EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR
Paige Prostate

DECISION SUMMARY

A.   DEN Number:

DEN200080

B.   Purpose for Submission:

De Novo request for evaluation of automatic class III designation for the Paige Prostate

C.   Measurands:

Not applicable

D.   Type of Test:

Software device to identify digital histopathology images of prostate needle biopsies that are suspicious for cancer and to localize a focus with the highest probability for cancer

E.   Applicant:

Paige.AI, Inc.

F.   Proprietary and Established Names:

Paige Prostate

G.   Regulatory Information:

1. Regulation section:

    21 CFR 864.3750

2. Classification:

    Class II

3. Product code:

    QPN

4. Panel:

    88 – PATHOLOGY

H. Indications for use:

1. Indications for use:

   Paige Prostate is a software only device intended to assist pathologists in the detection of foci that are suspicious for cancer during the review of scanned whole slide images (WSI) from prostate needle biopsies prepared from hematoxylin & eosin (H&E) stained formalin-fixed paraffin embedded (FFPE) tissue. After initial diagnostic review of the WSI by the pathologist, if Paige Prostate detects tissue morphology suspicious for cancer, it provides coordinates (X,Y) on a single location on the image with the highest likelihood of having cancer for further review by the pathologist.

   Paige Prostate is intended to be used with slide images digitized with Philips Ultra Fast Scanner and visualized with Paige FullFocus WSI viewing software.

   Paige Prostate is an adjunctive computer-assisted methodology and its output should not be used as the primary diagnosis. Pathologists should only use Paige Prostate in conjunction with their complete standard of care evaluation of the slide image.

2. Special conditions for use statement(s):

   For prescription use only

   For *in vitro* diagnostic (IVD) use only

3. Special instrument requirements:

   Philips IntelliSite Ultra Fast Scanner

   FullFocus image viewing software

I. Device Description:

Paige Prostate is an *in vitro* diagnostic medical device software, derived from a deterministic deep learning system that has been developed with digitized WSIs of H&E stained prostate needle biopsy slides.

Paige Prostate utilizes several accessory devices as shown in Figure 1 below, for automated ingestion of the input. The device identifies areas suspicious for cancer on the input WSIs. For each input WSI, Paige Prostate automatically analyzes the WSI and outputs the following:

- Binary classification of suspicious or not suspicious for cancer based on a pre-defined threshold on the neural network output.
- If the slide is classified as suspicious for cancer, a single coordinate (X,Y) of the location with the highest probability of cancer on an image determined to be suspicious for cancer.

- If the slide is classified as not suspicious for cancer, no additional output will be available by Paige Prostate. The Paige FullFocus WSI viewer will display "Not Suspicious for Cancer – Area of Interest Not Available".
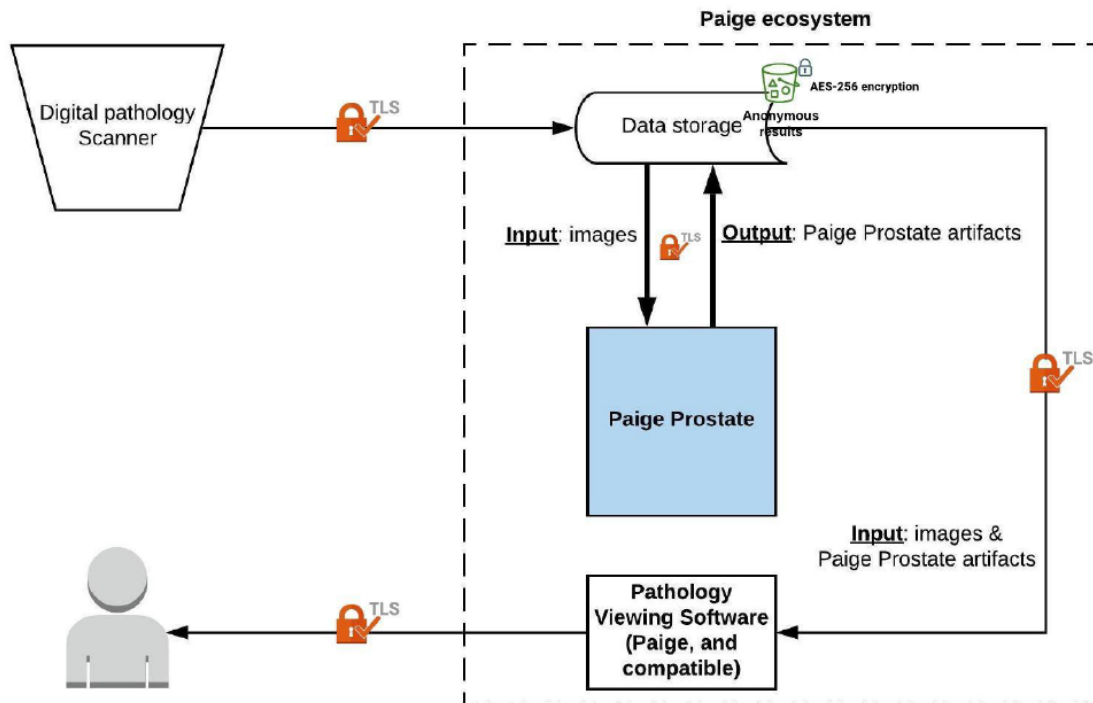

Figure 1: Dataflow and Input/Output Devices for Paige Prostate: (Lock icon refers to the transport layer security (TLS) encryption used for all data transfer between services within the Paige Ecosystem. Data storage is encrypted at rest as indicated by the locked green storage icon).
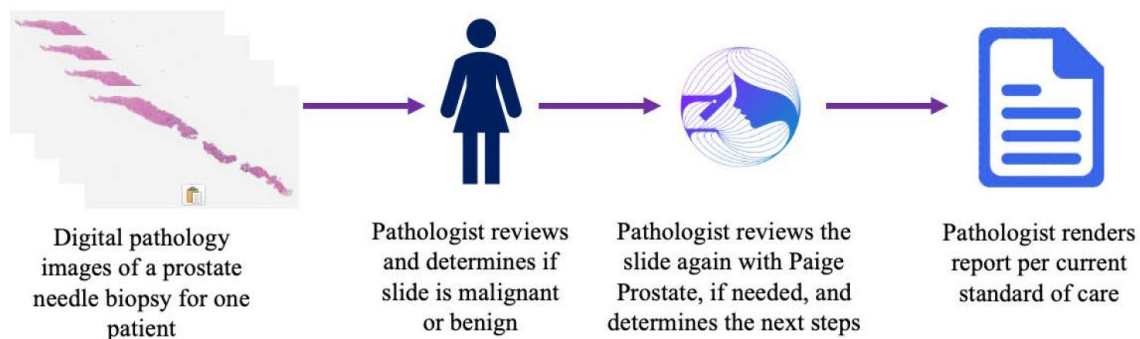

Figure 2: Paige Prostate Pathologist Workflow

Algorithm development: Paige Prostate algorithm development was performed on training, tuning, and test datasets. Each dataset contained slides from unique patients ensuring that training, tuning,

and test datasets do not have any slides, cases, or patients in common. De-identified slides were labeled as benign or cancer based on the synoptic diagnostic pathology report. These datasets were completely independent from the validation dataset.

