STATISTICAL REPORT AND ANALYSIS PLAN

Title: A Multi-Center Trial of the ProLung Test™ (Transthoracic Bioconductance Measurement) as an Adjunct to CT Chest Scans for the Risk Stratification of Patients with Pulmonary Lesions Suspicious for Lung Cancer (Protocol PL-208)

Study Device: ProLung Test, an algorithm driven transthoracic bioconductance assay that stratifies the risk of lung cancer in patients with indeterminate CT scans.

Sponsor: ProLung, 757 East South Temple, Suite 150, Salt Lake City, Utah, 84102 USA

Author: Richard P. Chiacchierini, Ph.D.

Date: January 17, 2019

Signature: Richard P. Chiacchierini Date: 1/17/2019

R.P. Chiacchierini Consulting, 17003 Horn Point Drive, Gaithersburg, MD 20878
Phone: (240) 683-3738; Fax: (301 527-0515 E-mail: rpc@rpcaconsulting.com

Analysis Summary

Study Title: A Multi-Center Trial of the ProLung Test (Transthoracic Bioconductance Measurement) as an Adjunct to CT Chest Scans for the Risk Stratification of Patients with Pulmonary Lesions Suspicious for Lung Cancer (Protocol PL-208)

Primary Objective: To demonstrate that the ProLung Test can assist in the assessment of patients presenting with indeterminant CT scans.

Primary Endpoint: The coprimary endpoints are the Positive Predictive Value (PPV) and Negative Predictive Value (NPV)

Secondary Endpoints: Co-secondary endpoints of sensitivity and specificity of the ProLung Test and the sum of PPV and NPV

Additional Endpoints: Additional Effectiveness Endpoint Analyses - Subgroup Analyses of NPV, PPV, sensitivity and specificity.

Patient Population: Males and females aged 18 years and older with a CT scan of the lungs that indicates one or more nodules or lesions suspicious for lung cancer.

Study Design: A single arm multicenter study of subjects with CT scans with nodules of undetermined significance and known diagnosis to be used to validate the ProLung Test to stratify the positive risk of lung cancer.

Number of Patients: 174 subjects with paired ProLung test result and a biopsy or radiological stability diagnosis for lung cancer.

Statistical Procedures

I. Introduction

Lung cancer is the deadliest form of cancer, with a 5-year survival rate in the United States of 16% (Jemal, Ahmedin et al. (2010)). Currently, it is difficult to detect and diagnose lung cancer early enough for treatment to be effective. New, immediate, and non-invasive methods are needed to stratify the risk of this disease in its earliest stages. Bioconductance technology of ProLung Inc. (formerly Fresh Medical Laboratories) has demonstrated the potential to meet these requirements.

The current standard of care for the detection of lung cancer involves the use of CT chest scans to evaluate patients with lesions suspicious for lung cancer. While these scans provide important information about the location and size of lung lesions and raise suspicion of lung cancer, they have limited capacity to determine immediately the malignant or benign nature of the lesions they identify. In the screening setting, between 23% to 33% of all CT screens are "positive" for an indeterminate nodule; among these 94% of non-calcified nodules are determined to be benign (Aberle et al. (2011). This presents a significant diagnostic challenge (Black (2000) and Welch and Black (2010)). Given the frequency of benign diagnosis, routine biopsy in an indeterminate lesion cannot usually be clinically justified. Because of these constraints, current guidelines (McMahon et al. (2017)) call for repeated CT scans over a period of months or years, in order to detect malignant growth in the detected lesion(s).

II. Study Design

This Study is a prospective cohort, single arm, multicenter study of subjects with indeterminate CT scans with known biopsy results or a follow-up CT scan showing radiological stability to be used to validate the ProLung Test to stratify the positive risk of lung cancer. The validation subjects are from the second phase of a two-part study to: 1) define the algorithm for cancer detection, i.e. the stabilization phase for algorithm development and cutoff determination and 2) the validation phase to evaluate the risk stratification. No subjects used in the validation portion of the study were used in the algorithm optimization.

III. Analysis Objectives

The primary Study hypothesis is that the ProLung Test will demonstrate safety and effectiveness in the PPV and NPV of patients with pulmonary lesions identified by CT that are suspicious for lung cancer. A statistically significant result will indicate that patients with a positive ProLung Test result will have a clinically relevant PPV.

