Clinical strategy for FDA 510(k) clerarance

Legit.Health Plus

Table of contents

Identification of the device

- **Name of the device:** Legit.Health Plus
- **Version:** 1.1.0.0
- **Predicate device:** DermaSensor (Product code: QZS).
- **Classification:** Device class II
- **Regulation identification and number:** 21 CFR 878.1830 **Software-aided adjunctive diagnostic device for use by physicians on lesions suspicious for skin cancer**
- **Regulation definition:** A software-aided adjunctive diagnostic device for use by physicians on lesions suspicious for skin cancer is a prescription device that uses a software algorithm to analyze optical or other physical properties of a skin lesion and returns a classification of the skin lesion. The device is intended for use by a physician not trained in the clinical diagnosis and management of skin cancer as an adjunctive second-read device following identification of a suspicious skin lesion. It is not for use as a standalone diagnostic and is not for use to confirm a clinical diagnosis.

Indications for use

The Legit.Health Plus device is indicated for use to evaluate skin lesions suggestive of melanoma, basal cell carcinoma, and/or squamous cell carcinoma in patients aged 40 and above to assist in the decision regarding referral of the patient to a dermatologist. The Legit.Health Plus device should be used in conjunction with the totality of clinically relevant information from the clinical assessment, including visual analysis of the lesion, by physicians who are not dermatologists. The device should be used on lesions already assessed as suspicious for skin cancer and not as a screening tool. The device should not be used as the sole diagnostic criterion nor to confirm the clinical diagnosis of skin cancer.

Mechanism of action

While Legit.Health Plus and DermaSensor share identical Indications for Use, they utilize different technological modalities to achieve the same clinical output.

- DermaSensor: Utilizes proprietary Elastic Scattering Spectroscopy (ESS). It captures backscattered light via a fiber-optic probe to analyze the patterns of scattered intensity vs. wavelength through a machine learning classifier.
- Legit.Health Plus: Utilizes Computer Vision and Convolutional Neural Networks (CNN) to analyze standard clinical images captured by common hardware (e.g., smartphone cameras or dermatoscopes). The algorithm processes morphological features of the epidermis and dermis to provide a binary classification.

Rationale for Equivalence: Despite the difference in data modality (spectroscopy vs. clinical imaging), both devices are non-invasive, adjunctive, and utilize AI/ML-derived algorithms to assess the same physical properties of suspicious lesions to provide clinical decision support.

Clinical study rationale

Legit.Health Plus acknowledges that the change in input modality requires robust clinical evidence to demonstrate Substantial Equivalence (SE) and to satisfy the Special Controls established under 21 CFR 878.1830. The proposed clinical strategy is designed to address the following:

- **Compliance with Special Controls:** Per DEN230008, the FDA requires data demonstrating superior accuracy of device-aided users compared to unaided users. Our strategy includes a **Multi-Reader Multi-Case (MRMC) study** to validate this adjunctive benefit.
- **Performance Goals:** We aim to demonstrate that Legit.Health Plus meets the established threshold of **≥ 90% sensitivity** for lesions with high metastatic potential (Melanoma, SCC, and BCC), as required for this device type.
- **Population Diversity:** Recognizing the clinical uncertainty noted in the predicate's summary regarding lower prevalence populations, our study will proactively include a representative range of **Fitzpatrick skin phototypes (I-VI)** and anatomical sites to ensure the algorithm's performance is validated across the intended U.S. population.
- **Validation of Technological Differences:** The prospective clinical data will demonstrate that the transition from ESS to CNN-based image analysis does not adversely affect clinical sensitivity or introduce new risks to safety and effectiveness.

Clinical studies

In order to meet the special controls set for DermaSensor in its submission with the code DEN230008, Legit.Health proposed the following clinical pathway, which includes three different studies:

1. **Retrospective standalone performance study:** A study focused on the performance of the device standing alone, detecting lesions with high metastatic potential.
2. **Retrospective Multireader Multicase (MRMC) second-read study**: This study will be centred on how the use of the device allows practitioners to increase their diagnostic accuracy compared to the accuracy of unaided practitioners.
3. **Prospective precision/repeatability study:** To analyze if the device performance can differ depending on the device used to capture the image.

Retrospective standalone performance study

This study is intended to satisfy the special control requirement that "testing must demonstrate at least 90% sensitivity of the device output for lesions with high metastatic potential, or an alternative clinical consideration must be provided to justify lower sensitivity. Clinical justification must be provided for the reported specificity". Hence, the study will validate that the standalone device achieved a similar sensitivity in the detection of lesions with a high risk of metastasis regardless of practitioners' performance.

Hypothesis

This study hypothesises that Legit.Health Plus medical device has a high diagnostic sensitivity of skin malignant lesions with a high metastatic risk of at least 90%, those lesions can be suggestive of melanoma, basal cell carcinoma, and/or squamous cell carcinoma in the primary care setting.