Table 1: Dataset Split for Training, Tuning and Test Sets

Algorithm Development		
Training Dataset	**Tuning Dataset**	**Test Datasets**
De-identified slides from cases prepared and diagnosed at internal site located in US, from 2013-2017, scanned with an Aperio Leica AT2 scanner Number of slide images: 33,543	Slides prepared and diagnosed at internal site*, scanned with an Aperio Leica AT2 scanner Number of slide images: 5,598	Same as tuning dataset, but scanned on Philips PIPS scanner Number of slide images: 5,598
		Slides prepared at external sites but diagnosed at internal site located in US, scanned with an Aperio Leica AT2 scanner Number of slide images: 10,605

Table 2: Distribution of slide images by race in algorithm development

Race	Training Dataset	Tuning Dataset	Test Dataset
White	27576 (82.21%)	7394 (82.33%)	8313 (78.45%)
Black or African American	2704 (8.06%)	669 (7.45%)	957 (9.03%)
Native American or American Indian	14 (0.04%)	14 (0.16%)	18 (0.17%)
Native Hawaiian or Pacific Islander	0 (0.00%)	0 (0.00%)	2 (0.02%)
Asian-Far East/Indian Subcontinent	1027 (3.06%)	289 (3.22%)	383 (3.61%)
Other	511 (1.52%)	171 (1.90%)	213 (2.01%)
Unknown race	1711 (5.10%)	444 (4.94%)	719 (6.78%)

J. Standard/Guidance Document Referenced:

- Guidance for the Content of Premarket Submission for Software Contained in Medical Devices; May 11, 2005
- CLSI document EP12-A2: User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline – Second Edition, 2008
- Content of Premarket Submissions for Management of Cybersecurity in Medical Devices; October 2, 2014
- The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)]; July 2014
- Guidance for Industry and FDA Staff: De Novo Classification Process (Evaluation of Automatic Class III Designation); October 30, 2017

- Acceptance of Clinical Data to Support Medical Device Applications and Submissions Frequently Asked Questions; February 2018
- Guidance for Industry and Food and Drug Administration Staff - Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approval and De Novo Classifications; August 30, 2019
- Guidance for Off-the-Shelf Software Use in Medical Devices; September 2019

K. Test Principle:

Paige Prostate is operated as follows:
1. Scanned digital images of prostate needle biopsies are acquired using the designated digital pathology scanner. Image and other related quality control steps are performed per the scanner instructions for use and any additional user site specifications. The scanned digital images are immediately processed by Paige Prostate in the background.
2. The pathologist selects a patient case and opens the whole slide image for review in the designated digital pathology viewing software.
3. After the pathologist has fully reviewed all areas on the digital image of a prostate core biopsy slide, and has decided upon a diagnosis of "cancer", "no cancer", or "defer", the pathologist must "activate the Paige Prostate" to view its output.
4. If Paige Prostate detects a region on the digital slide suggestive of carcinoma, it identifies the region with greatest likelihood of being cancer and overlays a mark on that region indicated by its coordinate (X,Y). This is a statistical determination and is not linked to other clinical assessments, such as Gleason score.
5. The pathologist can toggle Paige Prostate outputs on and off to allow unobstructed re-examination of any suspicious regions.
6. If the pathologist has already recognized cancer on the slide, no additional action is required. If the pathologist has indicated a diagnosis of "no cancer" or "defer" and the algorithm indicates a region suspicious for cancer, the pathologist is prompted to re-examine that slide image, focusing initially on the region indicated by the algorithm.
7. If the pathologist determines that the histologic findings warrant a change in diagnosis from "no cancer" to "cancer" or "defer", or from "defer" to "cancer", they then modify the original diagnosis to reflect the additional findings.
8. The final diagnosis of cancer is made by the pathologist based upon the histologic findings and should not be solely based on the algorithm's output.
9. Pathologists should follow standard of care to obtain any additional stains, other pathologists' opinions, and/or additional information, if needed, to render a final diagnosis.
10. The Paige Prostate device does not provide assistance with measuring or grading foci of cancer, whether detected initially by the pathologist or recognized after deployment of the algorithm.

The clinical workflow per prostate biopsy slide (WSI) is shown in Figure 3 below.
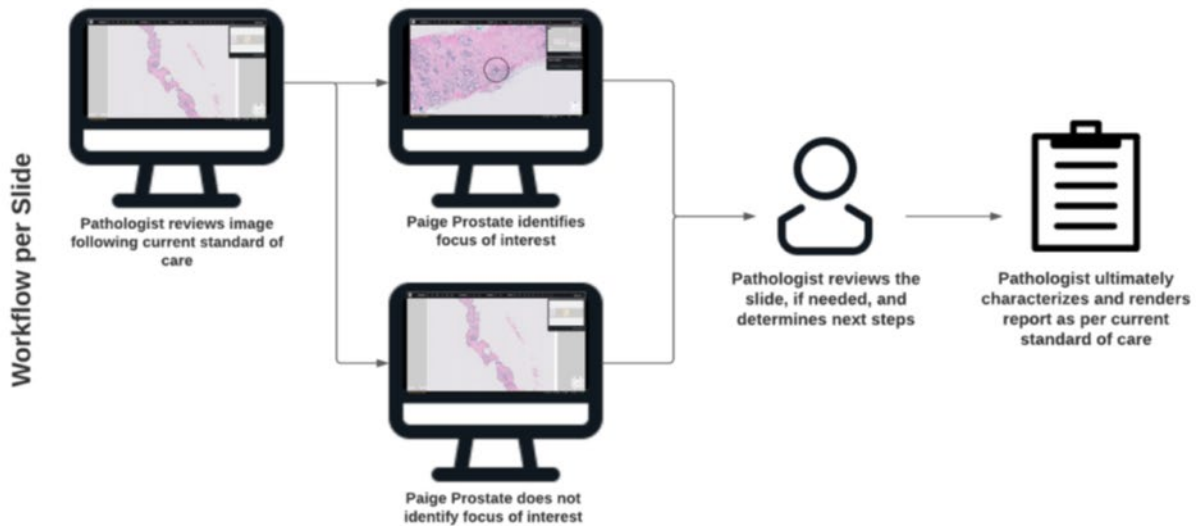
Figure 3: Clinical Workflow per Slide

L. Software:

The Paige Prostate device was identified to have a moderate level of concern as described in the FDA guidance document "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices." (May 11, 2005).

  a. Software Description: Paige.AI provided a general description of the features in the software documentation and in the device description. The description of the software is consistent with the device functionality described in the device description.

  b. Device Hazard Analysis: Paige.AI provided separate analyses of the device and cybersecurity concerns. The content of the hazard analysis is sufficient and assesses pre- and post-mitigation risks. The device hazard analysis includes:
    • identification of the hazard
    • cause of the hazard (hazardous situation)
    • probability of the hazard
    • severity of the hazard
    • method of control or mitigation
    • corrective measures taken, including an explanation of the aspects of the device design/requirements, that eliminate, reduce, or warn of a hazardous event verification of the control implementation, which is traceable through the enumerated traceability matrix.

