<div align="center">

DE NOVO CLASSIFICATION REQUEST FOR
GI GENIUS

</div>

REGULATORY INFORMATION

FDA identifies this generic type of device as:

> **Gastrointestinal lesion software detection system.** A gastrointestinal lesion software detection system is a computer-assisted detection device used in conjunction with endoscopy for the detection of abnormal lesions in the gastrointestinal tract. This device with advanced software algorithms brings attention to images to aid in the detection of lesions. The device may contain hardware to support interfacing with an endoscope.

> **NEW REGULATION NUMBER:** 21 CFR 876.1520

> **CLASSIFICATION:** Class II

> **PRODUCT CODE:** QNP

BACKGROUND

> **DEVICE NAME:** GI Genius

> **SUBMISSION NUMBER:** DEN200055

> **DATE DE NOVO RECEIVED:** September 8, 2020

> **SPONSOR INFORMATION:**

> > Cosmo Artificial Intelligence - AI, LTD
> > Riverside II, Sir John Rogerson's Quay
> > Dublin, Dublin 2 D02 KV60
> > Ireland

INDICATIONS FOR USE

The GI Genius is indicated as follows:

> The GI Genius System is a computer-assisted reading tool designed to aid endoscopists in detecting colonic mucosal lesions (such as polyps and adenomas) in real time during standard white-light endoscopy examinations of patients undergoing screening and surveillance endoscopic mucosal evaluations. The GI Genius computer-assisted detection device is limited for use with standard white-light endoscopy imaging only. This device is not intended to replace clinical decision making.

> **LIMITATIONS**

The sale, distribution, and use of the GI Genius are restricted to prescription use in accordance with 21 CFR 801.109.

The device is not intended to be used as a stand-alone diagnostic device.

The device is not intended to characterize lesions in a manner that would potentially replace biopsy sampling

The device is not intended to replace clinical decision making.

The device is not intended to be used with equipment that it was not tested against during validation activities.

The device has not been studied in patients with Inflammatory Bowel Disease (IBD), history of CRC, or previous colonic resection. The device performance may be negatively impacted by mucosal irregularities such as background inflammation from certain underlying disease.

PLEASE REFER TO THE LABELING FOR A COMPLETE LIST OF WARNINGS, PRECAUTIONS, AND CONTRAINDICATIONS.

DEVICE DESCRIPTION

The GI GENIUS™ is an artificial intelligence/machine learning (AI/ML) device system comprised of software, hardware, and accessories that is intended for polyp detection during standard white-light colonoscopy. The device system generates a video on the main endoscopy display that contains the original live video together with superimposed markers (in the form of green boxes) that appear when a lesion is detected.
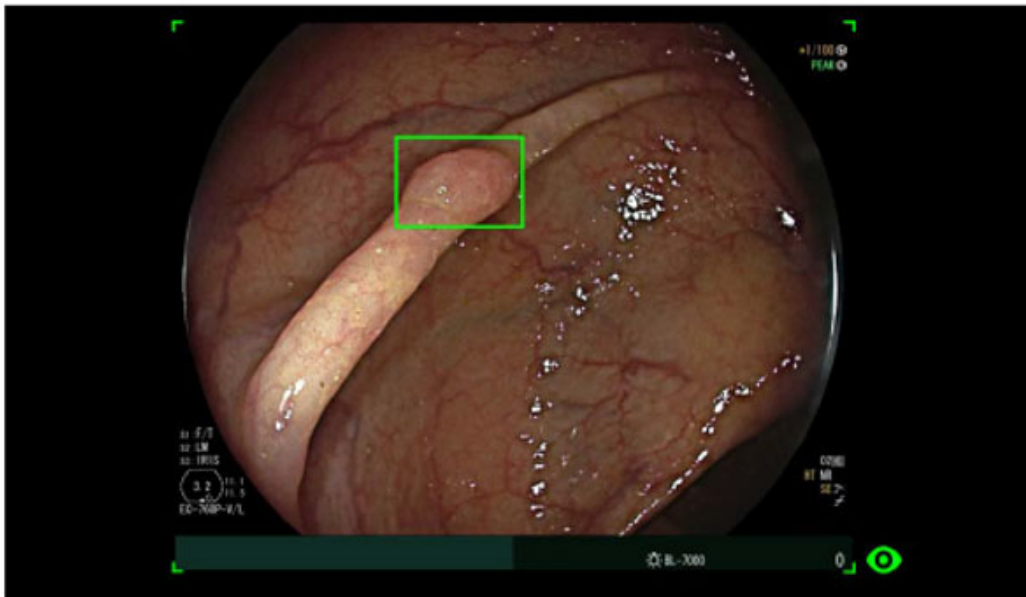

Figure 1. Example of a colonoscopy image in which the device has detected a lesion (green box)

The GI Genius takes the Serial Digital Interface (SDI) output stream from the video endoscope processor as an input and then generates an SDI output stream to the existing monitor/display system containing the original video stream with additional markers superimposed on it. In essence, the system is inserted into the video stream just prior to it being displayed to the user/operator. The technological characteristics of the GI Genius system are described below.

Figure 10-1 – GI GENIUS Compatible Hardware on Video Endoscopy Trolly

Figure 10-2 – GI GENIUS Front View

Figure 10-3 – GI GENIUS Rear View

Figure 2. Images of the GI Genius in relation to compatible hardware

The GI Genius software, operating in the described hardware, is comprised of the major functional software components (modules) described below:

The software's architecture is module-based, with the following modules:

• <u>Main Module:</u> This module starts the main application components and initializes the video acquisition components.

• <u>Video Capture Module:</u> This module handles frame collection, video frame management (e.g., colorspace transformations, cropping, and resizing) and providing the frames to the detection module.

• <u>AI/Detection Module:</u> This module is responsible for identifying potential mucosal lesions. The main component consists of a convoluted neural network.

• <u>Overlay Module:</u> This module generates markers and superimposes them on the endoscopic video stream.

• <u>Application Log Module:</u> This module traces events, such as overlay activation/deactivation and software errors.

• <u>GUI Handler Module:</u> This module generates the menu user interface, in order to allow user actions such as volume regulation, field of view setting and to check the system status.

• <u>Launcher Module:</u> This module provides integrity checks on the files necessary for the correct execution of the software.

The device description included a list of compatible hardware video processors, compatible endoscope characteristics, and the software architecture description. This includes the convolutional neural network (CNN) of the AI/ML algorithm.

SUMMARY OF NONCLINICAL/BENCH STUDIES

Test	Purpose	Method	Acceptance Criteria	Results
Video delay	To assess the time needed by the GI Genius to transmit the original colonoscopy video to the display	Measure the timing difference between the original endoscopic video and the GI Genius output of the same frame	Time is less than or equal to 5.75 milliseconds	(b) (4) microseconds (b) (4) %
Annotation delay	To determine the delay in annotating the video with the annotation box	Timing diagram of the frame capture, AI processing, AI overlay, and frame transmit pipeline	Time is less than or equal to 120 milliseconds	(b) (4) or (b) (4) milliseconds (b) (4 frames)
Video quality integrity test	To assess that the image quality did not degrade from the endoscopic video processor through the GI Genius and then the display	Pixel-wise comparison between the original endoscopic video and the GI Genius video with the marker overlays	There should be no degradation in image quality, meaning that all corresponding pixels (from the endoscopic video processor and from the GI Genius) are identical in the three color channels	The test data show no pixel-level discrepancies except for the pixels overlaid with markers

ELECTROMAGNETIC CAPABILITY (EMC) & ELECTROMAGNETIC SAFETY

The hardware components of GI Genius were tested per the FDA-recognized standards ANSI/AAMI/IEC 60601-1-2:2014 and IEC 60601-1:2005 + A1:2012 (Ed. 3.1). The results from the testing pass the acceptance criteria outlined in the EMC and Electromagnetic Safety standards. The device is electrically safe for use in its intended environment.

