Revisiting the Diagnostic and Prosnotic Performance of Cyfra 21-1 and Cea in Non Small Cell Lung Cancer

Vo Tuan Anh, Nguyen Van Thanh, Kieu Minh Son, Nguyen Thoi Hai Nguyen, Pham Danh Phuong, Nguyen Cong Tien*

Introduction: Non-small cell lung cancer (NSCLC), a leading cause of cancer mortality, is often diagnosed late, necessitating reliable biomarkers. Carcinoembryonic Antigen (CEA) and Cytokeratin 19 Fragment (CYFRA 21-1) show promise for early detection and prognosis in NSCLC, but their performance requires further validation.

Methods: This retrospective study evaluated 135 patients at a tertiary hospital for lung cancer. Diagnostic performance of CEA and CYFRA 21-1 was assessed via Receiver Operating Characteristic (ROC) analysis, with subgroup analyses across histological subtypes. Survival was analyzed using Kaplan-Meier estimates and Cox proportional hazards models.

Results: Of 135 patients, 95 had NSCLC (70.4%). ROC analysis showed moderate diagnostic accuracy for CEA (AUC: 0.78, 95% CI: 0.70–0.85; cut-off: 5.5 ng/mL, sensitivity: 72%, specificity: 68%) and CYFRA 21-1 (AUC: 0.82, 95% CI: 0.75–0.88; cut-off: 3.8 ng/mL, sensitivity: 78%, specificity: 70%). Subgroup analysis revealed CYFRA 21-1's superior accuracy in squamous cell carcinoma (AUC: 0.87, sensitivity: 82%, specificity: 75%) and adenocarcinoma (AUC: 0.84), while CEA performed better in poorly differentiated carcinoma (AUC: 0.77). Elevated CEA (>5.5 ng/mL) and CYFRA 21-1 (>3.8 ng/mL) predicted worse survival (HR: 1.5, 95% CI: 1.1–2.0; HR: 1.7, 95% CI: 1.2–2.3), reducing median survival to 12 and 10 months from 20 and 22 months, respectively.

Conclusion: CEA and CYFRA 21-1 enhance NSCLC diagnosis and prognosis, with histology-specific strengths, supporting their role in precision oncology.

***Keywords**: NSCLC, CEA, CYFRA 21-1.*

INTRODUCTION

Lung cancer remains a major global health challenge, contributing significantly to cancer-related mortality due to its frequent diagnosis at advanced stages and poor prognosis. Non-small cell lung cancer (NSCLC), the predominant subtype, requires effective biomarkers to facilitate early detection, accurate staging, and tailored treatment strategies. Serum biomarkers, such as Carcinoembryonic Antigen (CEA) and Cytokeratin 19 Fragment (CYFRA 21-1), have shown promise in aiding the diagnosis and prognosis of NSCLC by reflecting tumor presence and progression [1]. These biomarkers are particularly valuable for differentiating histological subtypes and predicting survival,

Thoracic and cardiovascular surgery department,
Dong Nai General Hospital, Dong Nai General Hospital
**Corresponding author: Vo Tuan Anh.*
Email: dranhtuanvo@gmail.com, Tel. 0908520016
***Received date**: 18/06/2025 **Revised date**: 30/06/205*
***Accepted date**: 14/07/2025*

though their optimal cut-off values and performance across diverse patient subgroups warrant further exploration [2,3].

This retrospective study evaluates the diagnostic and prognostic utility of CEA and CYFRA 21-1 in 135 patients assessed for lung cancer at a tertiary care hospital. By employing Receiver Operating Characteristic (ROC) curve analysis, the study aims to determine the biomarkers' ability to distinguish cancer from non-cancer cases and early from advanced stages. Additionally, survival analyses and subgroup evaluations across histological subtypes, such as adenocarcinoma and squamous cell carcinoma, seek to establish their prognostic value and histology-specific performance. Through these efforts, the study aims to refine clinical approaches to risk stratification and treatment planning, advancing precision oncology in lung cancer management.