  c. Software Requirement Specifications (SRS): The SRS includes user, engineering, algorithmic, cybersecurity, and various other types of requirements that give a full description of the functionality of the device. The SRS is consistent with the device description and software description.

d. Architecture Design Chart: Paige.AI provided the software overview and included flow diagrams representative of process flow for various features of the Paige Prostate device.

e. Software Design Specification (SDS): The SDS is traceable to the SRS and demonstrates how individual requirements are implemented in the software design and includes appropriate linkages to predefined verification testing.

f. Traceability Analysis/Matrix: Paige.AI Prostate provided traceability between all documents including the SRS, SDS, and subsequent verification and validation. Hazard mitigations are traceable throughout all documents.

g. Software Development Environment: Paige outlined the software development environment and the processes/procedures used for medical device software development. The content is consistent with expected quality system norms.

h. Verification and Validation Testing: The validation and system level verifications procedures are based upon the requirements with clearly defined test procedures and pass/fail criteria. All tests passed. Unit level test procedures, actual, and expected results are included for all design specifications.

i. Revision Level History: Version (v) 2.1.501 was released prior to its use in all the performance studies, including analytical (standalone and precision) and clinical reader study. Software v2.1.501 will remain locked for use with the authorized device and will not be continually trained and improved with each cohort analyzed in clinical practice, after marketing authorization.

j. Unresolved Anomalies: All identified anomalies were resolved prior to verification and validation of the software. There are no unresolved anomalies.

k. Cybersecurity: The cybersecurity documentation is consistent with the recommendations for information that should be included in premarket submissions outlined in the FDA guidance document "Content of Premarket Submissions for Management of Cybersecurity in Medical Devices: Guidance for Industry and Food and Drug Administration Staff" (issued October 2, 2014). Information related to cybersecurity reviewed included:

- Hazard analysis related to cybersecurity risks,
- Traceability documentation linking cybersecurity controls to risks considered,
- Summary plan for validating software updates and patches throughout the lifecycle of the medical device,
- Summary describing controls in place to ensure that the medical device will maintain its integrity, and
- Device instructions for use and product specifications related to recommended cybersecurity controls appropriate for the intended use of the device.

M. Performance Characteristics

1.   Analytical Performance
The sponsor provided data from the following two studies to support the analytical performance of the device:
>   a. Algorithm Localization (X,Y Coordinate) and Accuracy Study
>   b. Precision Study

a. Algorithm Localization (X,Y Coordinate) and Accuracy Study:
The performance of Paige Prostate in identifying digital histopathology images of prostate needle biopsies that are suspicious for cancer and localizing one specific focus (X,Y) coordinate) with the highest suspicion for cancer was evaluated. The (X,Y) coordinates identified by Paige Prostate were evaluated against manual annotations of regions drawn by 3 study pathologists who were blinded to the Paige Prostate results. These study pathologists did not participate in the clinical reader study.

The study sample set originally consisted of 847 scanned digital WSIs of prostate needle biopsy slides (353 cancer and 494 benign) stained with hematoxylin and eosin (H&E). The scanned images were obtained using the previously FDA-cleared Phillips UFS scanner. Out of 847 WSIs, 42 WSIs with cancer and 77 WSIs that were benign did not represent unique patients, i.e., WSIs from multiple different cases but from the same patient. In order to avoid any bias due to case-level overlap in slides, only unique patient level cases were used in the data analysis, i.e., all slides were unique at patient-level compared to the development dataset. Therefore, the final sample set consisted of 728 WSIs (311 WSIs from cancer slides and 417 WSIs from benign slides). There were 3 study pathologists that annotated the image crops as described below in the localization assessment procedure section.

The distribution of the slide images by diagnosis, source of slides and race is provided in Table 3 below.

Table 3: Distribution of case categories in algorithm localization and accuracy study

Characteristic	Cancer (N=311)	Benign (N=417)
Case Category		
ASAP[a]	**11 (3.5%)**	NA
Atrophy Present	0 (0.0%)	**17 (4.1%)**
High-Grade Prostate intraepithelial neoplasia (PIN) Present	See Cancer Category below	**20 (4.8%)**
Treated: tissue with treatment related changes	2 (0.6%)	**14 (3.4%)**
Cancer: tumor size larger than 0.5mm[b]	**153 (49.2%)**	NA
PIN Present	8 (2.6%)	NA
Cancer: tumor size equal or less than 0.5mm[c]	**147 (47.3%)**	NA
PIN Present	6 (1.9%)	NA
Benign (without Atrophy, PIN and Treated)	NA	**366 (87.8%)**
Source of Slides		
Internal Site[*]	136 (43.7%)	183 (43.9%)
External Sites[**]	175 (56.3%)	234 (56.1%)
Race		
Asian-Far East/Indian Subcontinent	11 (3.5%)	11 (2.6%)
Black or African American	26 (8.4%)	32 (7.7%)
Native Hawaiian or Pacific Islander	1 (0.3%)	0 (0.0%)
White	251 (80.7%)	347 (83.2%)
Other	13 (4.2%)	8 (1.9%)
Unknown race	9 (2.9%)	19 (4.6%)

[a] Atypical small acinar proliferation (ASAP) represents suspicious glands without adequate histologic atypia for a definitive diagnosis of prostate adenocarcinoma. However, they were included in the "cancer" category.
[b] Consecutive tumors
[c] Challenging tumors with minimal tumor burden
[*] Internal site located in US
[**] External sites include 217 different sites located throughout the world (including US)
NA: Not Applicable

The study set consisted of deidentified WSIs from:
- Consecutive prostate cancer slides from internal site
- Challenging cancer slides (slides with ≤0.5mm tumor) from internal site
- Consecutive cancer slides submitted from external sites
- Challenging cancer slides submitted from external sites
- Benign slides from consecutive prostate biopsy cases from internal site
- Consecutive benign slides submitted from external sites.External sites refer to a prostate biopsy case (slides) prepared by an external site and submitted to the internal site for expert consultation purposes that were subsequently read by the internal site pathologists.

For consecutive cancer cases, one slide with minimal tumor volume was selected per case per patient. For challenging cancer cases, slides with ≤0.5mm tumor were selected. The dataset was

enriched with 50% challenging cancer slides, which were defined as slides with minimal tumor burden (≤0.5mm). Benign parts were from cases that included cancer parts and represented unique patients in the dataset.

Exclusion Criterion: Slides used during development (algorithm training, tuning and testing) of the Paige Prostate were not used in this study.

Study set of slide images included representation from various races. Table 3 above shows the distribution of slides by race.

Slide-Level Cancer Ground Truth Determination: The synoptic pathology diagnostic reports from the internal site were used to generate the ground truth label for each slide as either cancer or no cancer.

Localization Ground Truth: The (X,Y) coordinates identified by Paige Prostate were evaluated against manual annotations of regions drawn by pathologists that were blinded to the device results. Localization ground truths were determined by 3 US board-certified pathologists (2 completed anatomic pathology fellowship and 1 sub-specialized genitourinary pathologist) annotated image patches.