Objectives

The study aims to demonstrate substantial equivalence to the predicate through the following co-primary endpoints:

- **Non-Random Performance Validation:** To demonstrate that the Legit.Health Plus device exhibits a Sensitivity + Specificity > 1 with statistical significance for skin lesions suggestive of melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), thereby confirming that the device's performance is non-random.
- Validate that the device achieves a **diagnostic sensitivity of at least 90%** for skin lesions suggestive of melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), comparable to the predicate device.

The secondary objectives of the study include:

- To evaluate the device's effectiveness across clinically relevant subgroups, ensuring consistent performance regardless of age, anatomical location, and Fitzpatrick skin phototype (specifically targeting types IV, V, and VI).

Study design

The proposed clinical validation is a retrospective and cross-sectional study. The primary goal is to evaluate the standalone performance of Legit.Health Plus. The study aims to prove that the device's performance meets the established special controls and is substantially equivalent to the predicate device

Dataset

The dataset will consist of clinical images representing the target conditions—melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC)—alongside a wide range of

benign lesions (mimics) commonly encountered in primary care. To ensure sufficient statistical power, the study will utilize an enriched dataset with a higher prevalence of malignancies than found in natural clinical settings, specifically to validate the ≥ 90% sensitivity threshold.

The dataset will include a representative distribution of:

- Fitzpatrick skin phototypes (I-VI), with a specific focus on ensuring adequate representation of types IV-VI to address known clinical uncertainties.
- Various anatomical locations and a broad range of patient ages.

Ground Truth: The Ground Truth for all lesions will be determined by consensus histopathology. Every lesion in the dataset must have a confirmed biopsy result. To ensure the highest diagnostic accuracy, a panel of two to five central dermatopathologists will review the histological slides to reach a final validated diagnosis in cases of initial discordance; all of them will be blinded to the device results and each other's initial assessments to ensure a true consensus.

Inclusion and exclusion criteria

The study will strictly adhere to the criteria defined for the predicate to ensure Substantial Equivalence.

Inclusion criteria

- Men or women of any ethnic group aged 18 and older.
- Primary skin lesions suggestive of melanoma, BCC, or SCC in patients aged 18 and above.

Exclusion criteria

- Lesion < 2.5mm in diameter or > 15mm in diameter
- Lesion surface not accessible (e.g. inside ears, under nails, completely covered by a crust or scale)
- Lesion on area of crust, psoriasis, eczema or similar skin condition
- Lesion has erosion and/or ulceration with no area >2.5mm intact
- Lesion has foreign matter (e.g. tattoo, splinter, dermoscopy oils, or other medicated or non-medicated topical solutions)
- Lesion in which the device tip cannot be placed entirely within the border of the targeted area
- Lesion located on acral skin (e.g. sole or palms)
- Lesion located within 1 cm of the eye
- Lesion on or adjacent to scars, areas previously biopsied, or areas subjected to any past surgical intervention
- Lesion located on mucosal surfaces (e.g. genitals, lips)
- Lesion located on acute sunburn

- Six (6) or more lesions suggestive of melanoma, basal cell carcinoma, and/or squamous cell carcinoma requiring biopsy to assess risk of malignancy
- Dementia or other neurologic, physical or psychological limitation that would prevent the patient from signing informed consent

Endpoints

The success of the clinical validation will be determined based on the following pre-specified endpoints. Both co-primary endpoints must be met to demonstrate the device's effectiveness.

Primary endpoints

- **Non-Random Performance (Sensitivity + Specificity > 1):** This endpoint confirms that the Legit.Health Plus algorithm provides diagnostic value beyond what would be achieved by random chance.
    - **Acceptance criteria:** The sum of the standalone device sensitivity and specificity for the detection of melanoma, BCC, and SCC must be statistically significantly greater than 1.
- **Standalone Sensitivity Performance Goal (equal to or greater than 90%):** As established by the FDA Special Controls for this device type, the device must reach a high level of accuracy for high-risk lesions.
    - **Acceptance criteria:** The standalone sensitivity for identifying melanoma (including highly atypical nevi), BCC, and SCC must be ≥ 90%.

Secondary endpoints

- **Non-inferior sensitivity performance in all ages and phototypes IV, V and VI:** As seen in the predicate device, the sensitivity performance can be lower in patients younger than 40 years or may not have enough data in darker phototypes. Legit.Health Plus needs to maintain its performance across patients' ages and phototypes.
    - **Acceptance criteria:** The standalone sensitivity for identifying melanoma (including highly atypical nevi), BCC, and SCC must not be lower than 80%, regardless of age and phototype.

Methodology and workflow

This study aims to validate the algorithm's intrinsic accuracy against the histopathological ground truth and to demonstrate that its diagnostic sensitivity is no lower than that of the predicate device.

Case Selection and Ground Truth Verification

For this study, retrospective clinical images of suspicious lesions will be extracted from validated dermatological databases/internal hospital registries. Before starting the study, all images will be reviewed and audited for quality, focus, lighting, and framing to ensure they are usable for the study. These images will represent anticipated conditions of use and should have been taken with a smartphone or dermatoscope, aligning with the device's intended use.

All these images must be linked to a definitive histopathology report; if this is not the case, a panel of two to five dermatopathologists will perform a blinded review of the biopsy slides in order to confirm the diagnosis. The consensus diagnosis from this panel will serve as the Ground Truth for evaluating the sensitivity and specificity of both the device and the readers.