SOFTWARE/CYBERSECURITY

GI Genius was identified as having a moderate level of concern as defined in the FDA guidance document "*Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices.*" The software documentation included:

1. Software/Firmware Description
2. Device Hazard Analysis
3. Software Requirement Specifications
4. Architecture Design Chart
5. Software Design Specifications
6. Traceability
7. Software Development Environment Description
8. Verification and Validation Documentation
9. Revision Level History
10. Unresolved Anomalies

Risk analysis was provided for the software with a description of the hazards, their causes and severity as well as acceptable methods for control of the identified risks. GI Genius provided a description, with test protocols including pass/fail criteria and report of results, of acceptable verification and validation activities at the unit, integration and system level.

Regarding the cybersecurity, the documentation included all the recommended information from the FDA guidance document "Content of Premarket Submissions for Management of Cybersecurity in Medical Devices." This includes a threat model, cybersecurity mitigation information, a malware-free shipping plan, an upgrade plan, and other information for safeguarding the algorithms.

PERFORMANCE TESTING – BENCH – STANDALONE PERFORMANCE

The purpose of the standalone performance testing is to demonstrate that the object-level, frame-level and overall algorithmic performance is sufficient to fulfill the indications for use of the GI Genius. This involves verification and validation of not only the software, but also additional performance testing of the algorithm alone to verify that it achieves acceptable detection performance, both overall and in important sub-populations. Standalone testing is also used to benchmark that performance as part of device labeling. The results provide an adequate benchmark for improved lesion detection and valid

scientific evidence that the chosen variable labels, features, and classifiers are sufficient to provide clinicians an aid for improved lesion detection.

The dataset used for the standalone performance testing was also used for algorithm training, and was originally from a study titled, "The Safety and Efficacy of Methylene Blue MMX® Modified Release Tablets Administered to Subjects Undergoing Screening or Surveillance Colonoscopy" [NCT01694966]. A diagram describing the use of the videos from that study is included in Figure 3 below. The multi-arm study included colonoscopy videos in which methylene blue was used (725 videos) and a control arm in which no methylene blue was used (480 videos). Training of the GI Genius AI software was done on a subset of videos in which polyps were identified, either in the presence or absence of methylene blue. To correct for bias in the data set due to training on methylene blue, a second fine-tuning training was performed on a subset of that same dataset, using only those videos without methylene blue. The fine-tuning training is not illustrated in the figure below.

In addition to algorithm training, the sponsor performed an independent validation or Holdout Testing using a total of 150 colonoscopy videos, without methylene blue. Of those videos, 105 included a total of 338 excised polyps with histology confirmation. The remaining 45 videos did not include polyps or lesions. The testing on these 150 colonoscopy videos is collectively referred to as standalone performance testing, and is separate from the clinical performance testing of the GI Genius that is described further below. The 150 colonoscopy videos had a total of 5,805,587 frames.
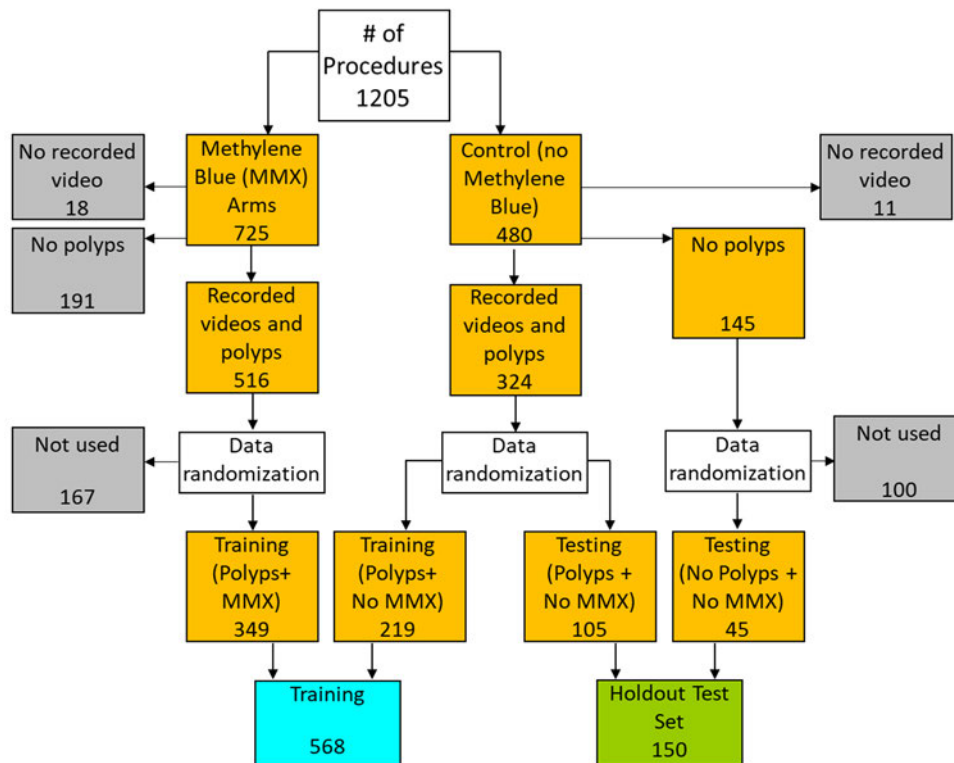

Figure 3. Diagram outlining the curation of the training and validation data

The patient demographics of the dataset used for training and the Holdout Test Set are provided in Table 1.

Table 1. Demographics information of MMX trial.

	Training (568 subjects)	**Holdout Test Set (150 subjects)**	**Overall (718 subjects)**
Mean Age, years (SD)	61.6 (6.58)	61.5 (6.32)	61.6 (6.59)
Sex, N (%)			
Male	370 (65.1%)	93 (62.0%)	463 (64.5%)
Female	198 (34.9%)	57 (38.0%)	255 (35.5%)
Indication for Colonoscopy, N (%)			
Screening	270 (47.5%)	73 (46.7%)	343 (47.8%)
Surveillance ≤ 2 years	43 (7.6%)	7 (4.7%)	50 (7.0%)
Surveillance > 2 years	255 (44.9%)	70 (48.7%)	325 (45.3%)
Race/Ethnicity			
White or Caucasian	522 (91.9%)	141 (94.0%)	663 (92.3%)
Black or African American	34 (6.0%)	5 (3.3%)	39 (5.4%)
Hispanic or Latino	7 (1.2%)	0 (0%)	7 (1.0%)
Asian	3 (0.5%)	3 (2.0%)	6 (0.8%)
Native Hawaiian or other Pacific Islander	1 (0.2%)	1 (0.7%)	2 (0.3%)

