agreement.

The twenty-one (21) study features and 9 sham features selected for this study are provided in Table 1. Each selected feature will be selected from at least three organs. There will be 6 FOVs per organ within each combination of feature-magnification. The 21 study features will be evaluated at their relevant magnification. The levels of viewing magnification will be 10x, 20x, and 40x. Each level of magnification will include seven study features. For each study feature, three organs will be selected. For each organ, six FOVs will be selected, and it is assumed that each FOV contains one primary feature.

Table 1: Features and Magnifications for Viewing

	Magnification / Rationale		
	10x	**20x**	**40x**
Feature #	**Tissue level**	**Cellular level**	**Subcellular level**
1	Small artery	Reed Sternberg cell	Cilia
2	Psammoma body	Polymorph neutrophil	Eosinophil with granules
3	Keratin pearl	Plasma cell	Mitotic Figure
4	Granuloma	Goblet cell	Infiltrating or metastatic lobular carcinoma
5	Adipocyte	Macrophage	Osteoid Matrix
6	Gland	Foreign body giant cell	Intercellular bridges
7	Necrosis	Clear Cell (of renal cell carcinoma)	Hemosiderin
8	Cartilage	Myxoid stroma	Intranuclear inclusion
9	Duct	Muscle cell	Melanin pigment
10	Nerve	Calcification	Crystals

Thus, the required number of study slides is 21 study features x 3 organs/study feature x 6 FOV/organ = 378 slides, which include selected study features and selected FOVs (see Table 2). An additional 162 sham slides will also be included.

The overall study design and required number of slides with selected features and selected shams are shown in Figure 2. The same slides will be used for all three sub-studies.

Table 2: Required Number of Slides with Selected Features and Selected FOVs

Feature	Organ	Slides with Selected Feature and Corresponding FOVs
1	1	1-6
1	2	7-12
1	3	13-18
2	1	19-24
2	2	25-30
2	3	31-36
.	.	.
.	.	.
.	.	.
21	1	361-366
21	2	367-372
21	3	373-378

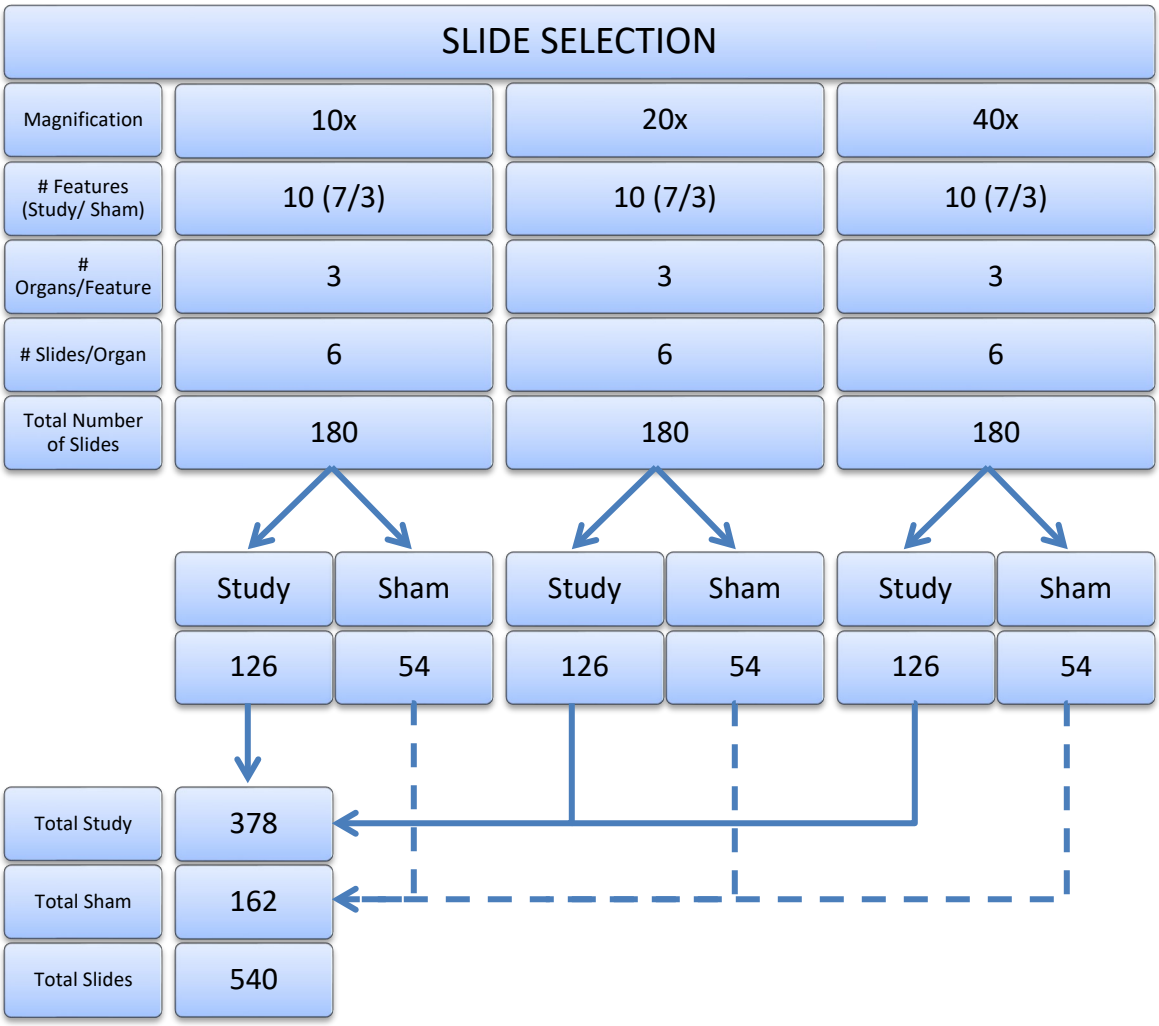


Figure 2. Slide Selection

3.1.1 Intra-Scanner Precision Sub-study

The precision for the intra-scanner sub-study will be based on the assessments from three reading pathologists' who will identify specific histologic features that they observe in TIFF images of the FOV obtained from three scans from three scanners during three readings sessions. The study is diagrammed in the first column of Figure 3.