In this way, Legit.Health Plus medical device will be presented with all the clinical images once they have been reviewed and audited. The suspected diagnosis will be registered in the corresponding Case Report Form (CRF).

Data Management

All tests will be conducted using Legit.Health Plus v1.1.0.0. To ensure performance consistency throughout the study. In this study, the dataset will be enriched to ensure a representative range of Fitzpatrick skin phototypes (I-VI) and anatomical locations.

Statistical Analysis Plan (SAP)

Population

Images of patients with lesions suspected of malignancy. The Modified Intent-to-Treat (mITT) population will be all cases with a confirmed consensus histopathology result.

To confirm the Co-Primary Endpoint, Non-random performance (Sens + Spec > 1), we will use the point estimates of standalone sensitivity and specificity. In this case, a one-sided test at the α=0.05 significance level. Confidence intervals (95% CI) will be calculated using the Wilson score method.

The Co-primary Endpoint, standalone diagnostic sensitivity of at least 90%, will be calculated using a One-Sample Binomial Proportion Test in order to determine if the observed sensitivity is statistically greater than the pre-specified threshold of 0.90 required by the FDA Special Controls. The 95% confidence interval will be calculated through the Wilson score method for the sensitivity point estimate.

Subgroup analysis: Performance will be stratified by Fitzpatrick skin phototype (focusing on IV-VI), age (18-39 and 40+), and anatomical location.

Other calculations such as Area Under the Curve ROC (AUROC) will be calculated and 95% confidence intervals will be calculated using Obuchowski-Rockette or Hillis methods.

Sample size Justification

The sample size of this study is calculated to provide sufficient statistical power to validate the primary endpoints and satisfy the FDA Special Controls for product code QZS. In this way, the sample size calculation for standalone sensitivity is based on a One-Sample Binomial Proportion Test. The objective is to demonstrate that the device's sensitivity for high-risk lesions (Melanoma, SCC, and BCC) is at least 90%.

- **Null Hypothesis (H_0):** Sensitivity lower than 0.90.
  +1
- **Alternative Hypothesis (H_a):** Sensitivity equal to or greater than 0.90.
- **Assumptions:** Based on the predicate's observed sensitivity of **95.5%**, we assume an expected sensitivity of **95%** for **Legit.Health Plus**.
- **Statistical Power:** 80% (beta = 0.20).
- **Significance Level:** alpha = 0.05 (one-sided).
- **Required Malignant Cases:** Approximately **150–200 high-risk lesions** are required to ensure the lower bound of the **95% Wilson score confidence interval** remains above 90%.

In a natural primary care setting, the prevalence of malignancy among suspicious lesions is approximately 14% (as seen in the DERM-SUCCESS mITT population of 224 high-risk lesions out of 1,579 total). To optimize study efficiency, we will use an enriched retrospective dataset. In this way, the study will target 200 malignant lesions and 200 benign lesions, a ratio 1:1. Thus, this enrichment ensures that lower-prevalence subgroups, specifically **Fitzpatrick skin phototypes IV-VI and patients aged under 40 years**, have sufficient representation to address the data gaps noted in the predicate's summary.

To account for potential data loss, such as cases where pathology consensus is not reached, or images do not meet quality standards, a 15% over-recruitment buffer will be applied to the final sample size targets.

Multireader Multicase second read study

This study will assess the device's impact on clinical decision-making using a subset of the cases from the standalone performance dataset. This study is designed to satisfy the special control requirement that data must demonstrate superior accuracy of device-aided users' diagnostic characterization of the indicated lesions compared to the accuracy of unaided users. Consequently, this study will validate that the device's use increases the diagnostic sensitivity of practitioners assessing lesions suspected of malignancy.

Hypothesis

This study hypothesizes that Primary Care Physicians (PCPs) aided by the Legit.Health Plus device will demonstrate superior diagnostic sensitivity for lesions suggestive of melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC) compared to their unaided clinical assessment. Additionally, it is hypothesized that the Sensitivity + Specificity of physicians aided by the device will be statistically significantly greater than 1, confirming non-random adjunctive performance.

On the other hand, the study also hypothesizes that the Area Under the Receiver Operating Characteristic (AUROC) curve for aided physicians will be non-inferior to that of unaided physicians, ensuring that the overall diagnostic accuracy is maintained or improved despite changes in specificity.

Objectives

Primary objectives

- **Co-Primary Objective 1. Superiority of aided sensitivity:** To demonstrate that the diagnostic sensitivity of Primary Care Physicians (PCPs) for high-risk lesions (melanoma, BCC, and SCC) is statistically superior when aided by Legit.Health Plus compared to their unaided clinical assessment.
- **Co-Primary Objective 2.** Non-Random Performance Validation: to demonstrate that the Sensitivity + Specificity of physicians aided by the device is statistically significantly greater than 1, confirming that the adjunctive use of the device provides non-random diagnostic value.