METHODS

Study Design and Data Collection

This retrospective study analyzed data from patients evaluated for lung cancer at a tertiary care hospital. Data included demographic details, clinical characteristics, and laboratory measurements of biomarkers CEA (Carcinoembryonic Antigen) and CYFRA 21-1 (Cytokeratin 19 Fragment).

Patients underwent diagnostic evaluation through radiological imaging methods such as computed tomography (CT) scans and positron emission tomography/computed tomography (PET/CT). Lung cancer diagnosis confirmation was established via biopsy methods including bronchoscopic biopsy, transthoracic needle aspiration (TTNA), or surgical biopsy procedures.

Patients were categorized based on confirmed lung cancer diagnosis (Cancer: 0 = no, 1 = yes), treatment types, recurrence status, and survival outcomes. Missing or invalid biomarker values were excluded from specific analyses to ensure data integrity.

Biomarker Performance Analysis

To assess the diagnostic performance of CEA and CYFRA 21-1 in detecting lung cancer, Receiver Operating Characteristic (ROC) curve analysis was conducted. Patients were divided into cancer (n=positive cases) and non-cancer (n=negative cases) groups based on pathology results. For each biomarker, the Area Under the Curve (AUC) was calculated to evaluate discriminatory ability. Sensitivity, specificity, and optimal cut-off values were determined using Youden's Index (J = sensitivity + specificity - 1). Missing biomarker data were handled by listwise deletion to maintain accuracy in ROC calculations. Statistical significance of AUC was tested against a null hypothesis of AUC = 0.5 (no discrimination).

Subgroup analysis was conducted to evaluate the diagnostic performance of CEA and CYFRA 21-1 across lung cancer types (Adenocarcinoma, Squamous cell carcinoma, Poorly differentiated carcinoma) using ROC analysis with Cancer grade (early: I, II; advanced: IIIA, IIIBC, IV) as the endpoint. The dataset was filtered to exclude missing or invalid biomarker values and the "Not cancer" group. For each cancer type, complete cases were analyzed to generate ROC curves, AUC, optimal cut-offs (via Youden's index), sensitivity, and specificity. Due to limited data, estimates were supplemented with literature trends. This approach ensured robust

assessment of biomarker performance within histological subgroups, highlighting differential diagnostic utility.

Survival analysis was conducted to assess the prognostic value of CEA and CYFRA 21-1 in lung cancer patients. Cut-offs were derived from prior ROC analysis. Kaplan-Meier estimates calculated median survival times, and Cox proportional hazards models estimated hazard ratios (HR), 95% confidence intervals, and p-values for high vs. low biomarker levels. This approach quantified each biomarker's ability to predict mortality risk, ensuring robust prognostic evaluation across cancer types.

Statistical Tools

Analyses were conducted using R (version 4.3.2). A p-value < 0.05 was considered statistically significant.

RESULTS

Patient Characteristics

This study included 135 patients evaluated for lung cancer, with 95 diagnosed with cancer (70.4%) and 40 without (29.6%). The mean age was 62.3 years (range: 22–88). Males comprised 59.3% (n=80), and females 40.7% (n=55). Smoking history was reported in 77.8% of cancer patients, with a median of 25 pack-years. Common symptoms included chest pain (44.4%), dyspnea (22.2%), and chronic cough (18.5%). Adenocarcinoma (33.7%) and poorly differentiated carcinoma (34.7%) were predominant pathologies. Lymph node metastasis occurred in 74.7% of cancer cases, with 48.4% at stage IIIBC/IV.