3.1.2 Inter-Scanner Precision Sub-study

The precision for the inter-scanner sub-study will be based on the assessments from three reading pathologists' who will identify specific histologic features that they observe in TIFF images of the FOV obtained from one scan from each of three scanners during three readings sessions. The study is diagrammed in the second column of Figure 3.

3.1.3 Inter-Site Precision Sub-study

The precision for the inter-site sub-study will be based on the assessments from one reading pathologist at each of three sites. The reading pathologists will identify specific histologic features that they observe in TIFF images of the FOV obtained from one scan from the one scanner assigned to their site during one reading sessions. The study is diagrammed in the third column of Figure 3.

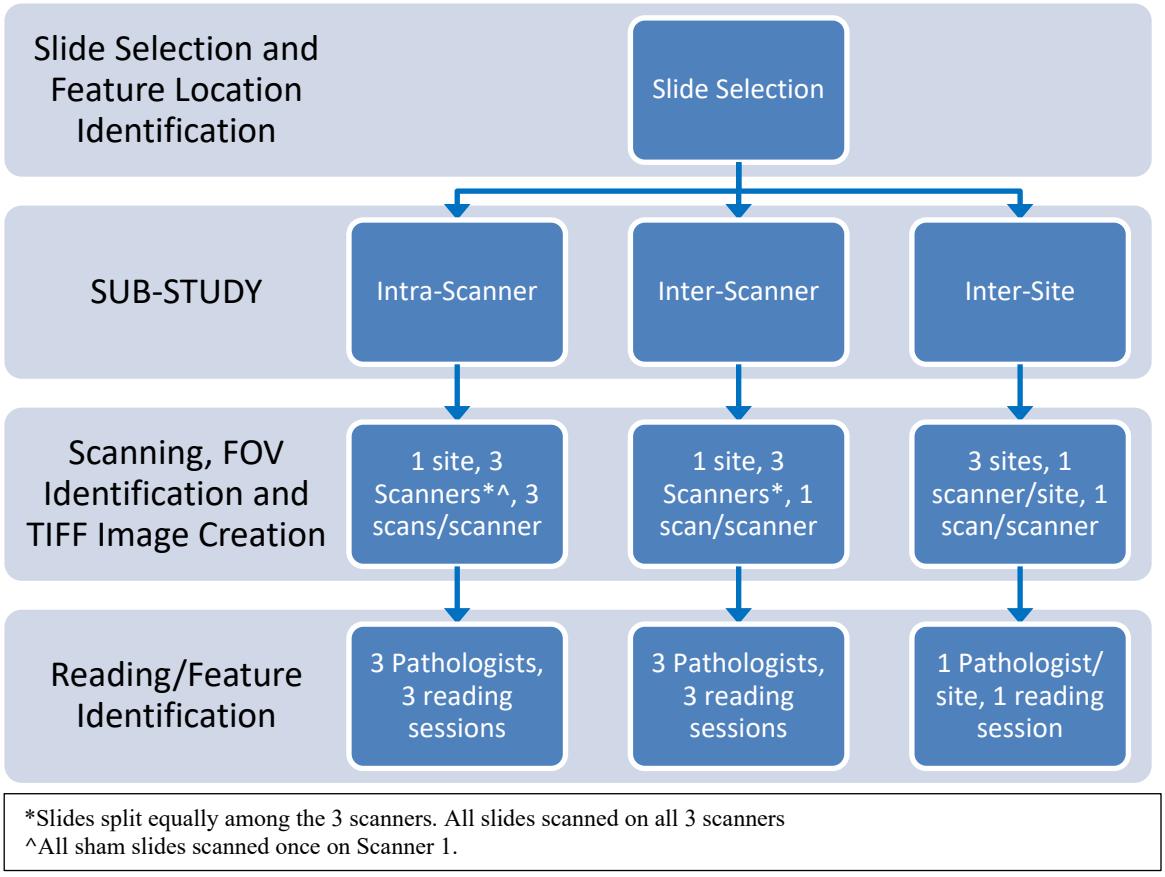


Figure 3. Study Schema

4. RANDOMIZATION AND BLINDING

To minimize bias due to recall of FOVs from previous reads, different Reading Pathologists (RPs; or Readers) will be used in the three different sub-studies.

4.1 Method of Assignment and Randomization

For the study, there are several levels of randomization:

- Assignment of slides to sets and FOV location (study-level randomization)
- Assignment of sets to scanners (sub-study level randomization)
- Assignment of scans (iterations) to rotation for reading sessions (sub-study level randomization)

4.1.1 Study-level Assignment of Slides to Sets

A stratified randomization will be used to assign slides to one of three sets, separately for study features and sham features. The 378 study FOVs will be divided into three sets of 126 slides such that 42 slides are from

each magnification. To assure this, the following factors will be considered in the stratified randomization: FOV magnification (10x, 20x, 40x), Feature, and Organ within feature. The 162 sham FOVs will be similarly divided into three sets using the same stratification factors. These three sets will remain constant for all three sub-studies. The assignment of FOVs to sets will not be blinded. FOVs will be indexed, for linkage with sub-study randomizations, as a two-digit feature number, a two-digit magnification number, a one-digit organ number, and a 1-digit slide number, all separated by a dash (e.g., 01-10-1-5 refers to small artery-10x magnification-organ number 1- slide 5). Additionally, a 4-by-4 grid will be used to randomly determine the location of the feature in the FOV as shown below where the 1-1 cell represents the upper left grid (Cell 1) and the 4-4 cell represent the lower right grid (Cell 16).

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

4.1.2 Intra-scanner Precision Sub-study Random Assignments

This sub-study includes one site with three scanners (called scanner A, scanner B, and scanner C). The random assignments required for this study are described in general below and in detail in Section 4.1.2.1 and 4.1.2.2.