IV. Analysis Populations

A. Intent-to-Treat (ITT) Population

The intent-to-treat population is every subject that is enrolled in the study.

B. Completed Cases (CC) Population

All subjects that have paired results of both ProLung Test and either confirmatory tissue biopsy (of cancer or benign condition) or radiologic stability to infer a benign diagnosis.. The primary and secondary analyses require paired evaluations and therefore will be conducted in the completed cases population.

V. Definition of Study Outcomes

A. Primary Effectiveness Endpoint

The coprimary effectiveness endpoints are PPV and NPV of the ProLung Test. The null and alternative hypotheses for PPV is presented below.

H_0: PPV $\leq PG_{Pos}$ versus H_a: PPV $\geq PG_{Pos}$

where PPV is the estimated Positive Predictive Value and PG_{Pos} is the performance goal for the PPV developed below.

The null and alternative hypotheses for NPV is presented below.

H_0: NPV $\leq$P G_{Neg} versus H_a: NPV $\geq PG_{Neg}$

where NPV is the estimated Negative Predictive Value and PG_{Neg} is the performance goal for the NPV developed below.

Each hypothesis above is tested with an exact binomial test. An alternative interval test will also be presented to demonstrate that the lower 95% confidence limit of the point estimate of PPV and NPV will lie above their respective performance goals.

1. Performance Goal Derivation and Sample Size

From the algorithm development data, the observed rate for PPV = 0.861, and for NPV =0.471. Note that this PPV and NPV estimate are based on a prevalence of 0.74 from the algorithm development cohort. Since the prevalence of the validation cohort has a lower prevalence (approximately 0.637), the PPV will be lower and the NPV will be higher. The estimate for PPV is 0.778 and for NPV is 0.551. The performance goals will be computed based on these estimates of PPV and NPV.

Recall that the denominator of the PPV is the number of total tests that were deemed positive from the ProLung Test. Likewise, the NPV computation uses as its denominator the total number of ProLung Tests that turned out negative. In the 200 patients used in the algorithm study, there are 148 positive patients and 52 negative patients 109/200 (54.5%) were true positives (TP), 39/200 (19.5%) were false negatives (FN), 16/200 (8.0%) were false positives (FP), and 36/200 (18.0%) were true negatives (TN). In the validation study, there are 111 patients with malignant lesions and 63 patients with benign lesions. The performance goals are computed below that are achievable with this distribution of positive and negative patients.

Statistical simulation was performed to determine the power and estimates of PPV and NPV for the validation cohort. From that simulation an average PPV = 0.778 and NPV=0.552. This results in the following estimates: 78 TP, 33 FN, 22 FP, and 41 TN cases. The number of positive test results are TP+FP=100, and the total number of negative test results is FN+TN=74. To achieve approximately 90% power for PPV with 111 positive test results and a point estimate of 0.778 will require PG_{Pos}=0.638. To achieve about 90% power for NPV with the 63 negative test results and a point estimate for NPV of 0.551, the PG_{Neg}=0.381. Note that 90% power is required because the coprimary status indicates that both tests must be achieved so the power of the joint hypothesis is 0.90*0.90=0.81, the usual power required for the tests. The power for the sum of PPV and NPV from the 10-simulation averaged 0.99.

B. Secondary Effectiveness Endpoints

There are two secondary endpoints: the estimation of sensitivity and specificity of the ProLung Test and the statistical analysis of the sum of PPV and NPV to demonstrate that the sum exceeds unity.

The co- secondary effectiveness endpoints sensitivity and specificity of the ProLung Test. There will be no hypothesis tests of these endpoints. Each will be presented with point estimates and 95% exact binomial confidence limits.

The sum of PPV and NPV will be tested with the null and alternative hypothesis below.

H_0: PPV+NPV≤1 versus H_a: PPV+NPV>1.

C. Primary Safety Endpoint

Since the device is not used in any patient treatment decisions in this study, there is no risk to the study subjects other than non-serious skin reactions to electrode patches used in testing. Therefore, this is a non-significant risk study. However, the frequency of occurrence of any adverse events will be reported to the extent these occur in the study.