Localization Assessment Procedure:
1. Images used in this study were generated by scanning slides with a single Philips Ultra Fast scanner.
2. Crops (regions) were generated from:
    - WSIs in which Paige Prostate predicted suspicion for cancer and the ground truth was cancer, and
    - WSIs in which Paige Prostate predicted no suspicion for cancer and the ground truth was no cancer. These slides (20% of the entire set)were included in the study for unbiased estimation of accuracy, but were not considered in the final analysis.
3. All image crops (mix of cancer and benign) of WSIs were reviewed by all 3 study pathologists independently and annotations were provided for image crops that were identified as having cancer. The study pathologists were blinded to each other's annotations and to the results provided by Paige Prostate, during their assessments.
5. All Pathologists were provided the following instructions before annotating the crops:
    a. The drawn boundary of the annotation must be reasonably tight such that it will be of minimal size to enclose the cancerous regions.
    b. Benign cells can be mixed in with cancerous cells since the purpose of the (X,Y) coordinate is to draw the Pathologist's attention to a focus in the region of interest.
6. The union of annotations between at least 2 of the 3 annotating pathologists was used as the localization ground truth.

Primary Endpoints: The following definitions were used to classify the algorithm outcome based on the type of slide:
• *True Positive (TP)*: the algorithm correctly classified that a slide is suspicious for cancer and the (X,Y) coordinate is within a prespecified distance of annotated pixels.
• *True Negative (TN):* the algorithm correctly classified that the slide is not suspicious for cancer.

- *False Positive (FP):* the algorithm incorrectly classified that a slide is suspicious for cancer.
- *False Negative (FN):* the algorithm incorrectly classified that a slide is not suspicious for cancer or the algorithm correctly classified that a slide is suspicious for cancer but the (X,Y) coordinate does not correctly identify a region suspicious for cancer.

Sensitivity and specificity along with 95% two-sided confidence intervals are calculated.

Results:

Sensitivity is the percentage of true positive out of the total number of positive slides.
Specificity is the percentage of true negative slides out of the total number of negative slides.
Overall sensitivity/specificity is as shown in **Table 4** below:

Table 4: Algorithm localization and accuracy study: overall sensitivity and specificity

Measure	Estimate	Counts	95% CI*
Sensitivity	94.5%	294/311	91.4%; 96.6%
Specificity	94.0%	392/417	91.3%; 95.9%

*Confidence intervals are calculated by a score method

The false negative localization errors are provided in Table 5 below.

Table 5: Paige Prostate results where localization errors are considered false negatives

Ground Truth: 311 Cancer	**Ground Truth:** 417 Benign
Paige Prostate Results: 294 True Positives (slide identified correctly as cancer and with correct localization) 17 False Negative: 9 False Negatives (slide identified as benign) 8 False Negatives (slide identified correctly as cancer but incorrect localization)	**Paige Prostate Results:** 392 True Negatives 25 False Positives
Total TP: 294 out of 311	**Total TN:** 392 out of 417
Total FN: 17 out of 311	**Total FP:** 25 out of 417

Sensitivity, specificity and localization accuracy stratified by source of slides:
Paige Prostate sensitivity and specificity were stratified by source of slides [internal site and external sites (217 sites)] as shown in tables 6 and 7, respectively.

Table 6: Summary of Algorithm Localization and Accuracy Study: Sensitivity, Stratified by Source of Slides

Source of Slides	Sensitivity Estimate	Counts	95% CI*
Internal site	94.1%	128/136	88.8%, 97.0%
External sites	94.9%	166/175	90.5%, 97.3%

*Confidence intervals are calculated by a score method

Among 17 slides with false negative results, 8 slides were from the internal site and 9 slides were from the external sites. Out of the 8 false negative slides from the internal site, 5 cancer slides were incorrectly classified as benign slides and 3 cancer slides had incorrect localization but classified as cancer slides. Out of the 9 false negatives from the external sites, 4 cancer slides were incorrectly classified as benign slides and 5 cancer slides had incorrect localization but classified as cancer slides.

Table 7: Summary of Algorithm Accuracy Study:
Specificity[**], Stratified by Source of Slides

Source of Slides	Specificity Estimate	Counts	95% CI*
Internal site	96.7%	177/183	93.0%, 98.5%
External sites	91.9%	215/234	87.7%, 94.7%

*Confidence intervals are calculated by a score method
** No device output [coordinate (X,Y)] for slides assessed as Benign by Paige Prostate

There were six false positive slides from the internal site and 19 from external sites. These false positive slides were attributed to: (1) scan blur in some slide image areas due to variation in tissue thickness based on different slide preparation techniques, (2) pen marks, and (3) other artifacts.

Differences in Paige Prostate's performance in correctly classifying benign slides as benign, between the slides sourced from the internal site and from 217 external sites, may be attributed to the difference in the diversity of the patient population and the difference in slide preparation techniques between these 217 external sites which include sites outside of the US. In tables above, localization was considered as correct if (X,Y) coordinate was within a prespecified distance to annotated pixels.

b. Precision Study:

The precision of Paige Prostate device in identifying digital histopathology images of prostate needle biopsies suspicious for cancer and its ability to localize a focus within a prespecified distance threshold with the highest suspicion for cancer was evaluated. Slides used in the precision studies were not slides used during development (algorithm training, tuning and testing) of the Paige Prostate device.
Precision was evaluated as follows:
- Within-scanner precision study: Glass slides were scanned three different times (3 repetitions) using one scanner and one operator.
- Reproducibility study (between-scanner and between-operator variability): Glass slides were scanned once with three different scanners, installed at different locations at the internal site and acquired by three different operators (one operator per scanner).
- Localization precision study: Glass slides were scanned 5 times (three scanners and three operators: 3 scans with scanner 1 and operator 1, 1 scan with scanner 2 and operator 2 and 1 scan with scanner 3 and operator 3).

The within-scanner precision and reproducibility studies consisted of 35 WSIs prostate cancer slides and 36 WSIs benign prostate slides from unique patients.

Table 8: Distribution of case categories in precision study

Characteristic	Cancer (N=35)	Benign (N=36)
Case Category		
ASAP[a]	0 (0.0%)	NA
Atrophy Present	0 (0.0%)	0 (0.0%)
High-Grade Prostate intraepithelial neoplasia (PIN) Present	See Cancer Category below	**2 (5.6%)**
Treated: tissue with treatment related changes	1 (2.9%)	**1 (2.8%)**
Cancer: tumor size larger than 0.5 mm[b]	**18 (51.4%)**	NA
PIN Present	1 (2.9%)	NA
Cancer: tumor size equal or less than 0.5mm[c]	**17 (45.7%)**	NA
PIN Present	1 (2.9%)	NA
Benign (without Atrophy, PIN and Treated)	NA	**33 (91.7%)**
Source of Slides		
Internal site[*]	16 (45.7%)	16 (44.4%)
External sites[**]	19 (54.3%)	20 (55.6%)
Race		
Asian-Far East/Indian Subcontinent	0 (0.0%)	1 (2.8%)
Black or African American	6 (17.6%)	2 (5.6%)
Native Hawaiian or Pacific Islander	0 (0.0%)	0 (0.0%)
White	27 (79.4%)	30 (83.3%)
Other	1 (2.9%)	0 (0.0%)
Unknown race	1 (2.9%)	3 (8.3%)

[a] Atypical small acinar proliferation (ASAP) represents suspicious glands without adequate histologic atypia for a definitive diagnosis of prostate adenocarcinoma
[b] Consecutive tumors
[c] Challenging tumors with minimal tumor burden
[*] Internal site located in US
[**] External sites include 217 different sites
NA: Not Applicable

Slide level classification ground truth: The synoptic diagnostic reports from the internal site were used to generate the ground truth for each WSI as either cancer or no cancer.

