Optical Sensor for Photodynamic Detection of Oral Pathology
NCT NCT00540774
3/19/2026

Statistical Analysis Plan (SAP) 2002-2805

The statistical analysis breaks down into

1. **Development of a screening algorithm for OC risk based on various imaging modalities in individuals with high-risk identifiers**
2. **Assessing the screening accuracy of the screening algorithm**

Screening Algorithm

Using histopathology-based specialist diagnosis, we will classify each dual-mode image set to one of four groups: healthy, OPML (mild dysplasia and moderate dysplasia), OPSCC (severe dysplasia and cancer), and nonneoplastic lesion. Since we will capture a relatively small number of images to train the neural networks, we will use several approaches to avoid overfitting—a common issue when the network is complex, and the training data set is small. (1). Pre-trained networks and transfer learning: In our preliminary study, we demonstrated the effectiveness of transfer learning by using the pre-trained VGG-CNN-M network. We will reuse the lower levels of our existing algorithm architecture, and only retrain the top decision layers that have a much smaller number of free parameters. (2). Data augmentation: In order to increase the number of image sets to train the neural networks, we will adopt a widely used data augmentation approach that has been proven to improve the classification performance of deep learning networks. We will enrich the data provided to the classifier by flipping and rotating original images—our probe images have no natural orientation—to generate seven additional images from each original image. Thus the total number of image sets will be (200+400+400+100) x 2 x 8 =17,600, with three-quarters used to train each neural network and one-quarter to evaluate its performance. (3). Regularization: We will use penalty terms that prevent the weights of the neural networks from becoming too large during training (e.g., L2 regularization). We will also use other forms of regularization such as dropout.

Evaluate screening modality and app's screening accuracy and referral decision. Diagnostic accuracy will be tested in at least 1000 high—risk individuals, including those with smoking, vaping, e-cigarette and smokeless tobacco habits. Subjects will be recruited until at least 100 OPSCCs, 400 OPMLs, 400 non-neoplastic lesions and 100 subjects with no visible mucosal changes have been accrued. The effect of the imaging system on OPSCC screening accuracy and long-term OPSCC outcomes as measured by referral decision will be assessed by comparing non-specialist diagnoses and referral decisions with and without scanner pen system vs the gold standard (dentist diagnosis and referral decision). Biopsy-based specialist diagnosis will also be used as gold standard when available.

Statistical Considerations and Data Analysis

Under Aim 1, We will first expand the AI-driven algorithm to include non-neoplastic lesions as a fourth category. Then we will compare the algorithm-generated screening result with specialist diagnosis to estimate percent agreement, sensitivity, specificity, and positive and negative predictive values of the AI algorithm. Acceptable agreement will be defined as sensitivity ≥80% and specificity ≥65% for (1) distinguishing OPSCC from healthy, (2) distinguishing OPSCC from OPML, (3) distinguishing OPSCC from non-neoplastic leuko/erythroplakia, (4) distinguishing OPML from healthy, (5) distinguishing OPML from non-neoplastic leuko/erythroplakia. With a sample size of 1000 images for final testing (including 200 OC, 300 OPML, 400 non-neoplastic leuko/erythroplakia and 100 healthy), we will have 95% power to detect sensitivity ≥80% against the null hypothesis value of 70% using a one-sided binomial test. We will have 89% power to detect specificity ≥65% against the null hypothesis value of 50% specificity using a one-sided binomial test. For comparisons (2) through (5), we will have power ≥94% to detect sensitivity ≥80% (against the null hypothesis sensitivity=70%) and power ≥91% to detect specificity ≥65% (against the null hypothesis specificity=50%). Positive and negative predictive values (and 95% CIs) and diagnostic accuracy will also be estimated.

Under Aim 2, we will evaluate diagnostic accuracy for the algorithm in a minimum of 1000 subjects including 100 OC, 400 OPML, 400 non-neoplastic leuko/erythroplakias, 100 healthy. We will estimate sensitivity and specificity for distinguishing (1) distinguishing OPSCC from healthy, (2) distinguishing OPSCC from OPML, (3) distinguishing OPSCC from non-neoplastic leuko/erythroplakia, (4) distinguishing OPML from healthy, (5) distinguishing OPML from non-neoplastic leuko/erythroplakia, against the gold standard of supervising dentist screening decision. Acceptable accuracy for the AI algorithm will be sensitivity >80% and specificity ≥65% compared to null hypothesis values of 68% for sensitivity and 50% for specificity based on conventional diagnosis by the dental clinician. With our sample, we will have 81% power to detect sensitivity ≥80% and 91% power to detect specificity ≥65% against the null hypothesis values of 68% sensitivity and 50% specificity using a one-sided binomial test for each. For comparing OPSCC to OPML, we will have 81% power to detect sensitivity≥80% and 99% power to detect specificity≥65% against null hypothesis values. For comparisons (3)-(5), we will have 99% power to detect sensitivity ≥80% and ≥91% power to detect specificity ≥65% against null hypothesis values. 95% confidence intervals will be calculated. We will also estimate screening accuracy and positive and negative predictive values. In addition, results will be analyzed by type of tobacco used. Compliance to referral and effect of tobacco type on algorithm performance will be explored using logistic regression methods.