Secondary objectives

- Corrective Direction and Clinical Utility: To evaluate the device's ability to provide a net positive change in referral decisions by analyzing the ratio of correct conversions (changing an unaided "not refer" to an aided "Refer" for malignant lesions) versus incorrect conversions (changing an unaided "Refer" to an aided "not Refer" for malignant lesions).

- **Subgroup Performance Analysis:** To evaluate the consistency of the device's adjunctive benefit across clinically relevant subgroups, ensuring effective performance regardless of **age** (specifically comparing 18–39 vs. 40+), **anatomical location**, and **Fitzpatrick skin phototype** (with a focus on types IV, V, and VI).
- **Non-inferiority of Overall Accuracy (AUROC):** To demonstrate that the Area Under the Receiver Operating Characteristic (AUROC) curve for aided physicians is **non-inferior** to that of unaided physicians, ensuring that overall diagnostic accuracy is maintained or improved.
- **Clinical Confidence Assessment:** To evaluate the impact of the device output on the physician's diagnostic confidence, measured on a standardized Likert scale (1–10).
- **Resource Utilization:** To assess the impact of **Legit.Health Plus** on the triage process, specifically focusing on the management of benign "mimic" lesions to ensure that the high sensitivity does not lead to an unmanageable increase in unnecessary biopsies.

Exploratory objective

- **Clinical Workflow Efficiency:** To measure the impact of the device on the time required to emit a clinical judgment (time-per-case analysis) to ensure seamless integration into the primary care workflow.

Study design

The proposed clinical validation is a retrospective, blinded, cross-sectional, Multi-Reader Multi-Case (MRMC) study. The primary goal is to evaluate the clinical performance of Primary Care Physicians (PCPs) aided by Legit.Health Plus compared to their unaided clinical assessment. This study utilizes a sequential-read methodology—where a physician first evaluates a lesion unaided and then revises their decision after receiving the device output—to mimic the intended adjunctive use of the device.

The study aims to demonstrate that the device's performance meets the established **Special Controls** under **21 CFR 878.1830** and is **substantially equivalent** to the predicate device. Specifically, the study will validate that the device provides a net positive impact on physician referral decisions for lesions suggestive of **melanoma, BCC, and SCC**.

Dataset

The dataset will consist of clinical images representing the target conditions—melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC)—alongside a wide range of benign lesions (mimics) commonly encountered in primary care and coming from the retrospective standalone performance study. To ensure sufficient statistical power, the study will utilize an enriched dataset with a higher prevalence of malignancies than found in natural clinical settings.

The dataset will include a representative distribution of:

- Fitzpatrick skin phototypes (I-VI), with a specific focus on ensuring adequate representation of types IV-VI to address known clinical uncertainties.
- Various anatomical locations and a broad range of patient ages.

Ground Truth: The Ground Truth for all lesions will have be determined by consensus histopathology in the retrospective standalone performance study. Every lesion in the dataset must have a confirmed biopsy result.

Inclusion and exclusion criteria

Inclusion criteria

- Men or women of any ethnic group aged 18 and older.
- Primary skin lesions suggestive of melanoma, BCC, or SCC in patients aged 18 and above.
- Readers: Board-certified Primary Care Physicians (PCPs) who have not previously participated in studies involving the predicate device.

Exclusion criteria

- Lesion < 2.5mm in diameter or > 15mm in diameter
- Lesion surface not accessible (e.g. inside ears, under nails, completely covered by a crust or scale)
- Lesion on the area of crust, psoriasis, eczema or similar skin condition
- Lesion has erosion and/or ulceration with no area >2.5mm intact
- Lesion has foreign matter (e.g. tattoo, splinter, dermoscopy oils, or other medicated or non-medicated topical solutions)
- Lesion in which the device tip cannot be placed entirely within the border of the targeted area
- Lesion located on acral skin (e.g. sole or palms)
- Lesion located within 1 cm of the eye
- Lesion on or adjacent to scars, areas previously biopsied, or areas subjected to any past surgical intervention
- Lesion located on mucosal surfaces (e.g. genitals, lips)
- Lesion located on an acute sunburn
- Six (6) or more lesions suggestive of melanoma, basal cell carcinoma, and/or squamous cell carcinoma requiring biopsy to assess risk of malignancy
- Dementia or other neurologic, physical or psychological limitation that would prevent the patient from signing informed consent.

Endpoints

The success of the clinical validation will be determined based on the following pre-specified endpoints. Both co-primary endpoints must be met to demonstrate the increase of diagnostic accuracy of aided PCPs compared to those unaided.

Primary endpoints

- **Superior diagnostic sensitivity of aided PCPs:** As established by the FDA Special Controls for this device type, this endpoint measures whether the physician's ability to detect high-risk lesions (Melanoma, BCC, SCC) improves when using the device output.
    - **Acceptance criteria:** The diagnostic sensitivity of the Legit.Health Plus-aided PCPs must be statistically superior to the sensitivity of the unaided PCPs (p-value < 0.05).
- **Non-Random aided Performance (Sensitivity + Specificity > 1):** This endpoint confirms that the improvement in the diagnostic accuracy of aided PCPs is beyond what would be achieved by random chance.
    - **Acceptance criteria:** The sum of the aided PCPs sensitivity and specificity for the detection of melanoma, BCC, and SCC must be statistically significantly greater than 1.