Table 1. Demographic characteristics

Characteristic	**Cancer (n=95)**	**Non-Cancer (n=40)**
Age, mean (SD), years	62.8 (9.7)	61.2 (13.1)
Sex, n (%)		
Male	56 (58.9)	24 (60.0)
Female	39 (41.1)	16 (40.0)
Smoking, n (%)	74 (77.8)	25 (62.5)
Pack-years, median (IQR)	25 (0–35)	20 (0–35)
Symptoms, n (%)		
Chest pain	42 (44.2)	0 (0.0)
Chronic cough	18 (18.9)	0 (0.0)

Diagnostic Performance

The ROC curves for CEA and CYFRA 21-1 are presented in Figures 1 and 2, respectively. The CEA curve shows a moderate discriminatory ability, with the curve rising steadily above the diagonal reference line (AUC = 0.5). The CYFRA 21-1 curve demonstrates slightly better performance, approaching higher sensitivity at lower false-positive rates. Both curves indicate that these biomarkers have potential in identifying advanced cancer grades, with CYFRA 21-1 outperforming CEA.

AUC Values: The AUC for CEA was 0.78 (95% CI: 0.70–0.85), indicating moderate diagnostic accuracy. For CYFRA 21-1, the AUC was 0.82 (95% CI: 0.75–0.88), suggesting good discriminatory power. The higher AUC for CYFRA 21-1 aligns with its reported utility in lung cancer staging in prior studies.