Within-scanner precision study results:
The agreement rate between three assessments from a single scanner and one operator was evaluated for 35 prostate cancer slides and 36 benign prostate slides as shown in Table 9 and the agreement rates for repetitions are shown in Table 10.

Table 9: Within-scanner precision: positive and negative agreements

Operator - Scanner	Agreement to Ground Truth by Slide Type			
	Cancer Images		Benign Images	
	PPA, Agreed % (n/N)	95% CI	NPA, Agreed % (n/N)	95% CI
1st Operator / Scanner 1 Repetition 1	100% (35/35)	90.0%, 100%	97.2% (35/36)	85.5%, 99.9%
1st Operator / Scanner 1 Repetition 2	100% (35/35)	90.0%, 100%	94.4% (34/36)	81.3%, 99.3%
1st Operator / Scanner 1 Repetition 3	97.1% (34/35)	85.1%, 99.9%	91.7% (33/36)	77.5%, 98.2%
Overall Average	**99.0%**	**97.1%, 100.0%***	**94.4%**	**88.9%, 99.1%***

*Confidence interval is calculated by bootstrap

Table 10: Within-scanner precision: agreement rates for scan repetitions

Type of Slide images	Total number of slide images	Percent of slide images with correct results with all 3 scan repetitions	Percent of correct results with 3 scan repetitions from all slide images, 95% CI*
Cancer	35	97.1% (34/35)	**99.0% (104/105),** 95%CI: (94.8%; 99.8%)
Benign	36	88.9% (32/36)	**94.4% (102/108),** 95%CI: (88.4%; 97.4%)

*Confidence intervals are calculated by a score method

For cancer slide images, probability that result is Cancer with repetitions with the same scanner and same operator is 99.0% with 95%CI: (94.8%; 99.8%).
For benign slide images, probability that result is Benign with repetitions with the same scanner and same operator is 94.4% with 95%CI (88.4%; 97.4%).

Reproducibility (between-scanner and between-operator variability) study results
The agreement rate between three assessments from three scanners (1 repetition per scanner) and 3 operators was evaluated for 35 prostate cancer slides and 36 benign prostate slides (Table 11). Also, the agreement rates between scanners/operators are shown in Table 12.

Table 11: Reproducibility (between-scanner and between-operator) precision: positive and negative agreement

Operator - Scanner	Agreement to Ground Truth by Slide Type	
	Cancer Slides	Benign Slides

	PPA Agreed % (n/N)	95% CI	NPA, Agreed % (n/N)	95% CI[*]
1st Operator / Scanner 1	100% (35/35)	90.1%; 100%	97.2% (35/36)	85.8%, 99.5%
2nd Operator / Scanner 2	100% (35/35)	90.1%, 100%	94.4% (34/36)	81.9%, 98.5%
3rd Operator / Scanner 3	100% (35/35)	90.1%, 100%	88.9% (32/36)	74.7%, 95.6%
Overall Average	**100%**	N/A	**93.5%**	88.0%, 98.1%[*]

*Confidence interval is calculated by bootstrap

Table 12: Reproducibility (between-scanner and between-operator): agreement rates for repetitions with 3 different scanners/operators

Type of slide images	Total number of slide images	Percent of slide images with correct results with all 3 scans	Percent of correct results with 3 scans from all slides images, 95%CI*
Cancer	35	100% (35/35)	**100% (105/105)**, 95%CI: (96.5%; 100%)
Benign	36	83.3% (30/36)	**93.5% (101/108)** 95%CI: (87.2%; 96.8%)

*Confidence intervals are calculated by a score method

For cancer slide images, probability that result is Cancer with scans with different scanners and different operators is 100% with 95%CI: (96.5%; 100%).
For benign slide images, probability that result is Benign with scans with different scanners and different operators is 93.5% with 95%CI: (87.2%; 96.8%).

Localization precision study

Localization precision study was conducted by randomly selecting 19 cancer and 4 benign WSIs out of the 35 cancer and 36 benign WSIs from the within-scanner precision and reproducibility studies. There were 95 crops from 19 cancer slides. Two crops were excluded because 1 crop was too close to the edge of the slide and 1 crop was not annotated by the pathologist. Analysis of the localization precision study data was performed only with cancer slides and is summarized in Table 13 below.

Table 13: Localization precision study results

| Operator - Scanner | N | Localization Cases | | %Correct Location | 95%CI** |
		Location Correct	Location Incorrect		
3 scans of Operator 1/Scanner 1	57	56	1	98.2% (56/57)	(90.7%; 99.7%)
3 scanners (1 scan per scanner) Operator 1/Scanner 1 Operator 2/Scanner 2 Operator 3/Scanner 3	55*	53	2	96.4% (53/55)	(87.7%; 99.0%)

*2 crops were not included in the calculations because 1 crop was not annotated by pathologists for ground truth and 1 crop was no visible tissue on the crop due to being too close to edge of the slide.
**Confidence intervals are calculated by a score method

2. Clinical Study:

The Paige.AI provided data from a clinical study to support the clinical performance of the Paige Prostate device. The study evaluated pathologists' slide-level diagnostic performance in identifying cancer in WSIs of prostate needle biopsy slides, with and without the assistance of the Paige Prostate device.

Clinical Study Design:

Paige.AI conducted a retrospective clinical study to evaluate the effectiveness of Paige Prostate in improving the diagnostic accuracy of pathologists. The clinical study included 527 WSIs from 171 prostate cancer slides and 356 benign slides from prostate biopsies and 16 pathologist readers [2 genitourinary (GU) subspecialists and 14 general specialists, with a median of 6 years of experience (range: 2-34 years)]. Pathologists completed the study reads in their current (remote or on-site office) work environment and study staff conducted remote monitoring for study sites. Three of the pathologists (generalists) completed the study on-site at a Clinical Laboratory Improvement Amendments (CLIA) -certified pathology laboratory and 13 pathologists (11 generalists and 2 subspecialists) completed the study remotely from home (offices confirmed to meet remote work environment requirements set by Centers for Medicare and Medicaid Services (CMS); Ref: QSO-20-21-CLIA[1].

Each pathologist was provided with an FDA-cleared monitor (Philips PP27QHD). Pathologists used their own computers to connect to the monitor and complete their reads. After pathologists were trained, they provided proof that the monitor was installed appropriately prior to starting the study.

---

[1] https://www.cms.gov/medicareprovider-enrollment-and-certificationsurveycertificationgeninfopolicy-and-memos-states-and/clinical-laboratory-improvement-amendments-clia-laboratory-guidance-during-covid-19-public-health

Results were displayed and pathologists conducted their assessments with an FDA-cleared whole slide image viewer (FullFocus™).