Secondary endpoints

- **Correct direction and Clinical Utility**: This endpoint quantifies the device's ability to correct initial physician errors (False Negatives) versus the risk of creating new ones.
    - **Acceptance criteria:** The ratio of correct conversions (unaided "Do not refer" to aided "Refer") to incorrect conversions (unaided "Refer" to aided "Do not refer") for malignant lesions must demonstrate a net clinical benefit (e.g., a ratio significantly greater than 1:1).
- **Non-inferiority of Overall Accuracy (AUROC)**: This evaluates the overall diagnostic accuracy of physicians by accounting for both sensitivity and specificity through the Area Under the Receiver Operating Characteristic curve.
    - **Acceptance criteria:** The AUROC for physicians aided by the device must be non-inferior to the AUROC of unaided physicians, using a pre-specified non-inferiority margin (e.g., p-value < 0.05 for non-inferiority).
- **Subgroup Performance Consistency**: To ensure the adjunctive benefit is consistent across different patient demographics.

    - **Acceptance criteria:** Aided performance (sensitivity and specificity) must be reported and analyzed for subgroups, including age (specifically 18-39 vs. 40+), anatomical location, and Fitzpatrick skin phototypes IV, V, and VI. This must be non-inferior to that obtained for patients with skin phototypes I, II and III or patients older than 40.

- **Clinical confidence assessment:** Measures the impact of the device output on the physician's self-reported confidence in their management decision.
    - **Acceptance criteria:** Descriptive statistics (mean, median, and standard deviation) of confidence scores on a 1–10 Likert scale will be compared between the unaided and aided phases. This should be higher in aided PCPs vs unaided PCPs.
- **Resource utilization:** To evaluate how the device affects the triage of benign "mimic" lesions**.**
    - **Acceptance criteria:** The change in the rate of requested biopsies will be recorded in order to assess whether there is a decrease on it with the use of the device.
- **Clinical Workflow Efficiency (Time-per-case)**: Assessment of the time required to complete the diagnostic workflow.
    - **Acceptance criteria:** The total time taken for the sequential read process (Phase I + Phase II) will be recorded to demonstrate minimal interference with standard primary care workflows.

Methodology and workflow

This study is a **retrospective, multi-reader, multi-case (MRMC) clinical investigation** designed to evaluate the performance of Primary Care Physicians (PCPs) aided by **Legit.Health Plus** compared to their unaided assessment.

Case Set Selection

The study will use a retrospective case set enriched with a high prevalence of malignancies (Melanoma, BCC, and SCC) and representative benign "mimics" to ensure statistical power with a1:1 ratio. Furthermore, to be consistent with DEN230008, the gold standard for every case is consensus histopathology. A panel of two to five central dermatopathologists will provide the final validated diagnosis in the retrospective standalone performance study. It is important to highlight that the case set will include diverse anatomical locations and a representative range of **Fitzpatrick skin phototypes (I-VI)**.

Selection of Readers

The study intends to recruit at least 30 board-certified Primary Care Physicians (Internal Medicine or Family Practice) in order to assess the images of skin lesions suspected of

malignancy. Readers will be blinded to the histopathology results, the device output and the device's standalone performance on the cases before the study.

Sequential Read Workflow

The study workflow will consist of two phases, which will mimic the intended adjunctive use:

- **Phase I Unaided assessment:** Readers will be presented with the clinical images to analyze and the clinical information, all of this without knowing the output of the device. In this way, the reader will provide a Top-5 list of possible diagnostics ordered according to their suspicion, an initial management plan "Refer" (suggestive of malignancy) or "Monitor" (low risk). Additionally, readers will also record their diagnostic confidence on a 1-10 scale and if they consider requesting a biopsy.
- **Phase II Aided assessment:** Readers will again be presented with the clinical images in random order, but this time they will be revealed on the device output. After that, the reader is asked to provide a **second management decision**, either maintaining or revising their initial judgment based on the device output. Additionally, the reader should also provide a record of their diagnostic confidence and if they consider a biopsy necessary.

Data management

All testing will be performed using Legit.Health Plus v1.1.0.0 to ensure consistency. To prevent recall or order bias, cases may be presented in a randomized order to different readers. The primary focus is the Corrective Direction—measuring how often the device correctly changes a physician's "Monitor" decision to a "Refer" decision for a true malignancy.

Statistical Analysis Plan (SAP)

Population

The Intented-to-Treat (ITT) Population will be all the readers and cases of lesions suspected of malignancy assigned to the study. The Modified Intent-to-Treat (mITT) population will be all cases with a confirmed consensus histopathology result and completed Phase I and Phase II reader evaluations.

To confirm the Co-Primary Endpoint, Superior diagnostic Sensitivity of aided PCPs, we will use a **Generalized Estimating Equation (GEE)** model with a binary link function. This model accounts for the correlated nature of the MRMC data (multiple readers evaluating the same set of lesions. The 95% confidence interval will be calculated through the Wilson score method for the sensitivity point estimate.