The most relevant characteristics of the lesions used in the standalone testing are reported in Figure 4. The charts below also show a wide distribution of the polyps that are meant to be detected. Approximately half of the 338 lesions in the Holdout Test Set were confirmed adenomas, and half were non-adenomas. The lesions were found throughout the colon, from the cecum to the rectum. Less than two-thirds of lesions had polypoid morphology, and the remaining had non-polypoid morphology. Approximately 70% of lesions were diminutive (less than 5 mm), approximately 20% were small (6-9 mm) and about 10% were considered to be large polyps (≥10 mm).
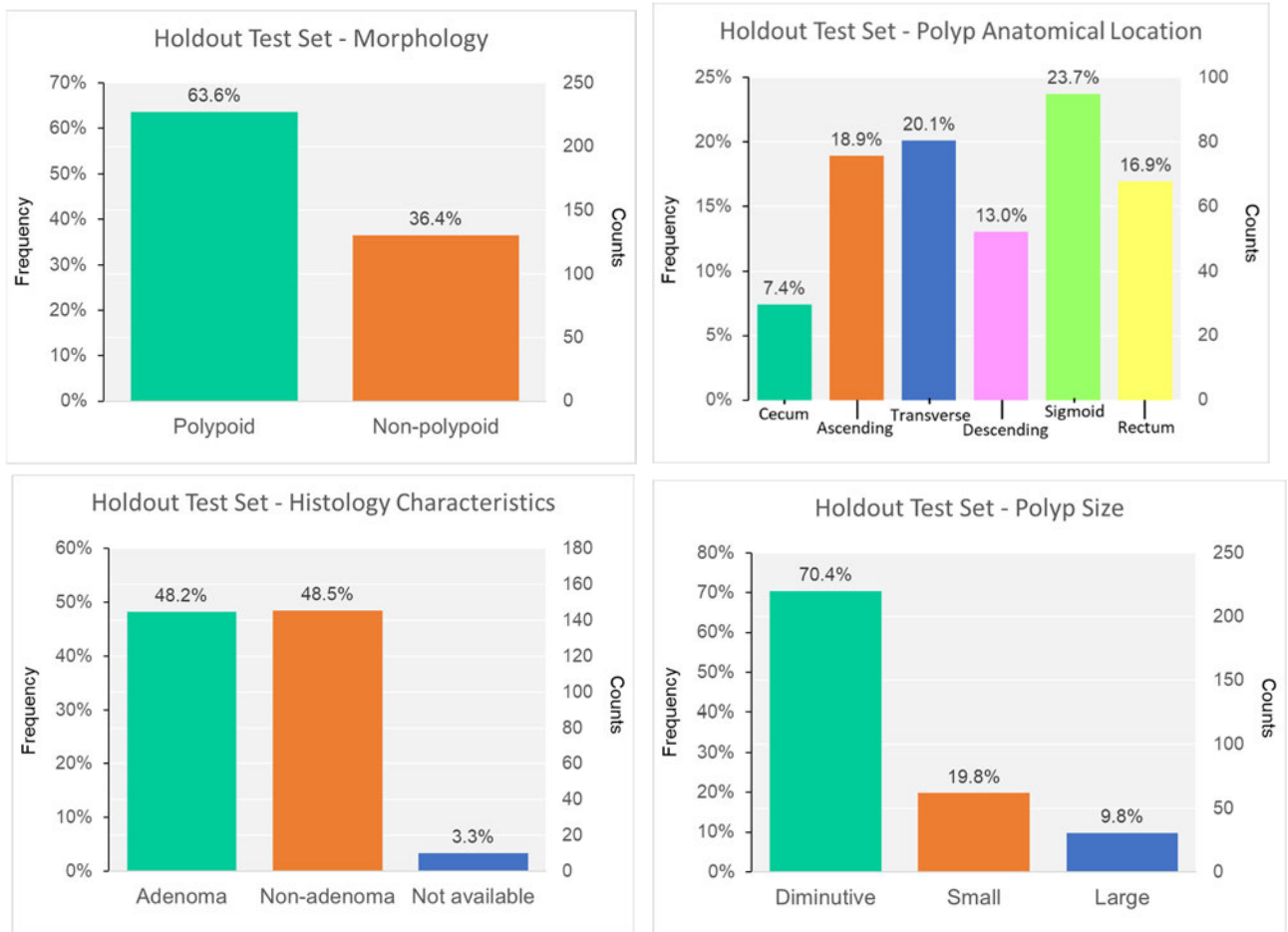
Figure 4. Charts summarizing the characteristics of the polyps used as part of the standalone performance dataset.

Standalone performance on the Holdout Test Set contained multiple elements including an assessment of the algorithm's activation time followed by an assessment of both object- and frame-level detection performance in terms of sensitivity, false positive rate and receiver operating characteristic (ROC) performance.

To assess true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), a reference standard was established. The standalone reference standard was created by having endoscopists review the video clips around all histologically confirmed polyps and placing an annotation box around the polyps visible in each frame. Those same video clips (without annotation) were analyzed by the GI Genius device; the device placed a marker on each frame in which the device identified a lesion. An assessment was then conducted to analyze the overlap between the endoscopists' annotation of lesions and the GI Genius marker for lesions, using an Intersection over Union (IoU) criterion, which is a metric of object detector accuracy.

Activation Time

The activation time refers to the time required for the device to detect a lesion and display a computer-aided detection (CAD) marker. For the device to function as intended, the device must detect and mark a lesion as it entered the field of view before the endoscopist identifies the lesion and before it exits the field of view.

To establish when the lesion is in this critical time frame, a panel of five expert endoscopists reviewed all the video clips containing polyps of the Holdout Test Set, along with an additional set of sham videoclips showing no polyps, in random order and recorded the moment of their first detection after accounting for endoscopists' base reaction time. GI Genius was found on average to have an activation time of 120 ms, which means that it detected a polyp 1270 ms (95% CI: 857 ms, 1684 ms) before the average endoscopist in this study. This result met the acceptance criterion of the device initializing and detecting a lesion faster than the reaction time of the endoscopists.

Object-Level Performance
The purpose of the object-level performance test is to measure the accuracy between the endoscopists' object detection and the GI Genius marker. To define when polyp detection is providing a true benefit, an experiment was first carried out to establish the offset in each excised polyp video-clip where the endoscopist did first spot the lesion. The endoscopist reviewed 338 video clips of the polyps after estimating each readers' Baseline Reaction Time using 15 calibration videoclips with only the last 10 used to estimate reaction time as a correction to the activation time. Each expert endoscopist reviewed all the videoclips of the activation time dataset in addition to 49 60-seconds sham videoclips showing no polyps. The endoscopist played each videoclip and stopped the playback when a polyp was detected. Then, the endoscopist had to localize the polyp. If the position was incorrect (based on a location < 10% of frame size from lesion center), the reaction was not calculable and the endoscopist was not asked to repeat the measurement. The endoscopists' mean detection performance was 300 ms (SD: 74 ms). 327/338 results were properly collected; none of the endoscopists were able to provide a measurable result for the remaining 11 cases.

Using this mean detection time of 300 ms, sensitivity analysis was performed by determining the fraction of lesions that were detected by the GI Genius in at least one frame before the average endoscopist detected them (number of video-clips where CDLlesionY<0). Therefore, a collection of False Positive objects are characterized as a function of cluster times.

The device sensitivity information in Table 2 and Figure 5 considers polyps as detected by the GI Genius only if they were marked at or before the average endoscopist detection time (the activation time). The following figure and table show how detection persistence in time (the duration of time a mark persists on the same target based on an IoU overlap criterion applied to the GI Genius marks across frames) correlates with polyp-based sensitivity and the number of False Positive targets: this testing below considers repeated marking overlays of the same target (polyps and false positives) as a single statistical event, instead of considering only markings in individual frames as a single statistical

event. This allows for an estimate of the number of unique targets (or objects) identified by GI Genius as a function of the time those targets persist in the field of view.

Table 2. Table demonstrating how detection persistence in time (the duration of a time a mark persists on the same target) relates to the number of False Positive objects and polyp-based Sensitivity.