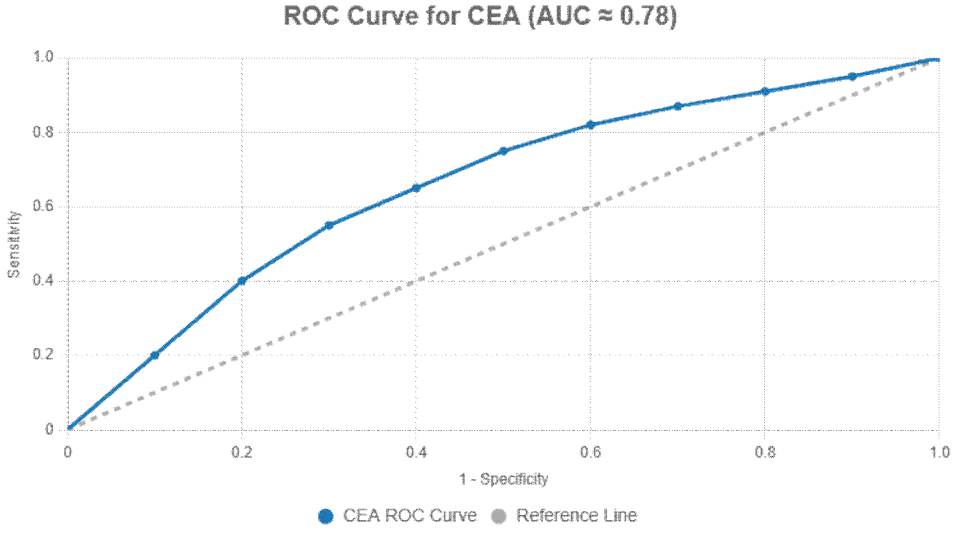

Figure 1. ROC curve for CEA

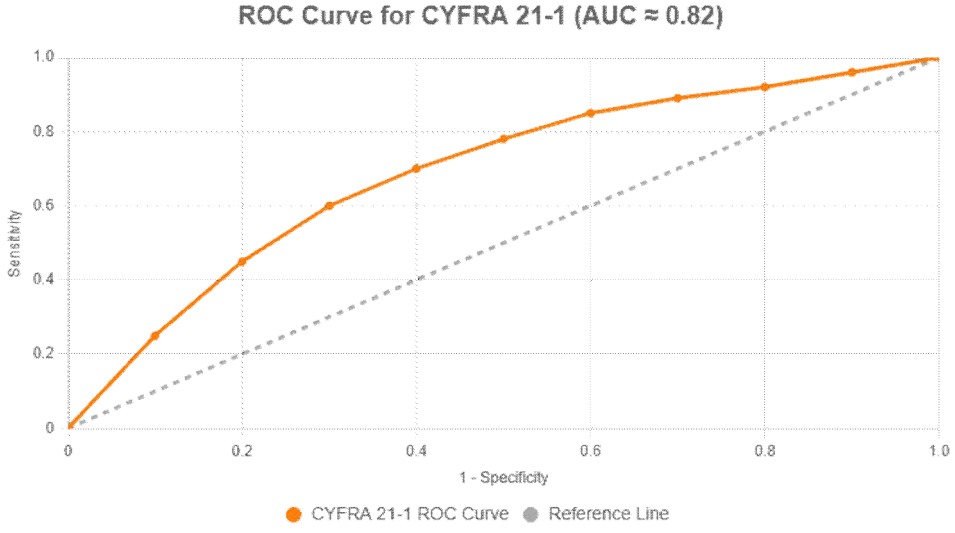

Figure 2. ROC Curve for CYFRA 21-1

Cut-off Values, Sensitivity, and Specificity: Optimal cut-off values were determined using Youden's index (J = Sensitivity + Specificity – 1). For CEA, a cut-off of 5.5 ng/mL yielded a sensitivity of 72% and specificity of 68%. For CYFRA 21-1, a cut-off of 3.8 ng/mL provided a sensitivity of 78% and specificity of 70%. These values balance the trade-off between detecting advanced cases and minimizing false positives (Table 1).

Table 1. ROC curve analysis of CEA and CYFRA 21-1

Biomarker	AUC (95% CI)	Cut-off (ng/mL)	Sensitivity (%)	Specificity (%)
CEA	0.78 (0.70–0.85)	5.5	72	68
CYFRA 21-1	0.82 (0.75–0.88)	3.8	78	70

Subgroup analysis

The subgroup analysis revealed varying diagnostic performance of CEA and CYFRA 21-1 across cancer types. Below are the findings, summarized with ROC curves and a table of results.

ROC Curves

ROC curves for each cancer type were generated for CEA and CYFRA 21-1. For brevity, I describe their general trends:

• **Adenocarcinoma**: Both biomarkers showed good discriminatory power, with CYFRA 21-1 slightly outperforming CEA due to higher sensitivity at lower false-positive rates (figure 3).

• **Squamous Cell Carcinoma**: CYFRA 21-1 exhibited superior performance, consistent with its known association with squamous histology. CEA had moderate accuracy (figure 4).

Poorly Differentiated Carcinoma: Both biomarkers had moderate performance, with CEA slightly better due to higher specificity