VI. Multiplicity

There are two coprimary hypothesis to be tested in this study, the descriptive analysis of the secondary endpoints using statistical estimation to provide exact 95% confidence limits for sensitivity and specificity, and the hypothesis test of the sum of PPV and NPV will preserve the nominal alpha level. Using the hierarchical closed form method to control alpha inflation, allows the descriptive presentation for the co-secondary endpoints of sensitivity and specificity and can use the nominal study alpha to present two-sided 95% confidence limits and the hypothesis test of the sum at the nominal alpha level as long as both co-primary hypothesis test results in statistical significance. The coprimary endpoints must both be statistically significant so there is no alpha inflation for the two coprimary hypothesis tests.

VII. Character of Study Variables

As a routine evaluation of the data during analysis, consistency of the study variables to properties of statistical tests will be done. For continuous variables, equality of variance tests will be done to support analyses requiring this condition such as the Wilcoxon rank sum test. With categorical data, exact statistical test procedures will be used to minimize test assumptions. For tabulated continuous variables, the descriptive analyses will present the mean, standard deviation, median, minimum and maximum. For tabulated categorical variables, the number with the characteristic, the total number evaluated, the percent and the 95% exact binomial confidence intervals will be provided.

VIII. Comparability Analyses of the Patient Populations

Baseline characteristics for the subjects in the validation study will be presented descriptively. Quantitative variables will have the mean, standard deviation (SD), number evaluated, median, minimum and maximum presented. Qualitative variables will have the number with the characteristic, the number evaluated, the percentage, and the exact 95% confidence limits on the percentage presented.

In addition, an analysis of comparability across study sites will be carried out. For continuous variables, this will be done by parametric or non-parametric analysis of variance. For categorical variables, the Fisher-Freeman-Halton test will be used. Study site differences do not disallow pooling, but variables including study site and the variable found different need to be considered as covariates in subsequent multivariate analyses.

Because some study sites will have too few subjects to permit an adequate evaluation, it will be necessary to combine study sites with less than 8 subjects together to form pseudo-sites that will have adequate numbers of subjects to support analyses by pseudo-site. The combination will be done in an unbiased manner without regard to study results. The study site with the lowest site number with fewer than 8 subjects will be combined with the next site in numerical order with fewer than 8 subjects. This process will be continued until the pseudo-site formed has 8 or more subjects. Study sites with 8 or more subjects will be assigned a pseudo-site number to allow analysis. No pseudo-site created by this process will have a greater number of subjects than the median of the sites that have 8 or more subjects. Pseudo-sites will also be used for the site comparability analyses and the data pooling analysis below.

IX. Patient Accountability and Missing Data

A summary table will provide the total number of subjects enrolled and completed. Subjects withdrawn will be tabulated with their reasons for withdrawal and subjects missing either a diagnosis result or ProLung Test result will be tabulated with which of the two evaluations is missing.

Every effort will be made to collect all data points in the study. The sponsor plans to minimize the amount of missing data by appropriate management of the prospective clinical trial, proper screening of study subjects, and training of participating investigators, monitors and study coordinators. The sponsor will provide a list of subjects who do not complete the trial along with the best information available on why each had missing data.

The number of subjects with missing data is expected to be a small proportion of total subjects in the study. If the diagnosis result is missing it is not possible from subject characteristics to estimate or impute a result for any given subject. Likewise, if the ProLung Test result is missing it is not possible to determine from baseline subject characteristics to estimate or impute the missing result of the test. Thus, these subjects will not be used in the analysis. A descriptive comparison of the baseline characteristics of the ITT and CC subjects will be presented to demonstrate that there is little bias from the missing data.

A sensitivity analysis will be done for cases with missing data that assigns cases with missing results the worst-case scenario, i.e. that the ProLung result is negative if cancer is detected or that the ProLung result is positive when cancer is not detected. A second analysis will be done as a best-case scenario, i.e., the ProLung result will be deemed positive when cancer is present and will be deemed negative when cancer is not detected.

X. Effectiveness Analyses

A. Primary Effectiveness

The data from all study sites or pseudo-sites will be pooled and the primary analysis will be done by a one-sided exact binomial test of the performance goal for PPV and NPV win nominal P-value of 0.025. Alternatively, the lower one-sided 97.5% confidence limit for PPV and NPV will be computed and compared to each performance goal. If these limits both are greater than the performance goal, then the null hypothesis will be rejected.

An exploratory descriptive analysis will be provided for PPV and NPV by study site or pseudo-site. The point estimates and 95% two-sided confidence limits will be presented by study site or pseudo-site. Recall that because no site was intended to have adequate power for the primary endpoint, this analysis is unlikely to show that by-site lower confidence limit for PPV and NPV will be above the performance goals.