| Persistence of Markers in milliseconds (ms)	FP Objects/Patient [95% Confidence Interval]	Sensitivity [95% Confidence Interval]
> 0	156.31 [135.61; 177.00]	81.96% [77.35%; 85.97%]
> 100 ms	65.00 [54.92;75.08]	70.03% [64.75%; 74.95%]
> 200 ms	33.09 [27.12; 39.06]	59.33% [53.79%; 64.70%]
> 300 ms	19.47 [15.46; 23.49]	48.62% [43.09%; 54.19%]
> 400 ms	14.09 [10.96; 17.22]	42.81% [37.38%; 48.37%]
> 500 ms	9.89 [7.50; 12.28]	35.47% [30.29%; 40.93%]
> 1000 ms	3.92 [2.71; 5.13]	20.80% [16.53%; 25.6]
> 1500 ms	2.03 [1.34; 2.73]	12.84% [9.42%; 16.96%]
> 2000 ms	1.25 [0.80; 1.70]	10.40% [7.31%; 14.23%]
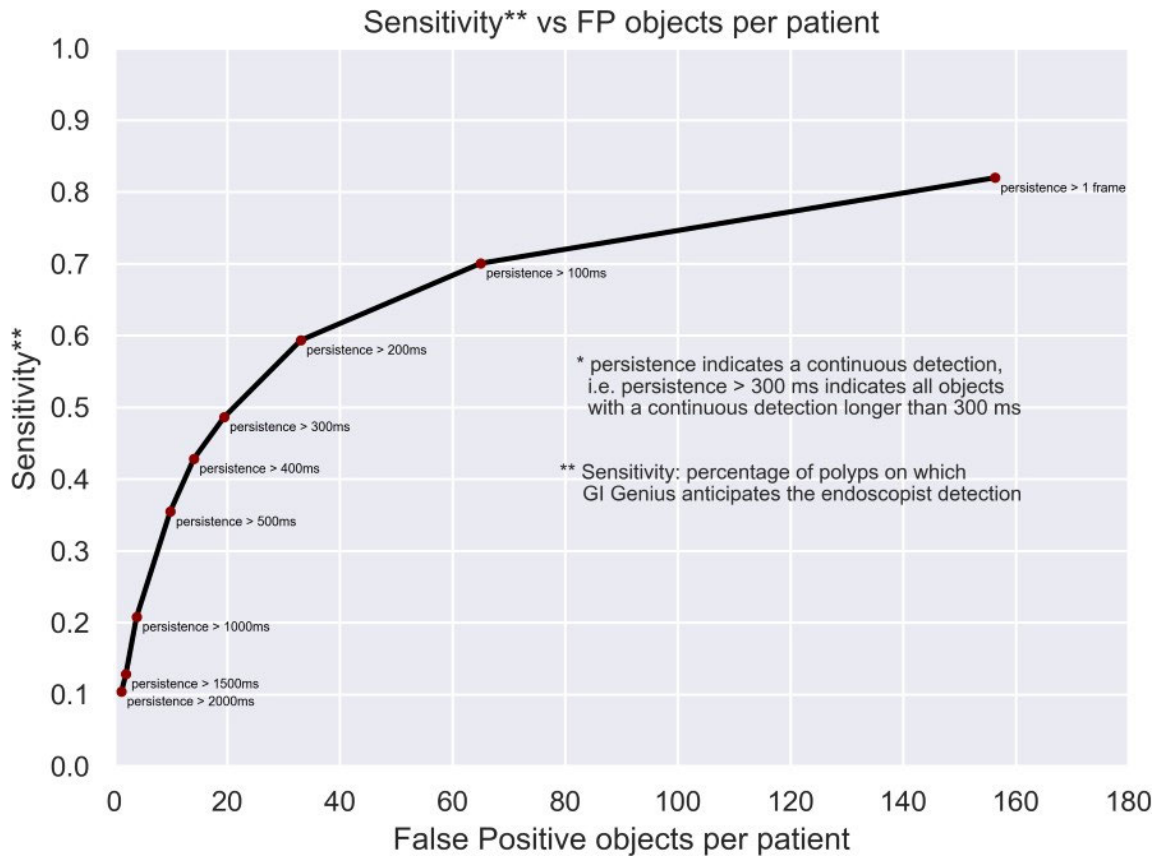
Figure 5. Graphical representation of Table 2, demonstrating the persistence of polyp-based Sensitivity and the number False Positive objects

The object-level performance shows that the GI Genius detects about 82% of the polyps before the average endoscopist detection time with about 156 false positive objects per colonoscopy exam. The number of false positive objects and true positive objects decreases as the length of time a target is marked increases. Many of the marks appear for a relatively small number of frames.

Frame-Level Performance
The frame-level performance is an assessment of the accuracy of the algorithm at sorting endoscopic images for quantification of false positives, false negatives, true positives, and true negatives. The sponsor expected a false positive rate per frame of 4.85% or lower based on an estimation from the video database in the Methylene Blue clinical trial.

Tables 3 and 4 summarize the performances of GI Genius when analyzing all 5,805,587 frames as individual statistical events in the Holdout Test Set calculated using two different statistical methods: Logistic regression mixed model analysis and Non-parametric cluster bootstrap analysis. Two different statistical methods were used for analysis, because the two models showed different performance levels between the two different analyses, and there is insufficient information to determine which analysis method is most appropriate.

The following definitions apply:
- True Positive Rate per Frame (TPR / FRAME) is the proportion of frames containing a polyp that were correctly detected by GI Genius;
- False Positive Rate per Frame (FPR / FRAME) is the proportion of frames not containing a polyp in which GI Genius did show a detection.

Table 3. Logistic Regression Mixed Model, with lesion random model for TPR and patient random model for FPR

Category		Mean Rate [95% Confidence Interval]	Percentage of Polyps Detected	Number of Videos
Overall		**TPR / Frame: 47.46% [42.51%; 52.45%]**	**99.70% (337/338)**	**105**
Histology	Adenoma	57.59% [50.62%; 64.26%]	99.39% (162/163)	69
	Non-Adenoma	38.68% [32.25%; 45.52%]	100.00% (163/163)	79
	Unknown	32.04% [14.30%; 57.11%]	100.00% (12/12)	9
Lesion Size	Diminutive (0-5 mm)	44.54% [38.88%; 50.35%]	99.60% (237/238)	92
	Small (6-9 mm)	64.95% [54.44%; 74.18%]	100.00% (67/67)	42
	Large (≥10 mm)	32.92% [20.76%; 47.89%]	100.00% (33/33)	25
Compatible Video Processors	Olympus CV-180	54.13% [46.45%; 61.62%]	99.28% (137/138)	44
	Olympus CV-190	42.6% [33.79%; 51.91%]	100.00% (93/93)	28
	Pentax EPK-i7000	25.31% [10.58%; 49.26%]	100.00% (12/12)	5
	Fujifilm VP-4450HD	45.9% [33.95%; 58.34%]	100.00% (52/52)	16
Overall		**FPR / Frame 1.44% [1.27%; 1.63%]**	**N/A**	**150**
Compatible Video Processors	Olympus CV-180	1.8% [1.50%; 2.16%]	N/A	63
	Olympus CV-190	1.26% 1.01%; 1.57%]	N/A	44
	Pentax EPK-i7000	1.45% [0.84%; 2.49%]	N/A	7

Fujifilm VP-4450HD	0.85% [0.62%; 1.15%]	N/A	23

* In this table, a polyp is considered detected if the GI Genius bounding box (overlay marker) adequately overlaps with the reference standard bounding box in at least one frame.

Table 4. Non-parametric Cluster Bootstrap analysis, considering within-patient correlation

Category		Mean Rate [95% Confidence Interval]	Percentage of Polyps Detected	Number of Videos
Overall		**TPR / Frame 49.57% [45.24%; 54.06%]**	**99.70% (337/338)**	**105**
Histology	Adenoma	55.24% [48.38%; 62.50%]	99.39% (162/163)	69
	Non-Adenoma	43.98% [38.25%; 49.91%]	100.00% (163/163)	79
	Unknown	53.63% [31.68%; 73.87%]	100.00% (12/12)	9
Lesion Size	Diminutive (0-5 mm)	45.91% [41.22%; 50.92%]	99.60% (237/238)	92
	Small (6-9 mm)	61.65% [53.87%; 69.75%]	100.00% (67/67)	42
	Large (≥10 mm)	50.18% [35.77%; 61.40%]	100.00% (33/33)	25
Compatible Video Processors	Olympus CV-180	55.44% [48.63%; 62.38%]	99.28% (137/138)	44
	Olympus CV-190	42.44% [36.14%; 50.43%]	100.00% (93/93)	28
	Pentax EPK-i7000	39.93% [21.07%; 63.06%]	100.00% (12/12)	5
	Fujifilm VP-4450HD	47.71% [40.88%; 59.24%]	100.00% (52/52)	16
Overall		**FPR / Frame 2.02% [1.72%; 2.35%]**	**N/A**	**150**
Compatible Video Processors	Olympus CV-180	2.22% [1.83%; 2.66%]	N/A	63
	Olympus CV-190	1.80% [1.28%; 2.49%]	N/A	44
	Pentax EPK-i7000	1.89% [0.97%; 3.17%]	N/A	7

| | Fujifilm VP-4450HD | 1.25% [0.76%; 1.98%] | N/A | 23 |

* In this table, a polyp is considered detected if the GI Genius bounding box (overlay marker) adequately overlaps with the reference standard overlay marker in at least one frame.