Each pathologist performed the following procedural steps on an individual basis.

    a. Pathologists were trained to use the FDA-cleared digital pathology image review system and the Paige Prostate device.

    b. WSIs of scanned prostate biopsy slides were displayed on an FDA-cleared monitor to each pathologist one at a time in a randomized order.

    c. The pathologists completed an unassisted read directly followed by an assisted read for every WSI.

- Unassisted Read: The pathologists reviewed the image, without Paige Prostate assistance, with the FDA-cleared pathology viewer.
- Assisted Read: The pathologists reviewed the image with Paige Prostate result coordinate (X,Y) overlaid on the same image. The result included:
- Paige Prostate slide level binary classification: suspicious for cancer or not suspicious for cancer.
- Coordinate (X, Y): If the slide was predicted to be suspicious for cancer, the algorithm identifies a coordinate (X, Y) of the region on the slide as having the highest likelihood for harboring cancer.

    d. Pathologists were instructed that they could choose to "defer for more information" during the study if they were unable to render a definitive diagnosis as either "cancer" or "no cancer."

    e. Classifications for each image were made without information from immunohistochemistry (IHC) stains. The pathologists performed a complete review of each WSI and recorded their classifications directly into the electronic database case report form (CRF):

(i) The pathologists classified each slide as:

- cancer,
- no cancer, or
- defer for more information.

(ii) For the deferral classification, the pathologists selected why they would defer, from the following options, which were all methods currently used in clinical practice today when a pathologist is not able to determine a diagnosis from an H&E slide:

- Additional stains
- Additional levels
- Seek another opinion
- Other: If the pathologists selected "Other" they would elaborate via a free text box in the CRF.

For each slide, sixteen pathologists completed an unassisted read directly followed by an assisted read with Paige Prostate for every image.

Study Sample Characteristics:

    a. The sample set originally consisted of 610 whole slide images of prostate needle biopsy slides (190 cancer and 420 benign) stained with hematoxylin and eosin (H&E) that were scanned using a single unit of the FDA-cleared Phillips Ultra Fast Scanner with Philips Image Management System (IMS) to upload the scanned images. Out of 610 WSIs, 19 WSIs from prostate cancer slides and 64 WSIs from benign slides from

prostate biopsies did not represent unique patients, i.e., WSIs were from multiple different cases, but from the same patient. To avoid any bias due to case-level overlap in slides, only unique patient level cases were considered for the final data analysis. Thus, the final sample set consisted of 527 WSIs from 171 prostate cancer slides and 356 benign slides from prostate biopsies.

b. Out of the 527 WSIs, 44.15% of the images were from cases prepared, reviewed, diagnosed, and digitized at the internal site, and 55.85% of the images were from cases prepared at 156 different external sites but reviewed, diagnosed, and digitized at the internal site.

c. No slide used during development of the Paige Prostate were used for this study.

d. Dataset was enriched with 50% challenging cancer slides, which were defined as slides with minimal tumor burden. Challenging cancer cases contained at least one slide with less than or equal to 0.5mm tumor; one with minimal tumor was selected.

e. Benign parts could come from a case that includes cancer parts, as long as the selected part represented a unique patient in the dataset. Benign parts could also come from a case without any cancer parts.

f. Slide-level cancer/benign ground truths were determined by reviewing the original diagnostic synoptic reports.

Table 14: Distribution of case categories in clinical study

Characteristic	Cancer (N=171)	Benign (N=356)
Case Category		
ASAP[a]	**8 (4.7%)**	NA
Atrophy Present	0 (0.0%)	**3 (0.8%)**
High-Grade PIN Present	See Cancer Category below	**18 (5.1%)**
Cancer: tumor size larger than 0.5 mm[b]	**84 (49.1%)**	NA
PIN Present	4 (2.3%)	NA
Cancer: tumor size equal or less than 0.5mm[c]	**79 (45.6%)**	NA
PIN Present	4 (2.3%)	NA
Treated: tissue with treatment related changes	1 (0.6%)	**12 (3.4%)**
Benign (without Atrophy, PIN and Treated)	NA	**323 (90.7%)**
Source of Slides		
Internal site*	77 (45.0%)	154 (43.3%)
External sites**	94 (55.0%)	202 (56.7%)
Race		
Asian-Far East/Indian Subcontinent	8 (4.7%)	11 (3.1%)
Black or African American	14 (8.2%)	28 (7.9%)
Native Hawaiian or Pacific Islander	1 (0.6%)	0 (0.0%)
White	135 (78.9%)	302 (84.8%)
Other	10 (5.8%)	7 (2.0%)

Unknown race	3 (1.8%)	8 (2.2%)

[a] Atypical small acinar proliferation (ASAP) represents suspicious glands without adequate histologic atypia for a definitive diagnosis of prostate adenocarcinoma
[b] Consecutive tumors
[c] Challenging tumors with minimal tumor burden
*Internal site located in US
** External sites include 156 different sites
NA: Not Applicable

Exclusion Criteria:

Slides from the following categories were excluded:
a. Any slide used during development (algorithm training, tuning and testing) of the Paige Prostate device.
b. Any slide with scanning quality control issues, as determined by a pathologist, such as (included blur/out of focus areas, folded tissue, scanning artifacts or other artifacts that compromised the ability to interpret the findings on the tissue.
c. Any slide that was not H&E-stained.

Pathologist Qualifications:

a. US Board certified anatomic pathologists from six sites outside of the internal site including community, academic, and private practices were included in the study. The study included 14 general pathologists with greater than one year of experience, and 2 subspecialized genitourinary pathologists. Study pathologists underwent training for FDA-cleared digital pathology system.

Clinical Performance Measures:

Diagnoses of 'deferred' or 'cancer' was considered as 'positive' and diagnosis 'benign' was considered as 'negative'. Sensitivity and specificity along with 95% confidence intervals were provided.

Study Results:

The clinical study demonstrated improvements in sensitivity and small differences in specificities between assisted and unassisted reads.

Table 15: Summary of sensitivity and specificity by pathologist (specialist and generalist) and location (on-site and remote) with positive=cancer or defer and negative=benign