A random assignment of sets to order of scanning on the Scanners will be generated. To accomplish this, a random ordering of 'First' 'Second' 'Third', three times will be performed and merged with the assignment of sets to scanners for each reader.

The random assignment of sets to scanners is specific to each reader. To accomplish this, a 3-by-3 Latin Square will be utilized where rows refer to RPs and columns refer to Scanners.

Once the three scans are generated, these iterations from each scanner are then assigned to a reading session such that in each reading session, one scan (or iteration) from each scanner is included in each reading session. To accomplish this, a second 3-by-3 Latin Square will be utilized where the rows represent the iteration and the columns represent the reading session. Further, FOVs may be randomly rotated in 90-degree intervals to minimize bias due to recall of the FOV; in addition, the order of the FOVs to be displayed within each session may be randomly ordered.

4.1.2.1 Assignment of Sets to Scanners

The order of Set 1, then Set 2, and then Set 3 can be represented as 123; order of Set 2 then Set 3 then Set 1 can be represented as 231; and, the order of Set 3, then Set 1, then Set 2 can be represented as 312. These orders represent three rows of a 3 x 3 Table where the columns represent the Sets. Then this will create the Latin Square shown in Table 3.

Table 3: Example of a Latin Square

	Column 1	Column 2	Column 3
Row 1	1	2	3
Row 2	***2***	***3***	***1***
Row 3	3	1	2

Each number appears once in each row and column. The number refers to a set of 126 slides.

By definition, a 3-by-3 Latin Square has the property that the numbers 1, 2, 3 (or the letters A, B, C) are represented once in each column and one in each row. The above Latin Square is one of twelve (12) possible Latin Squares that can be generated from the numbers 1, 2 and 3. The 12 possible Latin Squares are given in Table 4.

Table 4: List of All Possible 3-by-3 Latin Squares

$$\begin{bmatrix}1&2&3\\2&3&1\\3&1&2\end{bmatrix}, \quad \begin{bmatrix}1&2&3\\3&1&2\\2&3&1\end{bmatrix}, \quad \begin{bmatrix}1&3&2\\2&1&3\\3&2&1\end{bmatrix}, \quad \begin{bmatrix}1&3&2\\3&2&1\\2&1&3\end{bmatrix},$$

$$\begin{bmatrix}2&1&3\\1&3&2\\3&2&1\end{bmatrix}, \quad \begin{bmatrix}2&1&3\\3&2&1\\1&3&2\end{bmatrix}, \quad \begin{bmatrix}2&3&1\\1&2&3\\3&1&2\end{bmatrix}, \quad \begin{bmatrix}2&3&1\\3&1&2\\1&2&3\end{bmatrix},$$

$$\begin{bmatrix}3&2&1\\1&3&2\\2&1&3\end{bmatrix}, \quad \begin{bmatrix}3&2&1\\2&1&3\\1&3&2\end{bmatrix}, \quad \begin{bmatrix}3&1&2\\1&2&3\\2&3&1\end{bmatrix}, \quad \begin{bmatrix}3&1&2\\2&3&1\\1&2&3\end{bmatrix},$$

One of the 12 Latin Squares will be selected by random and the rows randomly assigned to a RP. For example, for the following selection,

$$\begin{matrix}2&3&1\\1&2&3\\3&1&2\end{matrix}$$

there are 6 possible choices for assigning the rows to the 3 readers. The choices are shown in Table 5.

Table 5: Possible Assignments of Latin Square Rows to Readers

Pathologist	Choices for Assigning Each Combination to a Reader					
	1	**2**	**3**	**4**	**5**	**6**
Reader 1	231	*231*	123	312	123	312
Reader 2	312	*123*	231	231	312	123
Reader 3	123	*312*	312	123	231	231

One combination (the selected Latin Square) will be randomly chosen and assigned to readers. Suppose the selected combination is choice 2. Then combination 231 will be assigned to Reader 1, combination 123 will be assigned to Reader 2 and combination 312 will be assigned to Reader 3.

Once the combination is assigned to each reader, the three scanners will be assigned in such a way that the first number of the combination is assigned to Scanner A, the second number is assigned to Scanner B and the third number is assigned to Scanner C. For choice 2 the assignment of readers and scanners for the selected Latin Square is given in Table 6.

Table 6: Assignment of Latin Square Columns to Scanners

	Scanner A	**Scanner B**	**Scanner C**
Reader 1	2	3	1
Reader 2	1	2	3
Reader 3	3	1	2

The number refers to a set of 126 slides.

Once the sets are assigned to Scanners for a Reader, the given set will be scanned three times (called Iteration 1 [or Scan 1], Iteration 2 [or Scan 2], or Iteration 3 [or Scan 3]) on the specified Scanner, with at least six hours downtime between Iterations to ensure full cool down. The resulting scan can be indexed by the set number and iteration number as given in Table 7 for **Reader 1.**

Table 7: Indexing of Iteration to Set or Scanner for Reader 1

Scanner	Set	Iteration		
		Iteration 1	**Iteration 2**	**Iteration 3**
Scanner A	2	2.1 [Set 2, Iteration 1]	2.2	2.3
Scanner B	3	3.1	3.2	3.3
Scanner C	1	1.1	1.2	1.3

All 162 sham slides will be scanned once on Scanner A.

4.1.2.2 Assignment of Iterations to Reading Sessions

Once the slides are scanned three times and the FOVs created, the resulting FOVs need to be randomly assigned to a reading session for each Reader. From the above scanner by iteration Table (Table 7), one of

the 12 possible Latin Squares listed in Table 4 can be selected to assign the set-specific iterations to the three reading sessions for a given reader. The random assignment is such that for a given reading session each reading set comes from a different scanner and a different iteration. Continuing the example from the previous section, one possible Latin Square is given in Table 8 for the three reading sessions for **Reader 1**.