The algorithm Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC) and 95% confidence interval are shown in Figure 6, calculated with two different statistical methods:
- On the left: using a Logistic Regression Mixed Model, with lesion random model for TPR and patient random model for FPR
- On the right: using a Non-Parametric Cluster Bootstrap analysis, considering within-patient correlation
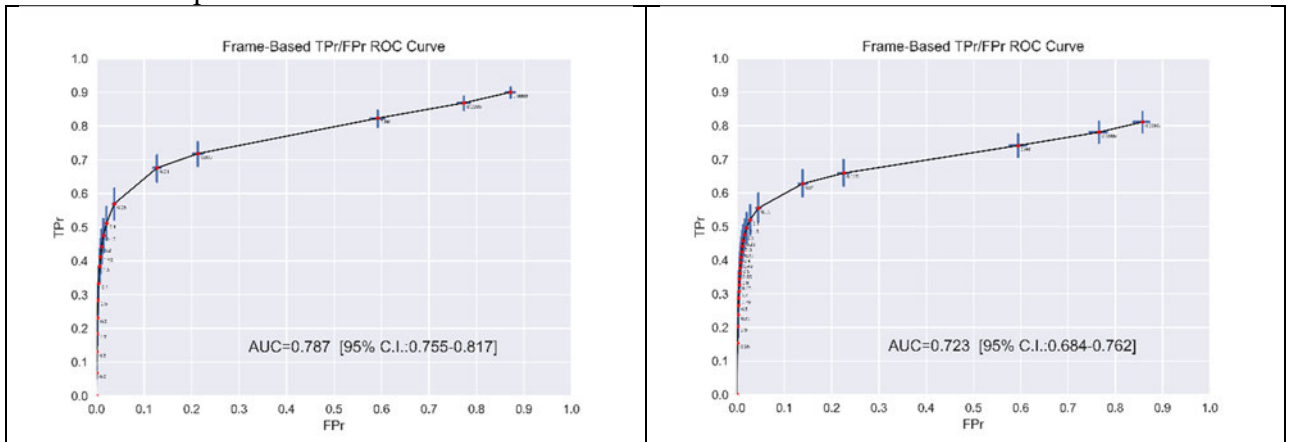

Figure 6. ROC and AUC curves for the TPR and FPR for the two models.

The frame-level performance shows that the device performs adequately for polyp detection. The overall false positive rate for the GI Genius was 1.44% and 2.02% per each of the frames in the Logistic Regression Mixed Model and Non-Parametric Cluster Bootstrap model, respectively. The results also demonstrate the TP and FP rates with a variety of endoscope video processors, although caution should be applied when interpreting the results when there is a small sample size. The frame-based results show a ~45% frame-detection rate for diminutive (≤5 mm) polyps, ~65% for small (6 mm-9 mm) polyps and ~33% for large (≥10 mm) polyps. This indicates that the GI Genius detects large polyps at a lower rate than diminutive and small polyps; however, these large polyps are also less likely to be missed by the endoscopists.

Standalone Performance Conclusions
Based on the above results, the standalone testing met the pre-defined performance criteria and were found to be adequate for benchmarking the GI Genius object- and frame-level performance overall and in relevant subgroups, as shown in the figures and tables above.

SUMMARY OF CLINICAL INFORMATION

GI Genius was tested in a randomized, prospective, multicenter, controlled clinical investigation performed in Italy, titled "The AID Study: Artificial Intelligence for Colorectal Adenoma

Detection" (Clinicaltrials.gov identifier: NCT04079478). The study was conducted at three medical centers in Italy (Humanitas Research Hospital [Milan], Nuovo Regina Margherita Hospital [Rome], and Valduce Hospital [Como]). Each medical center included two investigator endoscopists for a total of six investigators in the study. The study compared the performances of colonoscopies with the aid of GI Genius against standard colonoscopies with white light only. The study enrolled subjects between 40 and 80 years of age who were undergoing colonoscopies for primary colorectal cancer (CRC) screening or post-polypectomy surveillance, as well as for workup following fecal immunohistochemical test (FIT) positivity or for gastrointestinal symptoms. Patients were excluded in cases of personal history of CRC, inflammatory bowel disease, previous colonic resection, antithrombotic therapy precluding polyp resection, or lack of informed written consent. Eligible patients were randomized (1:1) between colonoscopy with the aid of GI Genius and standard colonoscopy. Six endoscopists with moderate endoscopy expertise (defined as an Adenoma Detection Rate [ADR] between 25 to 40%) conducted equal numbers of colonoscopy procedures with the GI Genius and standard unaided colonoscopy. (Note: standard colonoscopy is defined in this document as colonoscopy without the use of the GI Genius)

Study endpoints
The primary endpoint of the study was the Adenoma Detection Rate* (ADR*), defined in this study with an asterisk (*) as the proportion of patients with at least one histologically confirmed Adenoma or Carcinoma detected. The ADR, as typically used clinically and referenced in literature, is a validated quality indicator for colonoscopies, which is defined as the proportion of patients with at least one histologically confirmed Adenoma (not including carcinomas) detected. ADR was designated a surrogate measure of colonoscopy performance quality by the U.S. Multi-Society Task Force (USMSTF) on CRC and a minimum target detection rate in average-risk individuals is ≥ 25% for men and ≥ 15% for women older than 50 years undergoing their first examinations.

The mean number of adenomas per colonoscopy (APC) provides an important complement to the ADR quality metric for colonoscopies, as it provides greater discrimination between high performing and lower performing colonoscopists. Since endoscopists who conduct a perfunctory examination of the colon after removing the first adenoma ("one and done") could have the same ADR as an endoscopist who performs a thorough inspection, which yields more than one adenoma, the ADR metric can have significant variability and is prone to mischaracterizations. Therefore, a secondary endpoint of the study was adenomas per colonoscopy (APC). In addition, it was important to consider whether use of the GI Genius may result in unnecessary biopsy, and therefore the positive percent agreement (PPA) was also a secondary endpoint.

The statistical analysis plan was to demonstrate non-inferiority (10% margin) of the GI Genius in comparison to standard colonoscopy for the primary endpoint of ADR. After non-inferiority was met, a superiority analysis was conducted to evaluate GI Genius performance compared to standard colonoscopy. The statistical analysis plan was also intended to show superiority for APC, and non-inferiority (15% margin) for PPA.

The endpoints in the study are defined as:

Primary Endpoint:
- Adenoma Detection Rate* (ADR*): the proportion of patients with at least one histologically confirmed Adenoma <u>or Carcinoma</u> detected

Secondary Endpoints:
- Adenomas per Colonoscopy* (APC*): the total number of histologically confirmed Adenomas <u>and Carcinomas</u> detected, divided by the total number of colonoscopies;

- Positive Percent Agreement (PPA): the total number of histologically confirmed Clinically Significant Excised Lesions, divided by the total number of excisions.