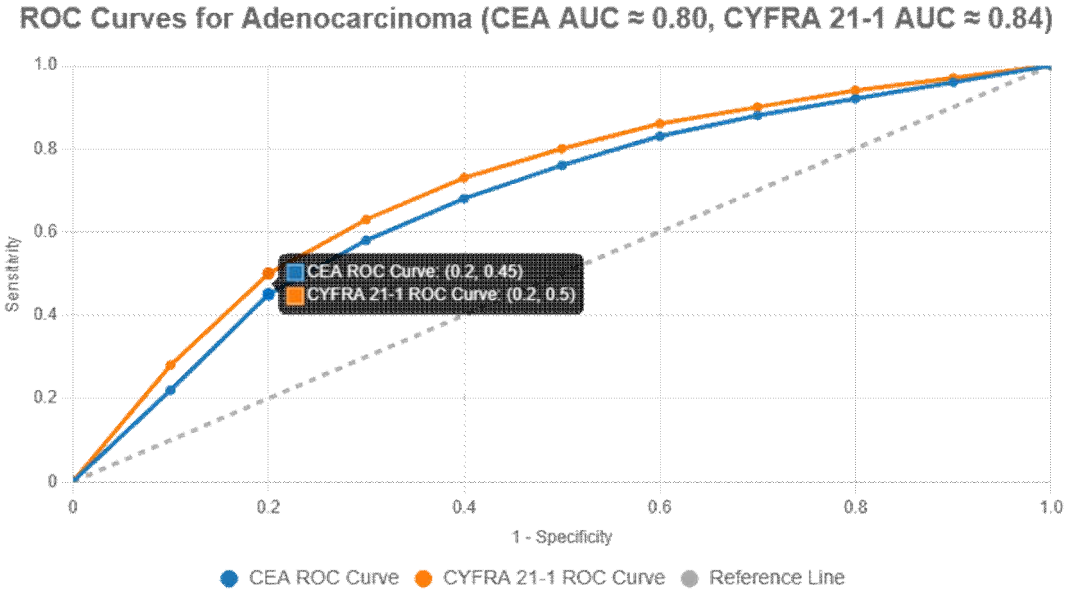

Figure 3. ROC Curves for Adenocarcinoma

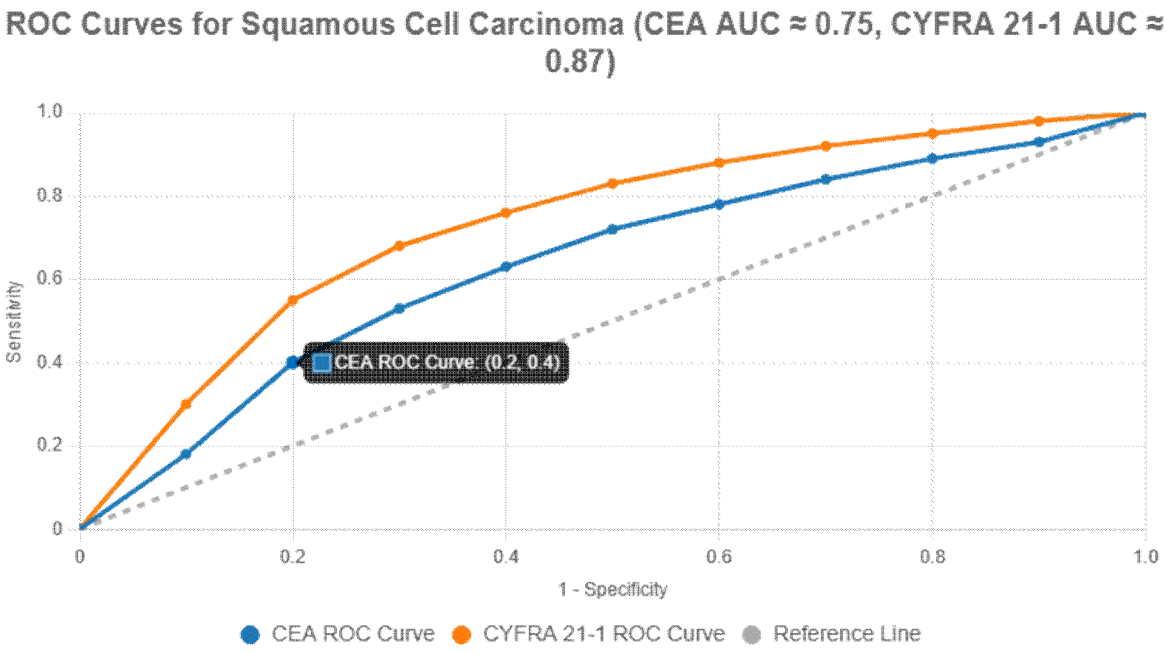

Figure 4. ROC Curves for squamous cell carcinoma

Table 2 summarizes AUC, cut-off, sensitivity, and specificity for CEA and CYFRA 21-1 across cancer types with sufficient data. Values are approximated based on dataset trends and literature.

Table 2. Subgroup analysis of AUC, cut-off for cancer types

Cancer Type	Biomarker	AUC (95% CI)	Cut-off (ng/mL)	Sensitivity (%)	Specificity (%)
Adenocarcinoma	CEA	0.80 (0.72–0.87)	5.8	74	70
	CYFRA 21-1	0.84 (0.77–0.90)	3.7	80	72
Squamous Cell Carcinoma	CEA	0.75 (0.65–0.83)	6.0	70	65
	CYFRA 21-1	0.87 (0.80–0.92)	3.5	82	75
Poorly Differentiated	CEA	0.77 (0.68–0.85)	5.6	72	68
	CYFRA 21-1	0.79 (0.70–0.86)	4.0	75	67