Table 8: Assignment of Iterations to Reading Sessions for Reader 1

		Reading Session		
		1	**2**	**3**
Scanning Set	***2***	**Scanner A, Iteration 1**, (2.1) (n=126)	**Scanner A, Iteration 3**, (2.3) (n=126)	**Scanner A, Iteration 2**, (2.2) (n=126)
	3	**Scanner B, Iteration 3**, (3.3) (n=126)	**Scanner B, Iteration 2**, (3.2) (n=126)	**Scanner B, Iteration 1**, (3.1) (n=126)
	1	**Scanner C, Iteration 2**, (1.2) (n=126)	**Scanner C, Iteration 1**, (1.1) (n=126)	**Scanner C, Iteration 3**, (1.3) (n=126)
Total		**378**	**378**	**378**

Set1, set2, and set3 do not need to be equal, but they need to add up to 378 slides and each set should have equal number of slides from each of 3 viewing magnifications

In this example, the first reading session will include the first iteration from Set 2 (Scanner A), the third iteration from Set 3 (Scanner B), and the second iteration from Set 1 (Scanner C) for Reader 1. Per reading session***, 54 different sham slides will be added to each reading session.*** The set assignment for the sham slides described in Section 4.1.2.1 will be used to assign slides to reading sessions; this assignment is built into the randomization schedule.

Independently, each iteration of each FOV for each Reader will be randomly rotated for presentation during the reading session. This will be performed by randomly selecting one of the four possible rotations: 0° (no rotation), +90°, -90°, or 180° with replacement. Sham slides and slides associated with Iteration 1 will not be rotated. This randomization schedule will then be linked to the above assignment of iterations to reading sessions to obtain the final randomization. Additionally, the FOV presentation order within a session for a Reader will be randomly ordered.

Three (3) separate reading sessions will be performed by each of three reading pathologists, with a washout period of at least two weeks between reading sessions.

The process described above for Reader 1 will be repeated for Readers 2 and 3. At the end, all 378 slides are scanned by each scanner and read by each reader.

4.1.3 Inter-scanner Precision Sub-study Random Assignments

This sub-study includes one site with three scanners (called scanner A, scanner B, and scanner C). The random assignments required for this study are described in general below and in detail in Section 4.1.3.1 and 4.1.3.2.

The random assignment of sets to scanners is to determine the order of scanning on the Scanners. To accomplish this, a 3-by-3 Latin Square will be utilized where rows refer to Sets and columns refer to Scanners.

Once the scans from the three scanners are generated, these scanner-specific scans are then assigned to a reading session such that in each reading session, one scan from each scanner is included in each reading

session. To accomplish this, a second 3-by-3 Latin Square will be utilized where the rows represent the iteration and the columns represent the reading session. Further, FOVs may be randomly rotated in 90-degree intervals to minimize bias due to recall of the FOV; and, the order of the FOVs to be displayed within each session may be randomly ordered.

4.1.3.1 Assignment of Sets to Scanners

Similar to the Intra-scanner Precision sub-study, one of the 12 Latin Squares depicted in Table 4 will be randomly chosen to assign Sets of study FOVs to Scanners, say

2 3 1
1 2 3
3 1 2

For this selected Latin Square, the assignment of sets to scanners and the scanning order is depicted in Table 9. Reading across the row for a given set, Set 1 is scanned on Scanner C first, then on Scanner A and then on Scanner B. Alternatively, reading down the column for a given scanner, Set 2 is scanned first on Scanner A, followed by Set 1 and then by Set 3.

Table 9: Order of Scanning Sets on Scanners

	Scanner A	**Scanner B**	**Scanner C**
Set 1	Second	Third	First
Set 2	First	Second	Third
Set 3	Third	First	Second

All 162 sham slides will be scanned once on Scanner A.

4.1.3.2 Assignment of Scans to Rotations and Reading Sessions

Once the FOVs are created, the resulting FOVs need to be randomly assigned to a reading session and a determination made as to which rotation of the FOV to display (0° [no rotation], 90°, 180°, or 270°) during the reading session. To accomplish the assignment to a reading session, a second 3-by-3 Latin Square will be utilized for each Reader where the rows represent sets, the columns represent the reading session, and the entries represent scanners, say.

3 1 2
1 2 3
2 3 1

For this selected Latin Square, the assignment of set from scanner to a reading session is depicted in Table 10. Reading across the first row, Set 1 scanned on Scanner C is read during Session 1, Set 1 scanned on Scanner A is read during Session 2, and Set 1 scanned on Scanner B is read during Session 3. Alternatively, reading down the first column, Reading Session 1 contains Set 1 scanned on Scanner C, Set 2 scanned on Scanner A, and Set 3 scanned on Scanner B. Per reading session***, 54 different sham slides will***

be added to each reading session. The set assignment for the sham slides described in Section 4.1.3.1 will be used to assign slides to reading sessions; this assignment is built into the randomization schedule.

Table 10: Assignment of Scanner-Set Combinations to Reading Sessions for Reader 1

		Reading Session		
		1	**2**	**3**
Set	***1***	**3 (Scanner C)** (n=126)	**1 (Scanner A)** (n=126)	**2 (Scanner B)** (n=126)
	2	**1 (Scanner A)** (n=126)	**2 (Scanner B)** (n=126)	**3 (Scanner C)** (n=126)
	3	**2 (Scanner B)** (n=126)	**3 (Scanner C)** (n=126)	**1 (Scanner A)** (n=126)
Total		**378**	**378**	**378**
Set1, set2, and set3 do not need to be equal, but they need to add up to 378 slides and each set should have equal number of slides from each of 3 viewing magnifications				

After generating Table 10 for Reader 2 and Reader 3 using the same procedure as for reader 1, then the set assignment for each scanner for a given reading session for a given reader can be summarized as shown in Table 11. As shown in the table, sets 1, 2, and 3 scanned on Scanner A, will be read by all 3 readers and similarly, sets 1, 2, and 3 scanned on Scanners B and C will be read by all 3 readers.