The Co-Primary Endpoint, Non-random aided performance (Sens + Spec > 1), the sum of the point estimates for Aided Sensitivity and Aided Specificity will be calculated. The lower bound of the 95% Confidence Interval for the sum (calculated via bootstrap or delta method for MRMC) must be > 1.

Secondary Endpoint Analysis

- **Non-inferiority of AUROC** in aided PCPs compared to unaided PCPs. In this analysis there will be a comparison of the Area Under the Receiver Operating Characteristic (AUROC) curve between aided and unaided PCPs. The Obuchowski-Rockette (OR) or Hillis method for MRMC data will be used to estimate the difference in AUROC and the corresponding 95% CI. The lower bound of the 95% CI for the difference AUROC of aided PCPs must be greater than the pre-specified non-inferiority margin (e.g., -0.05).
- **Corrective Direction Analysis (Clinical Utility):** A McNemar's test (or GEE equivalent) will be applied to the 2x2 contingency table of management decisions to compare: 1) Correct Conversions: Unaided "Monitor" -> Aided "Refer" (for malignant cases) and 2) Incorrect Conversions: Unaided "Refer" -> Aided "Monitor" (for malignant cases), all of this in order to demonstrate a statistically significant net increase in sensitivity.
- **Subgroup analysis and Multiplicity:** Sensitivity and specificity will be descriptively summarized by Fitzpatrick skin phototypes (I-VI) and Age groups (18–39 vs. 40+). No formal adjustment for multiplicity is planned for secondary endpoints, as they are intended to provide supportive evidence. However, both co-primary endpoints must be met to claim study success.
- **Resource Utilization:** The difference in biopsy rate will be calculated between the aided and unaided PCPs in order to determine whether the use of the device helps PCPs to avoid unnecessary interventional procedures, such as biopsy.
- **Clinical confidence assessment:** The confidence scores between aided and unaided PCPs will be assessed in order to determine whether the use of the device increases PCPs confidence in their diagnosis.

Sample size justification

The sample size for the Multi-Reader Multi-Case (MRMC) study is calculated to provide sufficient statistical power to validate the co-primary endpoints and demonstrate the adjunctive benefit of Legit.Health Plus. The calculation accounts for the variability among both readers and cases using the Obuchowski-Rockette (OR) or Generalized Estimating Equations (GEE) framework.

Case and reader requirements

The study will employ a fully crossed design where every reader evaluates every case in both unaided and aided modalities. A minimum of 30 board-certified Primary Care Physicians (PCPs) will be recruited. This sample size is consistent with the predicate's reader studies and ensures a representative range of clinical decision-making across the intended user population. On the other hand, the study will utilize the same enriched dataset of 100 cases (50 malignant, 50 benign) used in the standalone performance study.

Power analysis for primary Endpoints

Co-Primary Endpoint 1: Superior diagnostic sensitivity in aided PCPs

- Null Hypothesis: Sensitivity of aided PCPs - Sensitivity of unaided PCPs $\leqq 0$.
- Alternative Hypothesis: Sensitivity of aided PCPs - Sensitivity of unaided PCPs > 0.
- Assumptions: Based on the predicate's performance, we assume an unaided PCP sensitivity of 82.0% and a predicted aided sensitivity of 87.5% (a 5.5% improvement).
- Statistical Power: With 30 readers and 200 malignant cases, the study is powered at > 80% to detect this superiority margin at a significance level of alpha = 0.05.

Co-Primary Endpoint 2: Non-random performance (Sensitivity + Specificity > 1)

- The sample size of 100/200 total cases provides high statistical power to ensure that the sum of Aided Sensitivity + Aided Specificity is statistically significantly greater than 1.

Subgroup and diversity powering

The enrichment of the case set (1:1 malignancy ratio) is specifically designed to provide enough data points for:

- Fitzpatrick Skin Phototypes (IV-VI): By over-sampling these types relative to their natural clinical prevalence, the study ensures that the 95% confidence intervals for the adjunctive benefit in dark skin tones are sufficiently narrow.
- Age Group (<40): The sample will include a targeted number of younger patients to validate that the device maintains its "Corrective Direction" benefit in populations where clinical suspicion is typically lower.

Attrition and Reliability

- **Over-recruitment:** A 15% buffer will be applied to the case set (targeting 230/115 cases total) to account for technical failures or cases where the consensus histopathology (Ground Truth) is not achieved.
- **Reader Replacement:** If a reader fails to complete both the unaided and aided phases for all cases, their data will be excluded from the mITT population, and a replacement reader will be recruited to maintain the N=30 requirement.

System Reproducibility and Robustness Study

The main aim of this study is to address a unique challenge for image-based AI devices that did not apply to the predicate device QZS. Machine learning algorithms can be susceptible to differences in image acquisition technologies until data demonstrates otherwise. In this way, this study would be smaller compared to previous studies. Dermatologists would capture images of skin lesions using multiple smartphone models. The goal of this study is to gather extensive data on a relatively small number of lesions. Each lesion should be photographed with every device under multiple conditions, generating many images per lesion. The lesion dataset will consist of lesions that the device will encounter in a clinical setting, such as melanoma, BCC, SCC and benign lesions across skin phototypes and anatomic locations. So this study aims to demonstrate that the performance of the device is not affected by the image capture device.

