

Statistical Analysis Plan

ISRCTN Registry — Submission Document

Expression of CUG2, β IGH3 (TGFBI), and TGF- β in Bronchoscopic Biopsy Specimens of Lung Cancer Patients: Predictive Value for Metastasis and Differential Diagnosis from Other Thoracic Malignancies — A Pilot Study

Field	Details
Corresponding Author	Dr. Önder Öztürk
Institution	Süleyman Demirel University Faculty of Medicine, Isparta, Turkey
Department	Department of Pulmonology
Ethics Approval	Decision No. 239, 13 December 2018
Grant No.	TDK-2019-6958
Study Period	January 2019 – January 2022
Date of SAP	June 2026 (retrospective)
Version	1.0

Note: This statistical analysis plan is submitted retrospectively to the ISRCTN registry. All analyses described herein were pre-specified in the study protocol approved by the ethics committee and have been fully reported in the accompanying manuscript.

1. Study Overview

This pilot study prospectively enrolled 49 patients (final analytical cohort) who underwent diagnostic fiberoptic bronchoscopy (FOB) for suspected primary lung cancer or thoracic malignancy at Süleyman Demirel University Faculty of Medicine between January 2019 and January 2022. The study evaluated the diagnostic accuracy of three biomarkers — CUG2, β IGH3 (TGFBI), and TGF- β — measured by quantitative reverse transcription PCR (qRT-PCR) in bronchoscopic biopsy specimens.

2. Study Objectives

2.1. Primary Objective

- To evaluate the diagnostic accuracy of CUG2, β IGH3 (TGFBI), and TGF- β gene expression for predicting metastatic status at the time of lung cancer diagnosis (lung cancer subcohort, $n = 36$).

2.2. Secondary Objectives

- To assess biomarker utility for differential diagnosis of primary lung cancer from other thoracic malignancies (full cohort, $n = 49$).
- To examine correlations between biomarker expression levels and between biomarkers and clinical variables.
- To evaluate the prognostic significance of biomarker expression for overall survival.
- To compare biomarker expression across histopathological subtypes (SCC, adenocarcinoma, SCLC).

3. Analysis Populations

Population	Definition	n	Used For
Full Analytical Cohort	All enrolled patients after exclusions	49	Differential diagnosis; demographics
Lung Cancer Subcohort	Histopathologically confirmed primary lung cancer	36	Metastasis prediction; survival; subtype
Met+ Subgroup	Metastatic disease at diagnosis (PET/CT)	26	ROC; between-group
Met- Subgroup	No metastatic disease at diagnosis	10	ROC; between-group
Comparator Group	Other thoracic malignancies	13	Differential diagnosis

4. Statistical Software

Software	Version	Purpose
IBM SPSS Statistics	26.0	Descriptive statistics; between-group tests; survival
MedCalc Statistical Software	20.0	ROC curve analysis; AUC; bootstrap CI
Python (SciPy, scikit-learn, lifelines, scikit-posthocs)	3.10	Non-parametric tests; correlation; survival curves; FDR correction

5. General Statistical Principles

- All hypothesis tests were two-sided.
-
- 95% confidence intervals (CIs) were reported for all primary estimates.
- Bootstrap resampling (2000 iterations) was used for AUC and sensitivity/specificity CIs.
- Multiple testing correction: Benjamini–Hochberg false discovery rate (FDR) procedure applied to primary biomarker between-group comparisons; adjusted q-values reported alongside unadjusted p-values.
- No multiplicity correction was applied to secondary analyses (correlations, ROC pairwise comparisons, survival, subtype analyses), which were interpreted as exploratory.
- Given the pilot nature of the study, all results should be interpreted descriptively.

6. Descriptive Statistics

- Continuous normally distributed variables: mean $\pm$ standard deviation (SD). Normality assessed by Shapiro–Wilk test.
- Continuous non-normally distributed variables: median (interquartile range, IQR).
- Categorical variables: n (%).
- Biomarker expression values (CUG2, β IGH3/TGFBI, TGF- β): non-parametric methods used throughout (all distributions non-normal; Shapiro–Wilk $p < 0.001$).