Table 11: Slide Reading Schedule

Reader	Reading Session	Scanner A	Scanner B	Scanner C
1	1	Set 2	3	1
1	2	1	2	3
1	3	3	1	2
2	1	1	2	3
2	2	2	3	1
2	3	3	1	2
3	1	3	1	2
3	2	2	3	1
3	3	1	2	3
Total scans read		1134	1134	1134

Once the FOVs are assigned to a reading session, then a determination will be made as to which rotation will be presented using the same approach as for the Intra-Scanner precision sub-study. Sham slides and slides associated with scanning order “First” will not be rotated. These two randomization schedules will be combined to link order of presentation with rotation; the randomizations will be produced for each Reader separately.

4.1.4 Inter-site Precision Sub-study Random Assignments

This sub-study includes three sites with one scanner each. The random assignments required for this sub-study are not as complex as for the intra- and inter-scanner sub-studies. For this sub-study, all slides are scanned once at each site. For each reader (1 per site), the order of presentation of the FOVs will be random. No sham slides will be included.

The order in which the sets will be presented to the Reader will be randomly determined. This approach ensures that after every 126 slides, each magnification has had an equal number of features read.

4.2 Blinding and Unblinding

The randomization schedules will be kept blinded to the extent possible. The assignment of slides then sets and the location of the FOV will be known to personnel involved in the study.

5. WITHDRAWAL OR DISCONTINUATION

5.1 Slide Withdrawal

It is unlikely that slides will be withdrawn from the study after they satisfy inclusion and exclusion criteria. However, if during the trial, before all slides have been scanned, a slide somehow gets damaged or misplaced and cannot be used, or an inadvertent mix-up is discovered, this slide may be withdrawn from the study (or a sub-study) and a protocol deviation noted. If necessary, a replacement slide from a new case may need to be screened, enrolled and added into the study (or sub-study) in order to continue completing the required imaging to satisfy the overall number of features examined for the given sub-study, and thus the study.

If a sub-study has been completed, the new slide will only be included in the sub-study(ies) which have not yet completed. The new slide will need to be scanned, replacing the withdrawn slide.

5.2 Pathologist Withdrawal

If a reading pathologist is unable to complete study reads according to study schedules, a replacement pathologist may be selected.

For the intra-scanner (inter-scan) and inter-scanner sub-studies, because of how the slides are split across reading sessions, the replacement pathologist will need to re-read any **slides** read by the original pathologist such that the **three reading sessions of the same slide** are from the **same pathologist**.

For the inter-site sub-study, because slides are only read once, the replacement pathologist will need to re-read any slides read by the original pathologist.

The primary analysis will not use any readings from the original/previous pathologist.

6. STATISTICAL ANALYSIS

To minimize bias due to recall of FOVs from previous reads, different reader pathologists will be used in the 3 different sub-studies.

6.1 Pathologist Experience

Reader information (years of experience post-residency, categorized average number of cases per year [<5,000, 5,000 to 10,000, >10,000]) will be tabulated by sub-study and overall using descriptive statistics.

6.2 FOV Characteristics

The characteristics of the cases selected will be descriptively summarized using numbers and percentages by magnification and overall. The characteristics of the cases will include the features selected and the organ systems from which the features were selected.

6.3 Protocol Deviations

Protocol deviations will be listed by site, reader and sub-study. Additionally, the number and percentage of deviations will be summarized by site, reader and sub-study.

6.4 Intra-Scanner Precision Sub-study

To evaluate Intra-Scanner precision, the agreement rate of each reader for each scanner is first estimated, then averaged over all readers and scanners. Sham features are excluded from analysis. (see Section 4.1.2 for details of the randomization)

Each of three scanners will scan one-third of the slides three times; three Reading Pathologists (RPs; or Readers) will evaluate all FOVs across three reading sessions. Table 12 shows the number of comparison pairs for each scanner.

Table 12: Intra-Scanner Pairwise Comparison

Scanner	Number of Comparison Pairs
Scanner A	1134
Scanner B	1134
Scanner C	1134
Overall	**3402**

The number of comparison pairs is calculated based on three RPs where for each RP there are three pairwise comparisons, namely Iteration 1 vs. Iteration 2, Iteration 1 vs. Iteration 3, and Iteration 2 vs. Iteration 3. This results in 3 reader x 126 pairs per comparison x 3 pairwise comparisons =1134 (see Table 13) comparisons per scanner.

Table 13: Number of Comparison Pairs per Scanner

	Number of Pairwise Comparisons			
Reader	**Iteration 1 vs. Iteration 2**	**Iteration 1 vs. Iteration 3**	**Iteration 2 vs. Iteration 3**	**Total # of Pairwise Comparisons**
1	126	126	126	378
2	126	126	126	378
3	126	126	126	378
				1134

6.4.1 Percent Positive Agreement

For a given scanner, the pairwise comparisons among Iteration 1, Iteration 2 and Iteration 3 can be displayed as three (3) 2x2 tables for a given reader. Table 14 demonstrates a 2x2 Table for the pairwise comparison of Iteration 1 versus Iteration 2 for Reader1 for Scanner A. The result of reading for each slide is "Yes" or "No" if the primary feature was detected correctly. Each primary feature shall be considered equally important as the next primary feature; thus, the results of all primary features will be pooled.

Table 14: Agreement Table for Scanner A for Reader 1

		Feature X is correctly identified in Iteration 2		
		Yes	**No**	
Feature X is correctly identified in Iteration 1	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	126

The nine (9) 2x2 tables for a given scanner will be pooled across the 3 readers to create one 2x2 Table (see Table 15) for the agreement between Iteration *i* and Iteration *j*, where *i, j* = 1, 2, or 3; $i \neq j$; and, $j > i$.

Table 15: Agreement Table for Scanner A Across all Readers

		Feature X is correctly identified in Iteration *j*		
		Yes	**No**	
Feature X is correctly identified in Iteration *i*	**Yes**	a^1	b^1	a^1+b^1
	No	c^1	d^1	c^1+d^1
		a^1+c^1	b^1+d^1	1134

Then the Percent Positive Agreement (PPA) can be estimated as $a^i/1134$ for $i = 1$ to 3, corresponding to Scanner A, Scanner B and Scanner C. Similarly, the results across all three scanners can be summed to obtain the overall PPA (see Table 16).