FDA requested re-analysis of the primary (ADR*) and secondary (APC*) endpoints to utilize definitions for these endpoints that are more commonly used in the clinical setting:
- Adenoma Detection Rate (ADR): the proportion of patients with at least one histologically confirmed Adenoma detected

- Adenomas per Colonoscopy (APC): the total number of histologically confirmed Adenomas detected, divided by the total number of colonoscopies

Exploratory endpoints included:
- Polyps per Colonoscopy (PPC): the total number of histologically confirmed polyps detected, divided by the total number of colonoscopies
- Polyp Detection Rate (PDR): proportion of patients with at least one histologically confirmed polyp detected
- Serrated Lesions per Colonoscopy (SLPC): the number of histologically confirmed serrated lesions detected, divided by the total number of colonoscopies
- Serrated Lesions Detection Rate (SLDR): the proportion of patients with at least one histologically confirmed serrated lesion detected
- Advanced Adenoma Detection Rate (aADR), defined as proportion of patients with at least one histologically confirmed adenoma ≥ 10 cm or any adenoma < 10 mm, which was either of high-grade dysplasia (HGD) or villous or tubulovillous;
- Small Adenoma Detection Rate (sADR), defined as proportion of patients with at least one histologically confirmed adenoma smaller than 5 mm detected;
- Flat Adenoma Detection Rate (fADR), defined as the proportion of patients with at least one histologically confirmed non-polypoid adenoma detected;
- Proximal Adenoma Detection Rate (pADR), defined as the proportion of patients with at least one histologically confirmed adenoma detected in proximal colon;
- False Positive Rate, defined as the proportion of colorectal lesions resected or biopsied and subsequently not histologically confirmed to be clinically relevant colorectal polyps. All the biopsied or ablated specimens, which were histologically confirmed not to be polyps (e.g. normal mucosa, inflammatory tissue, stools or debris, etc.), were classified as False Positive.

Adenomas were defined as category 3 and 4.1 per revised Vienna classification (Table 5).

Table 5. The revised Vienna Classification.

CATEGORY	DESCRIPTION
1	Negative for neoplasia
2	Indefinite for neoplasia
3*	Mucosal low-grade neoplasia (low grade adenoma/dysplasia)
4	Mucosal high-grade neoplasia
	4.1* High-grade adenoma/dysplasia
	4.2 Non-invasive carcinoma (carcinoma in situ) †
	4.3 Suspicious for invasive carcinoma
	4.4 Intramucosal carcinoma ‡
5	Submucosal invasion of neoplasia (carcinoma invading the submucosa or beyond) or muscularis mucosae

† Non-invasive refers to the absence of evident invasion
‡ Intramucosal refers to invasion into the lamina propria
* For conventional adenomas, the histologist also specified whether the adenomas were: Tubular adenomas, Tubulovillous adenomas, Villous adenomas or other.

For calculating PPA, Clinically Significant Excised Lesions were defined as follows:
- Neoplastic lesions (classical adenomas and carcinomas);
- Sessile serrated lesions (SSL) classified according to the serrated lesion classification.
- Hyperplastic polyps (HP) of the proximal colon (caecum, ascending colon, hepatic flexure
  and transverse colon), classified according to the serrated lesion classification.

According to the World Health Organization (WHO), serrated lesions are currently classified into three main categories as follows:
1) hyperplastic polyps (HPs)
2) sessile serrated lesions (with or without dysplasia) (SSLs), and
3) traditional serrated adenomas (TSAs)

Study Population Demography
A total of 700 patients were screened and enrolled in the study: 350 patients were randomized to colonoscopy with GI Genius (GI Genius+colonoscopy) and 350 patients were randomized to standard colonoscopy. The study intent-to-treat (ITT) population included 40- to 80-year-old subjects undergoing colonoscopy for primary CRC screening or postpolypectomy surveillance, as well as for workup following fecal immunohistochemical test (FIT) positivity (cutoff = 20 mg Hb/g feces) or for symptoms/signs of CRC. Patients were excluded in case of personal history of CRC, or inflammatory bowel disease, inadequate bowel preparation (defined as Boston Bowel Preparation Scale > 2 in any colonic segment), previous colonic resection, or antithrombotic therapy precluding polyp resection. The primary analyses population (mITT, or modified Intent-to-Treat), which constituted the basis for the assessment of efficacy and safety of GI Genius was a subset of that population. The mITT population was limited to patients at low risk for CRC, i.e.

patients undergoing colonoscopy for primary screening of CRC or for surveillance within 3 to 10 years from previous colonoscopy. Limiting analysis to subjects at low risk for cancer is more likely to obtain consistency in the data and the two arms of the study, because even a few patients with high risk of cancer may have large numbers of polyps that can skew the results in that arm of the study. Furthermore, given that the prevalence of polyps in the low risk population is expected to be lower than in the high risk population (and, therefore, more difficult to detect), the assessment of performance in the low risk population is considered a "worst-case" testing scenario.

The mITT group comprised 263 patients in total, of whom 136 were randomized to GI Genius+colonoscopy and 127 were randomized to standard colonoscopy. Results of the study are presented only for the mITT group of subjects at low risk for CRC. The demographics of this low-risk group of patients is reported in the table below.

Table 6. Demographics information.

	GI Genius (136 subjects)	Standard Colonoscopy (127 subjects)	Overall (263 subjects)
Mean Age, years (SD)	60.6 (9.74)	59.9 (11.18)	60.3 (10.13)
Sex, N (%)			
Male	73 (53.7%)	62 (48.8%)	135 (51.3%)
Female	63 (46.3%)	65 (51.2%)	128 (48.7%)
Adequate bowel cleansing (total score ≥6 and no score <2 in any of the colon segments), N (%)	135 (99.3%)	126 (99.2%)	261 (99.2%)

Race/ethnicity information about study participants was not collected. However, the general populations accessing CRC care at these study sites in Italy were predominantly (>98%) Caucasian. It is assumed that the majority of study participants are Caucasian. We expect the performance as an aid to adenoma detection to be comparable in a US population, but the racial/ethnicity difference is an area of uncertainty and discussed further below in the Benefit-Risk Determination Section.

Study Results
In the original AID study, with the entire Intent to Treat population (n=700), the GI Genius+colonoscopy arm met the pre-specified 10% non-inferiority margin and subsequent superiority analysis for ADR, and demonstrated superiority for APC and 15% non-inferiority margin for PPA, compared to standard colonoscopy.

Re-analysis of the clinical study to limit the patient population to the mITT population (patients at low risk for CRC) and ADR*, ADR, APC*, APC, and PPA results are shown below.

a) Primary Endpoint – Adenoma Detection Rate (ADR)*
The ADR* was analyzed in the mITT set through a logistic regression mixed model, with treatment group, age (< 60, ≥60), reason for colonoscopy and sex as fixed effects, and endoscopist as random effect (random intercept). The estimates of ADR* after statistical adjustment were 55.1% (95% CI: 44.0% to 65.8%) in the GI Genius+colonoscopy group and

42.0% (95% CI: 31.3% to 53.4%) in the standard colonoscopy group.   The primary objective of the study was to assess the non-inferiority of GI Genius+colonoscopy versus standard colonoscopy in ADR*, with a 10% non-inferiority margin between the two arms prespecified as maximum acceptable difference. Since the non-inferiority claim was met (in other words, the lower bound of the confidence interval of the difference in ADR was larger than 10%), a superiority test was conducted. The superiority of the GI Genius+colonoscopy arm versus the standard colonoscopy arm was demonstrated, because the results met the lower bound of the confidence interval of the difference in ADR larger than 0.

Table 7. Primary endpoint information for ADR*.

Statistical Information	GI Genius (136 subjects)	Standard colonoscopy (127 subjects)
Adenoma Detection Rate (ADR*) (adjusted estimate, %) [95% C.I.]	**55.1** [44.0; 65.8]	**42.0** [31.3; 53.4]
Difference in ADR* between GI Genius and Standard Colonoscopy (adjusted estimate, %)	13.1 [0.09; 23.3]	
p-value for superiority	<0.05	

When analyzing the ADR excluding the carcinomas from the calculation, the results are the same and shown in Table 8. There was only one carcinoma detected in the mITT population.

Table 8. Primary endpoint information for ADR.

Statistical Information	GI Genius (136 subjects)	Standard colonoscopy (127 subjects)
Adenoma Detection Rate (ADR) (adjusted estimate, %) [95% C.I.]	**55.1** [44.0; 65.8]	**42.0** [31.3; 53.4]
Difference in ADR between GI Genius and Standard Colonoscopy (adjusted estimate, %) [95% C.I.]	13.1 [0.09; 23.3]	
p-value for superiority	<0.05	

b) Secondary Endpoints

Adenomas per Colonoscopy (APC*)
The APC* was analyzed in the mITT set through a negative binomial mixed model, with treatment group, age, reason for colonoscopy and sex as fixed effects, and endoscopist as random effect (random intercept).