Hypothesis

This study hypothesizes that the diagnostic sensitivity of Legit.Health Plus will not be altered regardless of the image capture device in lesions suggestive of melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC).

Objectives

Primary objective

The main aim of this study is to demonstrate the consistency of the device output across different acquisition systems and that the diagnostic sensitivity is not altered depending on the image capture system. The aim of this study is not to establish diagnostic sensitivity, which is addressed by the two previous studies, but to ensure homogeneity of device performance.

Study design

The proposed clinical validation is an observational, cross-sectional and method-comparison study. The primary goal is to assess the consistency of diagnostic performance of the device across different acquisition systems by gathering extensive data on a relatively small number of lesions. This study will use a sequential-capture approach, where the physician will capture first a lesion with one determined device, and then they will proceed to capture another image with a different device. The objective is to capture images of a specific lesion with the investigational image capture devices and analyze the Legit.Health Plus performance with them.

The study aims to demonstrate that the device's performance is consistent with different acquisition systems since machine learning algorithms can be susceptible to differences in image acquisition technologies.

Dataset

The dataset will consist of clinical images representing the target conditions—melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC)—alongside a wide range of benign lesions (mimics) commonly encountered in dermatology, all of them taken with different image capture devices, consequently it will generate many images per lesion.

For this study, a dermatology setting will be chosen, as dermatology clinics are naturally enriched with the malignant and suspicious lesion types that would be difficult to efficiently capture in primary care. To ensure benign lesions are adequately represented, dermatologists can also capture images of non-suspicious moles and other benign findings during the study.

Ground Truth: The Ground Truth for all lesions will have be determined by consensus histopathology. Every lesion in the dataset must have a confirmed biopsy result.

Inclusion and exclusion criteria

Inclusion criteria

- Men or women of any ethnic group aged 18 and older.
- Primary skin lesions suggestive of melanoma, BCC, or SCC or benign lesions in patients aged 18 and above.
- Practitioners Board-certified Dermatologists.

Exclusion criteria

- Lesion < 2.5mm in diameter or > 15mm in diameter
- Lesion surface not accessible (e.g. inside ears, under nails, completely covered by a crust or scale)
- Lesion on the area of crust, psoriasis, eczema or similar skin condition
- Lesion has erosion and/or ulceration with no area >2.5mm intact
- Lesion has foreign matter (e.g. tattoo, splinter, dermoscopy oils, or other medicated or non-medicated topical solutions)
- Lesion in which the device tip cannot be placed entirely within the border of the targeted area
- Lesion located on acral skin (e.g. sole or palms)
- Lesion located within 1 cm of the eye
- Lesion on or adjacent to scars, areas previously biopsied, or areas subjected to any past surgical intervention
- Lesion located on mucosal surfaces (e.g. genitals, lips)
- Lesion located on an acute sunburn
- Six (6) or more lesions suggestive of melanoma, basal cell carcinoma, and/or squamous cell carcinoma requiring biopsy to assess risk of malignancy

- Dementia or other neurologic, physical or psychological limitation that would prevent the patient from signing informed consent.

Endpoints

The success of the clinical validation will be determined based on the following pre-specified endpoints. Both primary and secondary endpoints must be met to demonstrate the consistency of Legit.Health Plus across different acquisition systems and types of skin lesions.

Primary endpoint

- **System Reproducibility (Inter-Instrument Reproducibility):** This endpoint measures the consistency of the Legit.Health Plus output across multiple supported acquisition systems. The goal is to demonstrate that hardware-induced variation (e.g., sensor noise, color processing, or lens optics) does not result in clinically significant changes to the device's risk classification.
    - **Acceptance criteria:** The device must achieve a Point Estimate of $\gtrsim$ 95% Overall Percent Agreement (OPA) between different validated acquisition systems. The lower bound of the 95% Confidence Interval (CI) must be $\gtrsim$ 88%. Additionally, the Intraclass Correlation Coefficient (ICC) for the underlying similarity scores must be $\gtrsim$ 0.80, indicating "excellent" reliability across hardware platforms.

Secondary endpoint

- **Subgroup-Specific Reproducibility:** To evaluate whether the level of agreement between acquisition systems is influenced by the underlying pathology (Melanoma, BCC, SCC, or Benign Mimics). This analysis ensures that hardware-induced variability does not disproportionately affect the detection of the most clinically significant lesions.
    - **Acceptance criteria:** The Percent Positive Agreement (PPA) for the "Malignant" subgroup (Melanoma, BCC, SCC) must maintain a point estimate of $\gtrsim$ 95% across all device pairings. The **Coefficient of Variation (CV)** of the similarity scores must not vary significantly (e.g., < 10% shift) across acquisition systems for high-risk lesions.

Methodology and workflow

This is a prospective, cross-sectional, instrument-comparison study. The primary objective is to evaluate the consistency of the Legit.Health Plus output when the same lesion is captured using multiple acquisition systems (different mobile handsets and dermatoscopes).