Table 16: Intra-Scanner Results Across 3 Reading Pathologists for Each Scanner and Overall

	Number of Pairwise Agreements	**Number of Comparison Pairs**	**Agreement Rate %**
Scanner A	a^1	1134	$(a^1)/1134$
Scanner B	a^2	1134	$(a^2)/1134$
Scanner C	a^3	1134	$(a^3)/1134$
Overall	a^*	3402	$(a^*)/3402$

To preserve the correlation structure of multiple readings of the same feature on an FOV, the bootstrap method will be used to derive a two-sided 95% confidence interval (CI) for the percent positive agreement. An FOV will be the bootstrap re-sampling unit.

The strategy in bootstrapping is to approximate the data generating mechanism (the data distribution function) using an empirical estimate. By design, the 378 study features are divided into 3 sets of 126 FOVs and each set is randomly assigned to a Scanner. Then 3 repeated scans (iterations) for each FOV and within each scanner are created for evaluation by each Reader. Since sets of FOVs are nested within scanners in this study, for bootstrap resampling the 3 scanners will be chosen **without** replacement (i.e., the same scanner cannot be sampled more than once).

After all data are collected, the database is cleaned and locked, the following data generation mechanism will be followed to preserve the correlation structure in this sub-study.

a. Choose 3 scanners **without** replacement (because each reader should read 378 FOVs (126 FOVs per scanner)
b. For each scanner, choose 126 FOVs **with** replacement from each set assigned to that scanner.
c. Identify the reader for each of the selected FOV-scanner combinations.
d. For each FOV-scanner-reader combination, calculate the within RP agreement status between reading sessions (1, 2), (1, 3), and (2, 3).
e. Estimate overall intra-scanner percent positive agreement using the sampled data.
f. Repeat step a to step e at least 1000 times to estimate a two-sided 95% confidence interval for the percent positive agreement.

Note: when the estimated agreement rate is at extreme values such as 0% or 100%, the arcsine variance transformation approach, using the Anscombe correction for the continuity [25] will be used to calculate the 95% confidence interval instead of using the bootstrap.

The study's acceptance criterion for percent positive agreement is 85%. The lower limit of the 95% CI for the percent positive agreement will be compared to 85%. If the lower limit is greater than 85%, then the study will have demonstrated intra-scanner precision.

6.4.2 Overall Agreement and Negative Agreement

In addition to percent positive agreement, percent negative agreement and the overall agreement rate will be estimated with their corresponding two-sided 95% confidence intervals.

From Table 15, which is all pairwise comparisons within Scanner A across all 3 reading pathologists, the overall agreement rate can be estimated as $(a^i + d^i)/1134$, and the percent negative agreement can be estimated as $d^i/1134$, for $i = 1$ to 3 corresponding to Scanner A, Scanner B and Scanner C. The results can be summed across all 3 readers to obtain the respective overall estimates.

Two-sided 95% confidence intervals for the percent negative agreement and the overall agreement rate will be estimated using bootstrapping. The rates will be estimated concurrently with positive agreement; no separate bootstrapping will be performed.

6.4.3 Secondary Features

The objective of this analysis is to investigate the repeatability/reproducibility of the readers' absence or presence calls of the secondary features. For a given magnification, each of the 7 features will be considered primary for 18 study FOVs (3 organs x 6 FOVs per organ) and will be considered secondary for the remaining 108 (6 primary features with the same magnification x 3 organs x 6 FOVs per organ) study FOVs.

It is possible that a given feature may be absent from the majority of the 108 study FOVs. Sham features are not considered.

The secondary features analysis will be based on the repeatability/reproducibility for each secondary feature separately. To perform the secondary analysis for a specific feature, say feature X, 3 scans (Iteration 1, Iteration 2, and Iteration 3) per scanner will be used. In the 3 scans from Set 1 on Scanner A, there are 6 FOVs primary for feature X among 42 FOVs and 36 FOVs secondary for feature X. Thus, 36 FOVs will be used to construct a 2x2 table for whether the secondary feature was observed.

Table 17: Secondary Feature Agreement Rate for Scanner A

		Secondary Feature X is Observed Using Iteration 2		
		Yes	**No**	
Secondary Feature X is Observed Using Iteration 1	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	36

As for overall positive agreement for the primary feature (see Table 15), nine (9) 2x2 tables will be pooled within Scanner A and across 3 readers for the 324 pairwise comparisons (see Table 18).

Table 18: Secondary Feature Agreement Rate for Scanner A Across 3 Readers

		Secondary Feature is Observed Using Iteration *j*		
		Yes	**No**	
Secondary Feature is Observed Using Iteration *i*	**Yes**	a^1	b^1	a^1+b^1
	No	c^1	d^1	c^1+d^1
		a^1+c^1	b^1+d^1	324

The required overall 2x2 table for a specific secondary feature will be constructed by pooling all 2x2 tables across the 3 reading pathologists and across the 3 scanners, as shown in Table 19.

Table 19: Overall Secondary Feature Agreement Rate

		Secondary Feature is Observed Using Iteration *j*		
		Yes	**No**	
Secondary Feature is Observed Using Iteration *i*	**Yes**	a^*	b^*	a^*+b^*
	No	c^*	d^*	c^*+d^*
		a^*+c^*	b^*+d^*	972

This overall contingency table (Table 19) will be generated for each of the 21 study features (7 features per magnification, 3 magnifications) that appear as a secondary feature and the positive, negative, and overall agreement rates for each secondary feature will be estimated as:

positive agreement	$a^*/972$
negative agreement	$d^*/972$
overall agreement	$(a^*+d^*)/972$

To preserve the correlation structure of multiple readings of the same secondary feature on an FOV, the bootstrap method will be used to derive a two-sided 95% confidence interval (CI) for the overall percent positive agreement. The following data generation mechanism will be followed to preserve the correlation structure in this sub-study.

a. Choose 3 scanners **without** replacement.
b. For each scanner, choose 36 FOVs **with** replacement from each set assigned to that scanner.
c. Identify the reader for each of the selected FOV-scanner combinations.
d. For each FOV-scanner-reader combination, calculate the within RP agreement status between reading sessions (1, 2), (1, 3), and (2, 3).
e. Estimate the overall intra-scanner percent positive agreement using the sampled data.
f. Repeat step a to step e at least 1000 times to estimate a two-sided 95% CI for the percent positive agreement.