The estimates of APC* after adjusting by age, sex, and indication for colonoscopy were 0.809 (95% CI: 0.567 to 1.154) in the GI Genius+colonoscopy group and 0.568 (95% CI: 0.393 to 0.820) in the standard colonoscopy group; the estimated counts (the APC*) ratio between the GI Genius+colonoscopy group and the standard colonoscopy group was 1.425 [1.027; 1.979]. Because the counts ratio is statistically significantly larger than 1, this indicates that the GI Genius+colonoscopy is superior to standard colonoscopy for this endpoint.

Table 9. Secondary endpoint information for APC*.

Statistical Information	GI Genius (136 subjects)	Standard colonoscopy (127 subjects)
Adenomas per Colonoscopy (APC*) (adjusted estimate) [95% C.I.]	**0.81** [0.57; 1.15]	**0.57** [0.39; 0.82]
Difference in APC* between GI Genius and Standard Colonoscopy (adjusted estimate)	0.24 [0.07; 0.53]	
Estimated counts ratio [95% C.I.]	1.43 [1.03; 1.98]	
p-value for superiority	0.03	

Table 10 provides the results for APC, which excludes carcinomas. There was only one carcinoma detected for APC*.

Table 10. Secondary endpoint information for APC.

Statistical Information	GI Genius (136 subjects)	Standard colonoscopy (127 subjects)
Adenomas per Colonoscopy (APC) (adjusted estimate) [95% C.I.]	**0.8** [0.56; 1.13]	**0.57** [0.39; 0.81]
Difference in APC between GI Genius and Standard Colonoscopy (adjusted estimate)	0.23 [0.06; 0.52]	
Estimated counts ratio [95% C.I.]	1.41 [1.02; 1.96]	
p-value for superiority	0.04	

Positive Percent Agreement (PPA)
The analysis of PPA in the mITT set was performed through a mixed effect logistic model for binomial data, with treatment group, age groups (< 60, ≥60), reason for colonoscopy (which include presence of GI symptoms, primary CRC screening, surveillance < 3 years, surveillance of 3-10 years, and FIT+) and sex as fixed effects, and endoscopist as random effect (random intercept). The objective was to assess the non-inferiority of GI Genius versus standard colonoscopy in PPA, with a 15% non-inferiority margin between the two arms prespecified as acceptable. The estimates of PPA after adjusting by age, sex, and indication for colonoscopy were 62.1% (95% CI: 43.4% to 77.8%) in the GI Genius+colonoscopy group and 65.2% (95% CI: 46.0% to 80.4%) in the standard colonoscopy group; the estimated difference in PPA is -3.1% (95% CI: -14.3% to 4.8%). As the lower bound of the difference in PPA is larger than -15%, the GI Genius+colonoscopy group was demonstrated to be non-inferior to the standard colonoscopy group.

Table 11. Secondary endpoint information for PPA.

Statistical Information	GI Genius (136 subjects)	Standard colonoscopy (127 subjects)

Positive Percent Agreement (PPA) (%) [95% C.I.]	**62.1** [43.4; 77.8]	**65.2** [46.0; 80.4]
Difference in PPA between GI Genius and Standard Colonoscopy (adjusted estimate, %)	-3.1 [-14.3; 4.8]	
p-value for non-inferiority of GI Genius versus standard colonoscopy	<0.05	

c) Exploratory Endpoints

Table 12. Exploratory endpoint information for clinically significant lesions.

Exploratory Endpoint	GI Genius (136 subjects)	Standard Colonoscopy (127 subjects7)
Polyps per Colonoscopy (PPC) [mean (SD)]	1.8 (1.80)	1.1 (1.21)
Polyp Detection Rate (PDR) [%]	77.2%	62.2%
Serrated Lesions per Colonoscopy (SLPC) [mean (SD)]	0.5 (1.02)	0.3 (0.64)
Serrated Lesions Detection Rate (SLDR) [%]	27.2%	24.4%
Advanced Adenoma Detection Rate (aADR) [%]	7.4%	4.7%
Small Adenoma Detection Rate (sADR) [%]	42.6%	33.9%
Flat Adenoma Detection Rate (fADR) [%]	23.5%	16.5%
Proximal Adenoma Detection Rate (pADR) [%]	34.6%	26.0%
False Positive Rate (FPR) [%]	0.9%	1.2%

* The reported exploratory analyses are not corrected for multiplicity such that each endpoint is assessed individually. These results are purely descriptive without taking uncertainty of the results into account.

The use of GI Genius resulted in no additional adverse events (i.e., perforations/bleeding due to additional biopsies, etc.) requiring new endoscopy or hospital admission during the clinical study. The study results demonstrated superiority for ADR* and APC*, and non-inferiority for PPA, which were the pre-specified statistical endpoints.

While the increase in the ADR* in the GI Genius group was modest, given the number of colonoscopies performed annually in patients with low risk for CRC, and given the fact that a single adenoma is an independent risk factor for cancer, the increase is clinically meaningful.

The ADR, APC, and PPA endpoints were met in both the original study population (n=700), which included patients at high risk and low risk for CRC, as well as the mITT population which was limited to patients at low risk for CRC. Demonstration of device performance in a population that consists solely of patients at low risk for CRC may provide a challenge to the GI Genius System as an aid to the clinician due to the expected decreased prevalence of polyps in a low risk population (as compared to a high risk population). The low risk population also reduces the potential for bias in patient and polyp populations across study arms. Furthermore, a study conducted solely on patients at high risk for CRC may not be representative of a low risk screening population, due to the increased scrutiny that high risk patients may be subjected to.

Pediatric Extrapolation

In this De Novo request, existing clinical data were not leveraged to support the use of the device in a pediatric patient population.

HUMAN FACTORS/USABILITY

A 19-point questionnaire was built into the clinical study to obtain usability data. The questions addressed the users' understanding of the labeling (e.g., initiating a feature of the GI Genius and configuring the field of view setting), assessment of the device (e.g., visibility of the markers, perceived accuracy of the markers, speed of the device in detection, distraction of the device), procedure-related information (e.g., endoscopes used, whether it was necessary to turn off the device during the procedure, reports of any malfunctions, withdrawal time observations), as well as the impact of the device on the overall procedure.

A total of (b) (4) users that are professionally trained in colonoscopy were tested. The users were comprised of the (b) (4) endoscopists from the clinical study and (b) (4) endoscopy nurses and personnel. Thirteen of the questions were semi-quantitative, and the answers to the questions were rated on a scale of 1 to 5. The remaining six questions were qualitative.

The acceptance criteria of the testing was for the threshold score for semi-quantitative questions to be above 3. Average scores for those questions ranged from 3.87 to 4.73. The threshold score was also at the lower margin of the two-sided 95% CI for each question score.

For the qualitative questions, no usability errors were reported, and there were no changes to the colonoscopy workflow when using the GI Genius in comparison to a standard colonoscopy procedure.

The usability assessment supports that the GI Genius does not have a negative impact on the clinical workflow, although a comprehensive quantitative assessment of any additional procedure and anesthesia time attributable to use of GI Genius (detection, biopsies, etc.) was not performed.

LABELING

The labeling includes a detailed description of the device and compatible products, description of the patient population for which the device is indicated for use, and instructions for use. The labeling also includes summary information on the non-clinical standalone performance testing and the clinical performance testing of the device.

The labeling includes warnings that prohibit the device from diagnosis or characterization of the lesions, and that the images and data acquired using the device are to be interpreted only by qualified medical professionals. There is a warning that the device should not replace clinician decision-making. There is also a warning regarding overreliance on the device.

RISKS TO HEALTH

The table below identifies the risks to health that may be associated with use of a gastrointestinal lesion software detection system and the measures necessary to mitigate these risks.