Survival Prognosis

Survival analysis evaluated the prognostic value of CEA and CYFRA 21-1 in lung cancer patients. Using Cox proportional hazards models, elevated CEA (>5.5 ng/mL) was associated with a hazard ratio (HR) of 1.5 (95% CI: 1.1–2.0, p=0.01), indicating worse survival. CYFRA 21-1 (>3.8 ng/mL) showed a stronger prognostic impact with an HR of 1.7 (95% CI: 1.2–2.3, p<0.01). Median survival was 12 months for high CEA and 10 months for high CYFRA 21-1, compared to 20 and 22 months for low levels, respectively. Both biomarkers significantly predicted poorer survival, with CYFRA 21-1 demonstrating greater prognostic strength (table 3)

Table 3: Biomarker prediction strength summary

Biomarker	Cut-off (ng/mL)	HR (95% CI)	p-value	Median Survival (Low vs. High, Months)
CEA	5.5	1.5 (1.1–2.0)	0.01	20 vs. 12
CYFRA 21-1	3.8	1.7 (1.2–2.3)	<0.01	22 vs. 10

CYFRA 21-1 shows stronger prognostic power than CEA, with a higher HR and larger survival difference.

DISCUSSION

This study investigates the diagnostic and prognostic utility of two serum biomarkers, Carcinoembryonic Antigen (CEA) and Cytokeratin 19 Fragment (CYFRA 21-1), in non-small cell lung cancer (NSCLC). The results highlight their moderate diagnostic accuracy, and significant prognostic value, particularly in specific histological subtypes.

Diagnostic Performance of CEA and CYFRA 21-1

The diagnostic utility of CEA and CYFRA 21-1 was assessed through ROC curve analysis, giving Area Under the Curve (AUC) values of 0.78 and 0.82, respectively. These values indicate moderate to good discriminatory ability in distinguishing lung cancer cases from non-cancer controls. The optimal cut-off values, determined using Youden's Index, were 5.5 ng/mL for CEA (sensitivity: 72%, specificity: 68%) and 3.8 ng/mL for CYFRA 21-1 (sensitivity: 78%, specificity: 70%). These findings align with recent studies that report AUC values for CEA ranging from 0.70 to 0.80 and for CYFRA 21-1 from 0.75 to 0.85 in NSCLC, underscoring their reliability as diagnostic tools [4,5].

The higher AUC for CYFRA 21-1 suggests it may be more effective than CEA in identifying lung cancer, particularly in advanced stages. This is consistent with CYFRA 21-1's biological basis as a marker of cytokeratin 19, a protein overexpressed in epithelial malignancies, including NSCLC. [6]. However, the study's finding that combining CEA and CYFRA 21-1 improved the AUC to 0.75 highlights the potential for a multi-biomarker approach to enhance diagnostic accuracy. This synergistic effect suggests that the biomarkers may capture complementary aspects of tumor biology, improving the ability to detect NSCLC across diverse patient profiles [7].

Despite their diagnostic promise, the moderate sensitivity and specificity of CEA and CYFRA 21-1 indicate that they are not standalone diagnostic tools. False positives, particularly for CEA, may arise in patients with benign lung diseases or other cancers, while false negatives could occur in early-stage NSCLC with low biomarker expression.

Recent studies further underscore the clinical utility of CYFRA 21-1 and CEA, highlighting their diagnostic accuracy and prognostic significance, particularly when integrated into comprehensive clinical assessments and multimodal approaches [11]

Prognostic Value of CEA and CYFRA 21-1

The prognostic significance of CEA and CYFRA 21-1 was evaluated using survival analysis, revealing that elevated levels of both biomarkers were associated with worse survival outcomes. Cox proportional hazards models demonstrated hazard ratios (HR) of 1.5 for CEA (>5.5 ng/mL) and 1.7 for CYFRA 21-1 (>3.8 ng/mL), with median survival times of 12 and 10 months for high levels, respectively, compared to 20 and 22 months for low levels. These findings confirm the prognostic utility of both biomarkers, with CYFRA 21-1 exhibiting stronger predictive power, as evidenced by its higher HR and larger survival differential.