Two-sided 95% CI for the negative and overall agreement rates will be estimated concurrently using the same bootstrap. When the estimated agreement rate is at extreme values such as 0% or 100%, the arcsine variance transformation approach, using the Anscombe correction for the continuity, will be used to calculate the 95% confidence interval instead of using the bootstrap.

6.5 Inter-Scanner Precision Sub-study

This sub-study includes one site with three scanners (scanner A, scanner B, and scanner C) and 3 Reading Pathologists (different from the Intra-Scanner pathologists). The same set of 378 slides with selected features and FOVs will be used in this sub-study and the same 162 sham slides will be used. Three (3) separate reading sessions will be performed by each pathologist with a washout period of at least two weeks between sessions. (see Section 4.1.3 for details of the randomization)

6.5.1 Percent Positive Agreement

To evaluate Inter-Scanner precision, a 2x2 table similar to Table 14 can be constructed for the pairwise comparisons of scanners for Reader 1. The result of reading for each FOV is "Yes" or "No" if the primary feature was detected correctly for the specified scanner (i.e., Scanner A versus Scanner B). Each primary feature is equally important as the next primary feature; thus, the results of all primary features will be pooled and can be represented as in Table 20.

Table 20: Agreement Table for Reader 1 Across all Scanners

		Feature is correctly identified in Scanner *j*		
		Yes	**No**	
	Yes	a^1	b^1	a^1+b^1

Feature is correctly identified in Scanner *i*	**No**	c^1	d^1	c^1+d^1
		a^1+c^1	b^1+d^1	1134

The inter-scanner precision will be estimated by $(a^i)/1134$ for reader *i*, where *i* = 1, 2, or 3. As for the intra-scanner precision sub-study, the overall inter-scanner precision will be estimated by averaging over all 3402 pairwise comparisons as $(a^*)/3402$.

To preserve the correlation structure of multiple readings of the same feature and possibly multiple features on an FOV, the bootstrap method will be used to derive a two-sided 95% CI for the percent positive agreement. An FOV will be the bootstrap re-sampling unit.

After all data are collected, database is cleaned, and locked; in order to preserve the correlation structure in resampling, we will mimic the following data generating mechanism in this sub-study.

a. Choose 3 scanners **without** replacement.
b. For each set, choose 126 FOVs **with** replacement.
c. For each selected FOV, choose 3 Reading Pathologists **with** replacement.
d. For each selected FOV-Reading Pathologist combination, select the corresponding results from each scanner (i.e., if RP1-FOV125 is the combination, then RP1's results from Scanner A, B and C for FOV125 will be selected).
e. Estimate the overall inter-scanner percent positive agreement using the sampled data.
f. Repeat step a to step e at least 1000 times to estimate a two-sided 95% CI for the percent positive agreement.

The study's acceptance criterion for percent positive agreement is 85%. The lower limit of the 95% CI for the percent positive agreement will be compared to 85%. If the lower limit is greater than 85%, then the study will have demonstrated inter-scanner precision.

6.5.2 Overall Agreement and Negative Agreement

In addition to the percent positive agreement, the percent negative agreement and the overall agreement rate will be calculated with their corresponding two-sided 95% CIs. The rates will be estimated from the overall 2x2 Table based on the 3402 pairwise comparisons of the scanners, across all Readers. The estimates are:

negative agreement	$d^*/3402$
overall agreement	$(a^*+d^*)/3402$

Two-sided 95% CI of the percent negative agreement and percent overall agreement will be estimated from bootstrapping as described above.

6.5.3 Secondary Features

The same methodology will be used to evaluate the secondary features as described in Section 6.4.3, with Scanner A, Scanner B, and Scanner C replacing Iteration 1, Iteration 2 and Iteration 3 respectively.

6.6 Inter-Site Precision Sub-study

In this sub-study, there will be three sites and one Reading Pathologist at each site. The same set of 378 slides with selected features and FOVs will be scanned once at each site. There will be just one reading session for each pathologist and no sham slides will be used.

6.6.1 Percent Positive Agreement

To evaluate inter-site precision, the same methodology as for the intra-scanner and inter-scanner sub-studies will be utilized. First, all pairwise comparisons of features between two sites will be determined (see Table 21). Each primary feature shall be considered as of equal importance as the next primary feature; thus, the results of all primary features will be pooled.

Table 21: Agreement Table for Pairwise Comparisons of Site

		Features are correctly identified at Site j		
		Yes	**No**	
Features are correctly identified at Site i	**Yes**	a^1	b^1	a^1+b^1
	No	c^1	d^1	c^1+d^1
		a^1+c^1	b^1+d^1	378

The inter-site agreement rate can be estimated by summing across all pairwise site comparisons as $(a^*)/1134$.

To preserve the correlation structure of multiple readings of the same feature and possibly multiple features on an FOV, the bootstrap method will be used to derive a two-sided 95% CI for the overall percent positive agreement. An FOV will be the bootstrap re-sampling unit.

After all data are collected, adjudications are completed, database is cleaned, and locked; in order to preserve the correlation structure in resampling, we will mimic the following data generating mechanism in this sub-study.

a. Choose all 3 sites **without** replacement (because each site has its own Reading Pathologist).
b. For each set (set 1, set 2, set 3), choose 126 FOVs **with** replacement.
c. For each selected FOV, select the corresponding results from each scanner (or reader)
d. Estimate overall inter-site percent positive agreement using the sampled data
e. Repeat step a to step d at least 1000 times to estimate a 95% two-sided CI for the percent positive agreement.