Case Set Selection

The study follows a "Repeated Measures" design where each lesion acts as its own control across different hardware platforms. A targeted subset of at least 50 lesions will be selected prospectively. To ensure a rigorous "stress test" of the algorithm, the set will be stratified: 25% Melanoma (High-stakes detection), 25% BCC/SCC and 50% Benign Mimics (Evaluating the risk of hardware-induced false positives).

For each identified lesion, a dermatologist will capture a sequence of images using a "Primary Reference Device" (e.g., a high-end iPhone with a validated dermatoscope), followed immediately by captures using "Secondary Test Devices" (e.g., mid-range Android, various mobile handsets, and alternative dermatoscopes). Captures will be performed under controlled clinical lighting to isolate hardware-specific variables (sensor noise, color processing, and optics) from environmental variables.

Ground Truth and Reference

The "Ground Truth" for the lesion remains consensus histopathology. The diagnostic output (Similarity Score and Classification) from the Primary Reference Device will serve as the benchmark for calculating the Overall Percent Agreement (OPA) of the secondary devices.

Data Management

All analyses will be performed using Legit.Health Plus v1.1.0.0. The primary analysis will consist of the calculation of Positive Percent Agreement (PPA) and Negative Percent Agreement (NPA) across hardware pairings. Performance will be analyzed by lesion type to ensure that hardware variability does not disproportionately impact the sensitivity for Melanoma compared to BCC or SCC. The Intraclass Correlation Coefficient (ICC) will be calculated for the similarity scores to assess the stability of the numerical output across sensors.

Statistical Analysis Plan (SAP)

Population

Includes all prospectively captured lesions (N=50) that have a confirmed consensus histopathology Ground Truth and a successful image capture across all tested hardware platforms. Any capture that triggers a "low quality" or "unable to process" error by the algorithm will be documented for the robustness analysis but excluded from the primary agreement calculation.

Primary endpoint analysis

The goal of this endpoint is to demonstrate that the device output is independent of the acquisition hardware. For that reason, an analysis of Overall Percent Agreement (OPA) will be performed, and the binary classification ("Monitor" vs. "Investigate") from the Secondary Test Devices will be compared against the Primary Reference Device. The Acceptance criteria will

be A Point Estimate of ≧ 95%, with the lower bound of the 95% Wilson Score Confidence Interval ≧ 88%. For this analysis, Cohen's Kappa will be selected to measure the agreement between hardware pairings beyond chance. An acceptance criterion of kappa ≧ 0.80 (indicating "Excellent" agreement) will be applied.

Secondary endpoint analysis

This endpoint aims to demonstrate the stability of the underlying numerical "Similarity Score" across different sensors. An ICC (model (2,1)) ≧ 0.85 is expected, demonstrating that the probability score does not shift significantly due to hardware noise.

On the other hand, the robustness subgroup analysis intends to ensure hardware variation does not disproportionately affect high-stakes lesions. The PPA (Positive Percent Agreement) will be calculated specifically for the Melanoma and BCC/SCC cohorts. We will analyze if any specific hardware platform causes a statistically significant drop in sensitivity compared to the reference device.

Sample size Justification

For this study, the selected sample size will be N=50. While smaller than the MRMC study, N=50 is the industry standard for hardware robustness in dermatology AI (similar to "Pilot" reproducibility studies). The 1:1 ratio (25 malignant, 25 benign) ensures that we test agreement at the "cut-off" thresholds for both positive and negative classifications.

Finally, using a minimum of x mobile devices (representing iOS and Android ecosystems) and x dermatoscopes creates a matrix of hundreds of comparison points, providing sufficient statistical power to detect hardware dependency.

FDA questions

1. **Acceptability of Retrospective Data and Training Data Independence.** Does the Agency agree that using a retrospective, enriched dataset is an acceptable clinical validation path for Legit.Health Plus, provided that we maintain strict independence between the training/tuning data and the final clinical validation set?

2. **Proposed MRMC Study Design (Readers, Cases, and Endpoints)**. Is the proposed Multi-Reader Multi-Case (MRMC) study design—including the recruitment of at least 30 board-certified Primary Care Physicians and the use of a sequential-read methodology—sufficient to demonstrate the adjunctive benefit required under the Special Controls for product code QZS?

3. **Statistical Analysis Plan (SAP) and Power Calculations.** Does the Agency agree with the proposed Statistical Analysis Plan, specifically the use of Generalized Estimating Equation (GEE) models and the 1:1 enrichment ratio to achieve 80% power for the $\gtrsim$ 90% sensitivity performance goal?

4. **Approach to Image Acquisition System Performance**. Does the Agency concur that a prospective robustness study utilizing at least 50 lesions across various iOS and Android platforms is appropriate to demonstrate that image acquisition hardware does not adversely affect the device's diagnostic consistency?

5. **Study Population Representativeness (U.S. Intent)**. Is the proposed inclusion of patients aged 18 and older, combined with the enrichment strategy for Fitzpatrick skin phototypes IV–VI, sufficient to demonstrate that the study population is representative of the U.S. intended use population?