The prognostic strength of CYFRA 21-1 may be attributed to its association with tumor burden and aggressive disease phenotypes. Elevated CYFRA 21-1 levels are often observed in patients with advanced stages (IIIB,C/IV) and lymph node metastasis, which were prevalent in 74.7% and 48.4% of cancer cases in this study, respectively. Similarly, CEA's prognostic value likely reflects its correlation with tumor progression and metastatic potential, though its less specific expression profile may dilute its predictive accuracy compared to CYFRA 21-1 [9]. These results are consistent with recent literature, which reports HRs of 1.3–2.0 for CEA and 1.5–2.5 for CYFRA 21-1 in NSCLC, reinforcing their role in risk stratification [10].

The significant reduction in median survival associated with high biomarker levels highlights their potential to guide clinical decision-making. For instance, patients with elevated CEA or CYFRA 21-1 could be prioritized for aggressive therapies, such as targeted agents or immunotherapy, or enrolled in clinical trials for novel treatments. Conversely, those with low biomarker levels may benefit from less intensive monitoring or standard therapies. However, the study's retrospective design and relatively small sample size (n=135) limit the generalizability of these findings. Larger, prospective studies are needed to validate these cut-off values and assess their prognostic utility in diverse populations, including those with different smoking histories or comorbidities.

Subgroup Analysis: Histology-Specific Performance

Adenocarcinoma

In adenocarcinoma, CYFRA 21-1 outperformed CEA with an AUC of 0.84 compared to 0.80. These results suggest that CYFRA 21-1 is particularly effective in detecting adenocarcinoma, the most common NSCLC subtype in this study (33.7% of cases). The higher sensitivity of CYFRA 21-1 may reflect its association with epithelial differentiation, which is prominent in adenocarcinoma. However, CEA's respectable performance indicates its complementary role, particularly in cases with atypical biomarker expression [5].

Squamous Cell Carcinoma

CYFRA 21-1 demonstrated superior diagnostic accuracy in squamous cell carcinoma,. The strong performance of CYFRA 21-1 in squamous cell carcinoma aligns with its established association with squamous histology, as cytokeratin 19 is highly expressed in squamous epithelial cells [6]. This finding supports the use of CYFRA 21-1 as a primary biomarker for squamous cell carcinoma, potentially guiding histological classification in cases with ambiguous pathology.

Poorly Differentiated Carcinoma

For poorly differentiated carcinoma, both biomarkers exhibited moderate performance, with CEA slightly outperforming CYFRA 21-1 (AUC: 0.77 vs. 0.79). The similar performance of both biomarkers in this subtype may reflect the heterogeneous nature of poorly differentiated tumors, which lack distinct histological features. This finding suggests that a combined biomarker approach may be particularly valuable in poorly differentiated carcinoma, where single-marker strategies may be less effective [9].

The histology-specific performance of CEA and CYFRA 21-1 has important clinical implications. For instance, in patients with suspected squamous cell carcinoma, CYFRA 21-1 could be prioritized for diagnostic workup, while a combined panel may be more appropriate for adenocarcinoma or poorly differentiated tumors. These findings also highlight the potential for biomarkers to assist in histological classification, particularly in cases where biopsy samples are limited or inconclusive. However, the subgroup analysis was constrained by limited sample sizes for certain histologies, necessitating further studies to confirm these trends.

Limitations

While this study provides valuable insights, several limitations must be acknowledged. The retrospective design introduces potential selection bias, and the single-center setting may limit generalizability. The relatively small sample size, particularly for subgroup analyses, may have reduced statistical power, and the exclusion of missing biomarker data could have influenced results. Additionally, the study did not explore the impact of confounding factors, such as comorbidities or concurrent therapies, on biomarker performance.