The study's acceptance criterion for percent positive agreement is 85%. The lower limit of the 95% CI for the percent positive agreement will be compared to 85%. If the lower limit is greater than 85%, then the study will have demonstrated inter-site precision.

6.6.2 Overall Agreement and Negative Agreement

In addition to the percent positive agreement, the overall agreement and percent negative agreement will be calculated along with their corresponding two-sided 95% CIs. The rates will be estimated from the overall 2x2 Table based on the 1134 pairwise comparisons of the sites. The estimates are:

negative agreement **$d^*/1134$**

overall agreement	$(a^*+d^*)/1134$

Two-sided 95% CI of the negative and overall agreement rates will be estimated from bootstrapping as described above.

6.6.3 Secondary Features

A similar methodology will be used to evaluate the secondary features as described in Section 6.4.3, but Site 1, Site 2, and Site 3 will replace Iteration 1, Iteration 2 and Iteration 3 respectively; and one level of summation is removed (only summation over sites will be performed). Thus there are 36 x 3 x 3 = 324 pairwise comparisons for each secondary feature.

6.7 Analysis Software

All analyses will be performed using SAS Software version 9.4 and R version 4.05.

7. REFERENCES

1. *Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center.* **Pantanowitz, L, et al.** 12, Dec 2013, Archives of Pathology & Laboratory Medicine, Vol. 137, pp. 1710–1722.

2. **FDA.** IntelliSite Pathology Solution (PIPS, Philips Medical Systems). *FDA.* [Online] April 12, 2017. [Cited: July 28, 2018.] https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm553358.htm.

3. —. FDA Device Summary. *FDA.* [Online] October 13, 2017. [Cited: July 28, 2018.] https://www.accessdata.fda.gov/cdrh_docs/pdf16/DEN160056.pdf. DEN 160056.

4. **Holody, M.** Presubmission for NanoZoomer S360 Digital Slide Scanner System. Q171418/A001.

5. **Cui, C.** Re: Written Feedback for Hamamatsu NanoZoomer S360 Digital Slide Scanner System.

6. —. Record of Meeting for NanoZoomer S360 Digital Slide System.

7. **Holody, M.** Meeting Minutes Disagreement-Q171418/A001 – Pre-Submission Meeting Minutes for NanoZoomer S360 Digital Slide System.

8. **Cui, C.** Re: Q171418/A002 Meeting Minutes Disagreement Received. December 28, 2017.

9. —. Re: Q171418/A002 Review Complete. January 3, 2018.

10. *Histologic grading of invasive lobular carcinoma: does use of a 2-tiered nuclear grading system improve interobserver variability?* **Adams, AL, et al.** 2009, Annals of Diagnostic Pathology, Vol. 13, pp. 223-225.

11. *Interobserver concordance in the assessment of features used for the diagnosis of cervical atypical squamous cells and squamous intraepithelial lesions (ASC-US, ASC-H, LSIL and HSIL).* **Bigras, G, et al.** 2013, Cytopathology, Vol. 24, pp. 44-51.

12. *Consistency in the observation of features used to classify duct carcinoma in situ (DCIS) of the breast.* **Douglas-Jones, A G, et al.** 2000, J Clin Pathol, Vol. 53, pp. 596-602.

13. *Interobserver Agreement for Endometrial Cancer Characteristics Evaluated on Biopsy Material.* **Nofech-Mozes, S, et al.** Article ID 414086, 2012, Obstetrics and Gynecology International, Vol. 2012, p. 6 pages.

14. *Interobserver Agreement on Pathologic Features of Liver Biopsy Tissue in Patients with Nonalcoholic Fatty Liver Disease.* **Jung, ES, et al.** 3, 2016, Journal of Pathology and Translational Medicine, Vol. 50, pp. 190-196.

15. *Mitosis Counting in Breast Cancer: Object- Level Interobserver Agreement and Comparison to an Automatic Method.* **Veta, Mitko, et al.** August 16, 2016, PLOS ONE, pp. 1-13.

16. *Mitotic figure recognition: Agreement among pathologists and computerized detector.* **Malona, Christopher, et al.** 2012, Analytical Cellular Pathology, Vol. 35, pp. 97–100.

17. *Gleason grade 4 prostate adenocarcinoma patterns: an interobserver agreement study among genitourinary pathologists.* **Kweldam, CF, et al.** 2016, Histopathology, Vol. 69, pp. 441-449.

18. *The Application of Digital Pathology to Improve Accuracy in Glomerular Enumeration in Renal Biopsies.* **Rosenberg, AZ, et al.** 6, 2016, PLoS ONE, Vol. 11.

19. *FDA PowerPoint Presentation: FDA Regulation of Digital Pathology Devices.* **Hu, Yun-Fu.**

20. *PowerPoint Presentation: Whole Slide Imaging (WSI) Devices – FDA perspective.* **Anderson, Nick.** June 6, 2014, FDA.

21. *FDA Device Summary, Evaluation of Automatic Class III Designation for Philips IntelliSite Pathology System (PIPS), , DEN 160056.* **FDA.** April 13, 2017. DEN 160056.

22. **Instrument Precision Study for Validation of Philips Dx. *ClinicalTrials.gov.* [Online] 2016. [Cited: Jan 21, 2019.] https://clinicaltrials.gov/ct2/show/NCT02699970?term=NCT02699970&rank=1. NCT02699970.**

23. *WSI and Primary Diagnosis: The Journey Begins; What FDA's Approval of WSI Means for You and Digital Pathology.* Abels, Ester. May 11, 2017, 2017 Digital Pathology Association Webinar Series, Session 2.

24. *Philips Presentation on Philips IntelliSite Pathology System.* Feldman, Michael. May 23, 2017, Pathology Informatics Summit, Clinical Studies DeNovo Sesson.

25. *Interval Estimators for Binomial Proprotion: Comparison of Twenty Methods.* Pires, A, and Amado, C. 2008, Statistcial Journal, pp. 165-197.