Identified Risks to Health	Mitigation Measures
Algorithm failure leading to: • False positives resulting in unnecessary patient treatment; or • False negatives resulting in delayed patient treatment	Clinical performance testing Non-clinical performance testing Software verification, validation, and hazard analysis Labeling
Failure to identify lesions, resulting in delayed patient treatment, due to software/hardware failure including: • Incompatibility with hardware and/or data source • Inadequate mapping of software architecture • Degradation of image quality • Prolonged delay of real-time endoscopic video	Software verification, validation, and hazard analysis Non-clinical performance testing Labeling Electromagnetic compatibility (EMC) Electrical safety, thermal safety, mechanical safety testing
False positive or false negative due to user overreliance on the device	Labeling Usability assessment

SPECIAL CONTROLS

In combination with the general controls of the FD&C Act, the gastrointestinal lesion software detection system is subject to the following special controls:

(1) Clinical performance testing must demonstrate that the device performs as intended under anticipated conditions of use, including detection of gastrointestinal lesions and evaluation of all adverse events.
(2) Non-clinical performance testing must demonstrate that the device performs as intended under anticipated conditions of use. Testing must include:
    (i) Standalone algorithm performance testing;
    (ii) Pixel-level comparison of degradation of image quality due to the device;
    (iii) Assessment of video delay due to marker annotation; and
    (iv) Assessment of real-time endoscopic video delay due to the device.
(3) Usability assessment must demonstrate that the intended user(s) can safely and correctly use the device.
(4) Performance data must demonstrate electromagnetic compatibility and electrical safety, mechanical safety, and thermal safety testing for any hardware components of the device.
(5) Software verification, validation, and hazard analysis must be provided. Software description must include a detailed, technical description including the impact of any software and hardware on the device's functions, the associated capabilities and limitations of each part, the associated inputs and outputs, mapping of the software architecture, and a description of the video signal pipeline.
(6) Labeling must include:

(i) Instructions for use, including a detailed description of the device and compatibility information;

(ii) Warnings to avoid overreliance on the device, that the device is not intended to be used for diagnosis or characterization of lesions, and that the device does not replace clinical decision-making;

(iii) A summary of the clinical performance testing conducted with the device, including detailed definitions of the study endpoints and statistical confidence intervals; and

(iv) A summary of the standalone performance testing and associated statistical analysis.

BENEFIT-RISK DETERMINATION

Risks and Other Factors
The risks of the device are based on data collected in a clinical study described above.

The use of GI Genius in addition to standard white light colonoscopy resulted in no additional adverse events (perforations/bleeding due to additional biopsies, etc.) during the clinical study, as compared to standard colonoscopy.

There is a risk that the use of the GI Genius may result in more unnecessary extractions. This was assessed by measuring the positive percent agreement (PPA) in the clinical study. Although the PPA value was lower for the GI Genius than standard colonoscopy, which is suggestive that use of the GI Genius resulted in slightly more unnecessary extractions, the difference in PPA was neither statistically nor clinically significant. The difference in PPA values and the lack of adverse event data related to additional biopsies suggest that there was not a substantially increased risk associated with non-essential extractions.

Benefits
The benefits of the device are based on data collected in a clinical study as described above.

The clinical study demonstrated a clinically meaningful and statistically significant improvement of ADR* with GI Genius compared to unaided standard colonoscopy. The primary endpoint was the ADR* defined as the proportion of patients with at least one histologically confirmed adenoma or carcinoma detected. The GI Genius procedures had an ADR* of 55.1% and the standard colonoscopy procedures had an ADR* of 42.0%. Literature suggests that increased ADR is associated with a reduced risk of interval CRC and death. (Kaminski, Michal F., et al. "Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death." Gastroenterology 153.1 (2017): 98-105.)

The clinical study also demonstrated a clinically meaningful and statistically significant improvement of APC*. The secondary endpoint of the adenomas per colonoscopy (APC*) was defined as the total number of histologically confirmed adenomas and carcinomas detected divided by the total number of colonoscopies. The APC* was 0.809 for the GI Genius patients and 0.568 for the standard colonoscopy patients.

As the primary purpose of colonoscopy in this patient population is the detection of clinically relevant pre-malignant lesions, both the primary and secondary endpoints support a more effective lesion detection performance in the GI Genius in comparison to the control.

Uncertainty

There were multiple areas of uncertainty associated with the data. The study design, which included subjects of limited diversity, has likely resulted in an overly optimistic performance of the device. We cannot exclude the possibility that differences in patient characteristics as they relate to ethnicity and race, would introduce more variables and impact the colonoscopy quality metrics. Because this device functions as an aid to clinicians and does not replace traditional colonoscopy, no patient population would be likely to experience outcomes that are worse than what is expected with the current standard of care, The training of the software included videos using methylene blue, which introduced uncertainty in the performance of the device when methylene blue is not used during colonoscopy. However, subsequent fine tuning and the clinical validation was done without the methylene blue, which supports the current indications for use. There is also uncertainty regarding the performance with different clinicians. The study included six endoscopists in Italy with ADR values between 25-40%. It is unknown if the study results would have been impacted if more endoscopists had been included, if the endoscopists were based in the US, or if the endoscopists had ADR values that were above or below 25-40%. Another area of uncertainty is the long-term usage of AI software devices, and the potential for overreliance on lesion detection software leading to errors in treatment and diagnosis, rather than the intended use of the device as an aid to clinicians.

Related to the primary and secondary endpoints, although changes in ADR have been linked in the clinical literature to clinical outcomes, the impact of differences in APC and PPA on clinical outcomes is unknown. Furthermore, some clinically significant lesions, such as serrated lesions, did not occur with sufficient frequency in the study to draw conclusions about the performance of the GI Genius at detecting those lesions; the performance of the GI Genius at detection of non-adenoma clinically significant lesions is uncertain. There were also few videos in the standalone performance testing used with some models of video processors; the low sample sizes with those video processors is another source of uncertainty.

Conclusion

Because colon polyps and adenomas are precursor lesions for CRC, colonoscopic detection and removal of adenomas are an important aspect of CRC prevention. Literature confirms that there is wide variation in endoscopists' success at detecting the precursor lesions, and the ADR is associated inversely with the risk of interval CRC (i.e., a cancer diagnosed before the next surveillance examination is due) and CRC death.

The GI Genius was shown to have improved detection rates compared to standard colonoscopy based on ADR* and APC*. Therefore, the improved ADR* and APC* might suggest a reduction in cancer risk, although this would need to be confirmed in further clinical trials. There is some uncertainty about the degree of benefit that will be seen in a more diverse population of patients as well as in the hands of endoscopists with different colonoscopy skills. Nonetheless,

the GI Genius software device demonstrates a benefit for adenoma detection as compared to standard colonoscopy and is likely to result in clinical benefit to patients. Despite the existence of uncertainty, there is unlikely to be a negative impact on the standard colonoscopy outcomes. Furthermore, there is likely to be an improvement in the adenoma detection rate with the use of the GI Genius.

Based on the above information, the probable benefits of the GI Genius outweigh the probable risks in light of the listed special controls and the general controls.

Patient Perspectives

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

Benefit/Risk Conclusion

In conclusion, given the available information above, for the following indication statement:

The GI Genius System is a computer-assisted reading tool designed to aid endoscopists in detecting colonic mucosal lesions (such as polyps and adenomas) in real time during standard white-light endoscopy examinations of patients undergoing screening and surveillance endoscopic mucosal evaluations. The GI Genius computer-assisted detection device is limited for use with standard white-light endoscopy imaging only.

The probable benefits outweigh the probable risks for the GI Genius. The device provides benefits and the risks can be mitigated by the use of general controls and the identified special controls.

CONCLUSION

The De Novo request for the GI Genius is granted, and the device is classified as follows:
> Product Code: QNP
> Device Type: Gastrointestinal lesion software detection system
> Regulation Number: 21 CFR 876.1520
> Class: II