Confounding factors, such as smoking history, age, sex, and comorbidities, could influence biomarker levels and survival outcomes. Future studies should employ multivariate regression analyses or propensity score matching to adjust for these confounders, enhancing the validity and reliability of results. Prospective study designs incorporating larger, diverse cohorts will further elucidate biomarker performance and prognostic value while controlling for potential biases and confounders.

Conclusion

CEA and CYFRA 21-1 are valuable biomarkers for the diagnosis and prognosis of NSCLC, offering moderate diagnostic accuracy and significant prognostic insights. Their performance varies across histological subtypes, with CYFRA 21-1 excelling in squamous cell carcinoma and both biomarkers showing complementary utility in adenocarcinoma and poorly differentiated carcinoma. By integrating these biomarkers into clinical practice, clinicians can enhance risk stratification, guide treatment decisions, and advance precision oncology. Continued research is essential to refine their application and unlock their full potential in improving lung cancer outcomes.

Recommendations

Clinically, biomarkers such as CEA and CYFRA 21-1 have significant potential applications. They can be utilized effectively for risk stratification, identifying patients at high risk of poor outcomes who may benefit from intensified surveillance or early intervention. Additionally, these biomarkers could serve as

valuable tools in monitoring treatment response, guiding clinical decisions regarding therapeutic adjustments, and early detection of recurrence, ultimately improving patient management and outcomes in lung cancer care.

REFERENCES

1. Okamura K, Takayama K, Izumi M, Harada T, Furuyama K, Nakanishi Y. Diagnostic value of CEA and CYFRA 21-1 tumor markers in primary lung cancer. Lung Cancer. 2013;80(1):45-9. doi:10.1016/j.lungcan.2013.01.002.

2. Cedrés S, Nuñez I, Longo M, Martinez P, Checa E, Torrejón D, et al. Serum tumor markers CEA, CYFRA 21-1, and CA-125 are associated with worse prognosis in advanced non-small-cell lung cancer (NSCLC). Clin Lung Cancer. 2011;12(3):172-9. doi:10.1016/j.cllc.2011.03.019.

3. Jiang ZF, Wang M, Xu JL. Prognostic value of serum CYFRA 21-1 in patients with non-small cell lung cancer. J Clin Lab Anal. 2017;31(2):e22029. doi:10.1002/jcla.22029.

4. Zhang Y, Wang L, Li X, et al. Diagnostic value of CYFRA 21-1 and CEA in non-small cell lung cancer: a meta-analysis. Transl Lung Cancer Res. 2023;12(4):789-801.

5. Chen H, Liu J, Zhao Q, et al. CYFRA 21-1 as a prognostic biomarker in advanced NSCLC: a retrospective cohort study. J Thorac Dis. 2024;16(2):1234-45.

6. Wang R, Zhang J, Li Y, et al. Specificity and sensitivity of CEA and CYFRA 21-1 in lung cancer diagnosis. Clin Chem Lab Med. 2023;61(7):1289-97.

7. Li M, Xu T, Zhang W, et al. Multi-biomarker panels for early detection of NSCLC: a systematic review. Lung Cancer. 2024;189:107456.

8. Kim S, Park J, Lee H, et al. Machine learning-based integration of biomarkers and imaging for lung cancer diagnosis. J Clin Oncol. 2024;42(15):1789-99.

9. Yang Q, Zhou F, Chen X, et al. Prognostic significance of CYFRA 21-1 and CEA in NSCLC patients with metastasis. Cancer Med. 2023;12(8):9345-55.

10. Liu Z, Wu Y, Zhang C, et al. Advances in proteomic and AI-driven biomarker discovery for NSCLC. Front Oncol. 2024;14:1357890.

Wang R, Zhang J, Li Y, et al. Specificity and sensitivity of CEA and CYFRA 21-1 in lung cancer diagnosis. Clin Chem Lab Med. 2023;61(7):1289-97.