7. Primary Analysis: Metastasis Prediction

7.1. Between-Group Comparison

- Test: Mann–Whitney U test (two independent groups: Met+ vs. Met–).
- Effect size: rank-biserial correlation ($r = |Z|/\sqrt{N}$).
- Applied to: CUG2, β IGH3 (TGFBI), TGF- β in lung cancer subcohort ($n = 36$).
- FDR correction (Benjamini–Hochberg) applied to the three p-values.

7.2. Diagnostic Accuracy (ROC Analysis)

- ROC curve analysis for each biomarker separately.
- AUC reported with 95% CI (bootstrap, 2000 iterations).
- Optimal cut-off: Youden Index ($J = \text{sensitivity} + \text{specificity} - 1$).
- At optimal cut-off: sensitivity, specificity, PPV, NPV, overall accuracy reported.
- Sensitivity and specificity CIs: bootstrap resampling (2000 iterations).
- Pairwise AUC comparisons: bootstrap-based test (2000 iterations); Δ AUC with two-sided p-value.
- Reference: AUC = 0.5 (no discrimination).

8. Secondary Analyses

8.1. Differential Diagnosis (Full Cohort, $n = 49$)

- Between-group comparison: Mann–Whitney U test (lung cancer vs. other malignancies).
- ROC curve analysis: AUC with 95% CI; Youden-optimal cut-off; sensitivity, specificity, PPV, NPV.
- FDR correction applied to primary between-group p-values.

8.2. Correlation Analysis

- Biomarker–biomarker correlations: Spearman rank correlation coefficient (ρ) with two-sided p-value.
- Biomarker vs. metastatic status (binary): point-biserial correlation framework.
- Biomarker vs. continuous clinical variables (age, pack-years, follow-up): Spearman ρ .

8.3. Survival Analysis

- Overall survival (OS): time from bronchoscopic biopsy to death from any cause; patients alive at last follow-up censored.
- Kaplan–Meier survival curves estimated for biomarker expression groups (above/below Youden cut-off) and metastatic status subgroups.
- Log-rank test for between-group survival comparison.
- Univariate Cox proportional hazards regression: hazard ratio (HR) with 95% CI for each prognostic variable.
- Multivariable Cox models: not fitted (limited events, $n = 20$).

8.4. Histopathological Subgroup Analysis

- Kruskal–Wallis H test for overall between-subtype comparison (SCC, adenocarcinoma, SCLC).
- Post hoc pairwise comparisons: Dunn’s test with Bonferroni adjustment.
- SCLC subgroup (n = 4): analyses considered exploratory; interpreted with caution.

9. Sample Size and Power

No formal a priori sample size or power calculation was performed. No published data were available for CUG2 expression in bronchoscopic biopsy specimens at the time of study design, precluding effect-size-based power analysis. The study was sized based on clinical feasibility within a single center over a four-year enrollment period.

Post hoc power analysis: 20 observed survival events (deaths) provided approximately 28% power to detect a hazard ratio of 2.0 at $\alpha = 0.05$ in univariate Cox regression, confirming that survival analyses are substantially underpowered and should be interpreted as exploratory. All other analyses are similarly hypothesis-generating.

10. Missing Data

- Biomarker data: specimens with insufficient RNA yield (A260/A280 outside 1.7–2.0) excluded prior to analysis (n = 4).
- Smoking status: unavailable for 2 lung cancer patients; ever-smoker analysis performed on 34 patients with known smoking status; denominator clearly stated.
- No imputation of missing data was performed.

11. Reporting

- Results reported in accordance with STARD 2015 guidelines (Bossuyt et al., BMJ 2015;351:h5527).
- CONSORT/STROBE not applicable (diagnostic accuracy pilot study).
- All AUC values reported with 95% CI. All HRs reported with 95% CI.
- Exact p-values reported to three decimal places; $p < 0.001$ reported as such.
- Effect sizes reported for all between-group comparisons.

Approved by:

Dr. Önder Öztürk — Principal Investigator

Department of Pulmonology, Süleyman Demirel University Faculty of Medicine

Isparta, Turkey

Date: June 2026