Pathologist	Pathologist Specialty	Pathologist Setting	Sensitivity (N=171 Cancer)			Specificity (N=356 Benign)		
			Assisted % (n)	Unassisted % (n)	Difference 95%CI	Assisted % (n)	Unassisted % (n)	Difference 95%CI
1	Generalist	Remote	95.9% (164)	86.5% (148)	9.4% (4.7%; 14.8%)	94.1% (335)	93.8% (334)	0.3% (-2.3%; 2.9%)
2	Generalist	Remote	98.8% (169)	89.5% (153)	9.4% (5.1%; 14.7%)	91.9% (327)	92.4% (329)	-0.6% (-2.9%; 1.7%)
3	Generalist	Remote	98.2% (168)	95.3% (163)	2.9% (-0.5%; 7.0%)	92.7% (330)	90.4% (322)	2.2% (-0.2%; 4.9%)
4	Generalist	Remote	93.6% (160)	87.7% (150)	5.8% (2.2%; 10.3%)	77.2% (275)	78.4% (279)	-1.1% (-2.5%; 0.2%)
5	Generalist	On-Site	97.1% (166)	93.6% (160)	3.5% (-0.8%; 8.2%)	96.1% (342)	94.9% (338)	1.1% (-0.7%; 3.1%)
6	Generalist	On-Site	98.2% (168)	93.6% (160)	4.7% (1.4%; 9.0%)	88.8% (316)	88.2% (314)	0.6% (-1.6%; 2.7%)
7	Generalist	On-Site	97.7% (167)	86.0% (147)	11.7% (6.7%; 17.6%)	91.0% (324)	89.6% (319)	1.4% (-1.3%; 4.2%)
8	Generalist	Remote	97.1% (166)	73.1% (125)	24.0% (17.4%; 31.0%)	94.9% (338)	81.2% (289)	13.8% (9.5%; 18.2%)
9	Generalist	Remote	94.2% (161)	74.9% (128)	19.3% (13.5%; 25.7%)	82.3% (293)	81.2% (289)	1.1% (-1.2%; 3.5%)
10	Generalist	Remote	99.4% (170)	96.5% (165)	2.9% (0.0%; 6.8%)	83.1% (296)	83.4% (297)	-0.3% (-2.0%; 1.4%)
11	Generalist	Remote	96.5% (165)	90.6% (155)	5.8% (1.9%; 10.7%)	91.6% (326)	89.3% (318)	2.2% (-0.2%; 4.9%)
12	Generalist	Remote	98.2% (168)	95.9% (164)	2.3% (-1.0%; 6.3%)	87.4% (311)	88.5% (315)	-1.1% (-2.7%; 0.3%)
13	Generalist	Remote	98.2% (168)	92.4% (158)	5.8% (2.3%; 10.5%)	81.5% (290)	82.9% (295)	-1.4% (-3.3%; 0.4%)

Pathologist	Pathologist Specialty	Pathologist Setting	Sensitivity (N=171 Cancer)			Specificity (N=356 Benign)		
			Assisted % (n)	Unassisted % (n)	Difference 95%CI	Assisted % (n)	Unassisted % (n)	Difference 95%CI
14	Generalist	Remote	95.9% (164)	91.8% (157)	4.1% (0.4%; 8.4%)	90.4% (322)	89.3% (318)	1.1% (-1.6%; 3.9%)
15	Specialist	Remote	94.2% (161)	93.6% (160)	0.6% (-2.3%; 3.6%)	94.9% (338)	95.8% (341)	-0.8% (-2.4%; 0.5%)
16	Specialist	Remote	95.9% (164)	90.6% (155)	5.3% (1.4%; 9.9%)	94.1% (335)	95.8% (341)	-1.7% (-3.9%; 0.2%)
Combined	Generalist or Specialist	On-site or Remote	96.8% (165.6)	89.5% (153.0)	7.3% (3.9%; 11.4%)	89.5% (318.6)	88.4% (314.7)	1.1% (-0.7%; 3.4%)

*Confidence intervals for differences in sensitivities and specificities are calculated by a score method for an individual pathologist and by bootstrap for combined data (averaged over all pathologists).

Details about reduction of false negative slides and false positive slide numbers on average (16 pathologists) are presented in **table 16** and **17**, respectively.

Table 16: Assessment of cancer slides, (171 slides) on average (16 pathologists)

		Classification for unassisted read			
		Cancer	Deferred	No Cancer	Total
Classification for assisted read	Cancer	128.56 (75.2%)	5.75 (3.4%)	2.75 (1.6%)	137.06 (80.2%)
	Deferred	0.44 (0.3%)	17.50 (10.2%)	10.56 (6.2%)	28.50 (16.67%)
	No cancer	0.25 (0.1%)	0.50 (0.3%)	4.69 (2.7%)	5.44 (3.18%)
	Total	129.25 (75.58%)	23.75 (13.89%)	18.00 (10.53%)	171 (100%)

Numbers in grey colors are numbers of slide images with the same classification in assisted and unassisted reads. Numbers in green color, 2.75 (1.6%) and 10.56 (6.2%), present a reduction in the number of false negative results for the cancer slide images because of use of the Paige Prostate device. Numbers in orange color, 0.25 (0.1%) and 0.50 (0.3%), present an increase in the number of false negative results because these cancer slide images had "No cancer" for assisted reads but had "Cancer" or "Deferred" for unassisted reads. Overall reduction in the number of false negative slides was 12.56 slides [=(2.75+10.56)-(0.25+0.50)] what is 7.34% (=12.56/171). Reduction in the number of false negative slides of 7.3% with 95% CI: (3.9%; 11.4%) that was statistically significant.

Table 17: Assessment of benign slides, (356 slides) on average (16 pathologists)

		Classification for unassisted read			
		Cancer	Deferred	No Cancer	Total
Classification for assisted read	Cancer	4.31 (1.2%)	1.12 (0.3%)	0.81 (0.2%)	6.25 (1.76%)
	Deferred	3.19 (0.9%)	22.75 (6.4%)	5.19 (1.5%)	31.12 (8.74%)
	No cancer	0.69 (0.2%)	9.06 (2.5%)	308.87 (86.8%)	318.62 (89.5%)
	Total	8.19 (2.3%)	32.94 (9.2%)	314.87 (88.45%)	356 (100%)

Numbers in grey colors are numbers of slide images with the same classification in assisted and unassisted reads. Numbers in green color, 0.69 (0.2%) and 9.06 (2.5%), present a reduction in the number of false positive results for the benign slide images because of use of the Paige Prostate device. Numbers in orange color, 0.81 (0.2%) and 5.19 (1.5%), present an increase in the number of false positive results because these benign slide images had "Cancer" assisted reads (0.2%) or "Deferred" assisted reads (1.5%) but had "No Cancer" for unassisted reads. Overall difference in the number of false positive slide images was 3.75 slides [=(0.69+9.06)-(0.81+5.19)] what is 1.05% (=3.75/356). Difference in the number of false positives slides of 1.1% with 95% CI: (-0.7%; 3.4%) was not statistically significant.

Analysis of sensitivity and specificity by pathologist specialty and location is presented in **Table 18** and **Figure 4** depicts the improvement in sensitivities between assisted and unassisted reads.

Table 18: Analysis of sensitivity and specificity by pathologist (specialist and generalist) and location (on-site and remote):

Pathologist Specialty	Pathologist Setting	Number of Pathologists	Improvement in Sensitivity			Difference in Specificity		
			Average	Median	Range	Average	Median	Range
Generalist	Remote	11	8.3%	5.8%	(2.3%; 24.0%)	1.5%	0.3%	(-1.4%; 0.8%)
Generalist	On-site	3	6.6%	4.7%	(3.5%; 11.7%)	1.0%	1.1	(0.6; 1.4)
Specialist	Remote	2	2.9%	2.9%	(0.6%; 5.3%)	-1.3%	-1.3%	(-1.7%; -0.8%)
Generalist	Remote or On-site	14	8.0%	5.8%	(2.3%; 24.0%)	1.4%	1.1%	(-1.4%; 13.8)
Specialist	Remote	2	2.9%	2.9%	(0.6%; 5.3%)	-1.3%	-1.3%	(-1.7%; -0.8%)