B. Secondary Effectiveness

The first secondary endpoints, co-secondary endpoints of sensitivity and specificity of the ProLung Test will be estimated. Sensitivity is defined as the proportion of subjects with ProLung Tests that are positive among subjects with positive biopsy results. Specificity is defined as the proportion of subjects with ProLung tests that are negative among the subjects with a negative biopsy or CT follow-up result. Each of these endpoints will be estimated from the data and presented with their 95% exact confidence intervals.

The second secondary endpoint the test of the sum of PPV and NPV will be tested with a continuity corrected z-statistic presented below. An estimate is needed for the standard error of the sum of PPV and NPV, because PPV and NPV are unlikely to have equal variances, the standard error of the sum can be estimated by the following expression for linear combinations of items that have unequal variances.

$$SE(PPV + NPV) = \sqrt{\frac{PPV(1-PPV)}{TP+FP} + \frac{NPV(1-NPV)}{TN+FN} + \rho\sqrt{\frac{PPV(1-PPV)}{TP+FP}}\sqrt{\frac{NPV(1-NPV)}{TN+FN}}}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, FN is the number of false negatives, and ρ is the estimate of the correlation coefficient between PPV and NPV. Note that the PPV has a negative correlation with NPV so ρ is negative reducing the standard error than if PPV and NPV were independent.

The test statistic is given by the formula below.

$$z = \frac{(PPV+NPV-\frac{1}{2}\left(\frac{1}{TP+FP}+\frac{1}{TN+FN}\right)-1)}{SE(PPV+NPV)}.$$

If z > 1.96 (P<0.025), then the sum of PPV and NPV will be considered to be greater than unity. Alternatively, if the lower one-sided 97.5% confidence limit of the sum is greater than 1, the null hypothesis will be rejected. The lower confidence limit is computed by the formula below:

$$LCL_{0.975} = (PPV + NPV) - 1.96 * SE(PPV + NPV).$$

C. Additional Effectiveness Endpoint Analyses - **Subgroup Analyses of Sensitivity and Specificity**

Descriptive presentation of sensitivity and specificity by selected subgroups will be presented. The subgroup variables and segments are presented below.

Age (<50, 50 to 59, 60-69, and ≥70)

Gender (Male and Female)

BMI (<20, 20-25, 26-30, ≥30).

XI. Safety Analyses

The adverse events in this study will be displayed descriptively. The number of subjects with at least one event, the total number of subjects evaluated, the percentage, the 95% confidence limits on the percentage, and the total number of events of a given type are to be presented.

XII. Scan Eligibility for Analysis

ProLung Tests are excluded from analysis for quality test issues which may include missing patient measurement points, instances when the operator did not follow device training protocol, and other quality issues. The duration of the scan more than 40 minutes will not be an exclusion factor.

XIII. Statistical Software

The primary analyses will be done using SAS, Version 9.4 or later for Personal Computers. The Fisher's exact tests, exact 95% confidence limits, and other categorical data computations will be done with StatXact for windows (Version 8 or later). Some preliminary descriptive analyses and figures may be done with Minitab Version 17 or later.

XIV. References:

1. Jemal Ahmedin et al. (2010). Cancer Statistics 2010, 60 Cancer Journal for Clinicians 5.
2. Aberle, D. et al., National Lung Screening Trial Research Team. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening, *N. Engl. J. Med*. 365: 395–409.

3. Black WC. (2000). Over diagnosis: an under recognized cause of confusion and hphase in cancer screening. *J of NCI* 92 (16): 1280-1282.
4. Welch HG, WC Black. (2010). Over diagnosis in cancer. *J Natl Cancer Inst* 102 (9): 605-613.
5. MacMahon et al. (2017). Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner society. *Radiology* 284; July 228-243.
6. Meinert, C, 1986. Clinical Trials: Design, Conduct, and Analysis. Oxford University Press, New York.
7. Breslow, N, and N. Day. (1980). The Analysis of Case-Control Studies. *IARC Scientific Publications No. 32.* Lyon, France.
8. Harrell, F. K. Lee, and D. Mark. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions, and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15:361-387.
9. Peduzzi, P., J. Concato, E. Kemper, T. Holford, and A. Feinstein. (1996). A simulation study of the number of events per variable in logistic regression. *Journal of Clinical Epidemiology,* 99:1373-1379.