	16	7.3%	5.6%	(0.6%: 24.0%)	1.1%	0.4%	(-1.7%; 13.8%)
Combined							

For combined data, an average improvement in sensitivity was 7.3% with 95%CI: (3.9%; 11.4%) (statistically significant), a median value was 5.6% and range was from 0.6% to 24.0%. An average difference in specificity was 1.1% with 95% CI: (-0.7%; 3.4% (not statistically significant), a median value was 0.4% and range was from -1.7% to 13.8%.
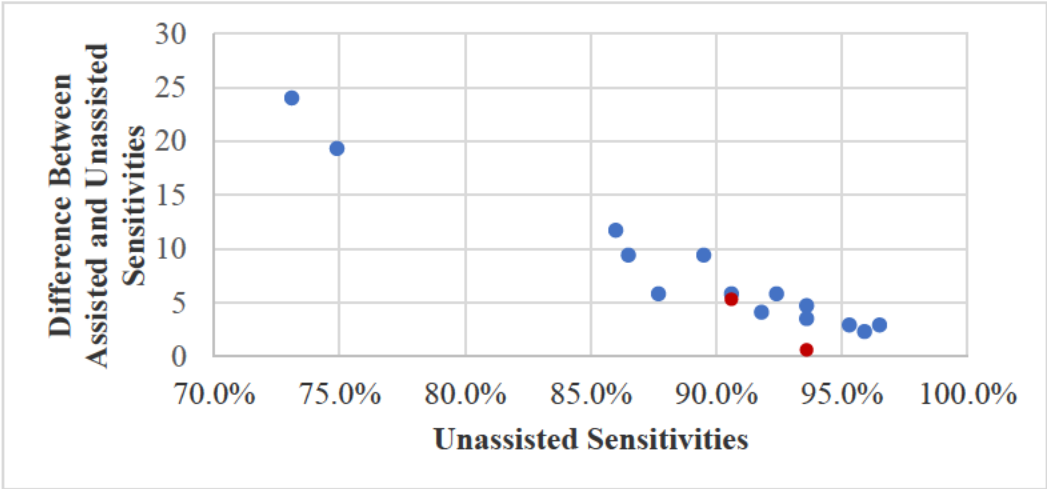

Figure 4: Improvement in assisted sensitivities vs Unassisted Sensitivity: Points in orange represent the improvement in sensitivities among specialists and points in blue represent the improvement in sensitivities among generalists.

It should be noted that:

- The pathologists' reviews in the clinical study were based on an initial interpretation of the slide images. In clinical practice, additional special studies are performed when there is any doubt about the diagnosis. In the clinical study, the initial interpretations were used and special studies were not permitted because an objective of the clinical study was to evaluate an improvement in accuracy of the prostate slide images using the Paige Prostate.

- The study analysis was on a per-biopsy basis, not on a per-patient basis. In a typical patient undergoing prostatic needle biopsy for evaluation of possible cancer, multiple core biopsies are obtained (often 12-14 biopsies), and many patients with prostate cancer have cancer in multiple biopsies.

Based on these two constraints of the clinical study, the expected benefit of the use of the Paige device on the final diagnosis in practice would likely be substantially lower than 7.3% when evaluated on a per-patient basis.

N. Labeling

The labeling supports the decision to grant the De Novo request for this device.

O. **Patient perspectives**

This submission did not include specific information on patient perspectives for this device.

P. Identified Risks to Health and Identified Mitigations

Identified Risks to Health	Mitigation Measures
False negative classification (loss of accuracy)	Certain design verification and validation, including certain device descriptions, certain analytical studies, and clinical studies. Certain labeling information, including certain device descriptions, certain performance information, and certain limitations.
False positive classification (loss of accuracy)	Certain design verification and validation, including certain device descriptions, certain analytical studies, and clinical studies. Certain labeling information, including certain device descriptions, certain performance information, and certain limitations.

Q. Benefit-Risk Determination

Summary of Benefits

The use of this device for the proposed IU population according to the proposed instructions for use is expected to benefit a small proportion of men who have undergone prostate biopsy in receiving a correct pathologic diagnosis of that biopsy. Although standard of care is expected to yield the correct diagnosis in the vast majority of such biopsies, there appears to be a small proportion of cases for which a small focus of carcinoma may be overlooked and consequently identified with the use of the device. In the pivotal clinical study, for 171 slides with cancer, the change from "unassisted benign" to "assisted defer" was 6.2%, and the change from "unassisted benign" to "assisted cancer" was 1.6%. Also, for 171 slides with cancer, the change from "unassisted defer" to "assisted benign" was 0.3% and the change from "unassisted cancer" to "assisted benign" was 0.1%. Therefore, in 7.3% of individual biopsy specimens with cancer, there is expected patient benefit in terms of an improvement in sensitivity ((6.2% + 1.6%) - (0.3% + 0.1%)). On average, improvement in specificity was 1.1% (specificity assisted = 89.50% minus specificity unassisted = 88.45%). It should be noted that this analysis is on a per-biopsy basis, not on a per patient basis. Since in a typical patient undergoing prostatic needle biopsy for evaluation of possible cancer, multiple core biopsies are obtained, and many patients with prostate biopsies have cancer in multiple biopsies, this expected benefit in practice would likely be substantially lower than 7.3% when evaluated on a per-patient basis. There is also some limited expected benefit in terms of time savings for the pathologist reviewing these biopsies.

Summary of Risks

The risk of use of this device for the proposed IU population according to the proposed instructions for use is the loss of accuracy leading to an incorrect diagnosis (false positive or false negative). Incorrect diagnosis is clearly harmful. This could be in the form of an incorrect diagnosis of cancer for which the patient may receive unnecessary treatment and psychologically harmful misinformation. An incorrect rendering of a benign diagnosis would likely cause a delay in the treatment of a cancer and would likely in some cases lead to increased morbidity and mortality.

Benefit/Risk Conclusion

Paige Prostate appears to provide a reasonable assurance of safety and effectiveness for diagnostic use by its intended users after taking into consideration the special controls. The clinical and analytical studies have shown that the risk of accuracy loss resulting in a false positive or false negative diagnosis, is minimal relative to the patient safety benefits, including new findings that would contribute to the correct diagnosis. This is contingent on the device being used according to the approved labeling, particularly that the end user must be fully aware of how to interpret and apply the device output.

The potential for false negative and false positive results is mitigated by special controls. Labeling requirements, which include certain device description information as well as certain limitations, ensure that users will employ all appropriate procedures and safeguards as specified, including use of the device as an adjunct rather than as the sole basis of making the diagnosis.
In addition, design verification and validation includes data on software performance as supported by the underlying software design, as well as software algorithm training and validation within the limits of the specified intended use. This also includes analytical validation (including precision studies) and clinical validation (including user validation studies and performance studies) studies.

The probable clinical benefits outweigh the potential risks when the standard of care is followed by qualified users, and appropriate mitigation of the risks is provided for through implementation of and adherence to the special controls. The combination of the general controls and established special controls support the assertion that the probable benefits outweigh the probable risks.

R. Conclusion

The De Novo request is granted, and the device is classified under the following and subject to the special controls identified in the letter granting the De Novo request:

    Product Code: QPN
    Device type: Software algorithm device to assist users in digital pathology
    Class: II
    Regulation: 21 CFR 864.3750