

Journal Pre-proof

CLINICAL APPLICATION OF EXPERT SOFTWARE BASED ON SIX TUMOUR BIOMARKERS TO STRATIFY THE RISK OF LUNG CANCER IN A PULMONARY RAPID DIAGNOSIS UNIT

Rafael Molina Jaume Trapé Adolfo Garrido Ernesto Salas Rosa M^a. Domínguez Gutierrez de Ceballos Simón Gundín Antonio León Justel Marta García de Burgos Juana Piedad Rivas M. Ángel Julián Ansón Manuel Matos Garrido Sofía Lorenzo García Ramón Marrades Nekane Múgica Atorrasagasti Estefanía Luque Crespo Cecilia López-Ramírez Fernando Gustavo Gutiérrez Herrero Carolina González-Fernández Victoria Villalta Robles M^a Cruz González Cocaño Mónica Ramos Álvarez José Carvalho Marta Clara Jiménez García Raquel Cabezón Vicente M^a. Pavón Masa Ruth Sáez de la Maleta Úbeda Oscar Bernadich Graciliano Estrada Trigueros Aurora Orejas J. Carlos Santana Astudillo J. Trapé-Úbeda Antonio Barco-Sánchez

PII: S0300-2896(26)00003-7

DOI: <https://doi.org/doi:10.1016/j.arbres.2025.12.007>

Reference: ARBRES 3934

To appear in: *Archivos de Bronconeumología*

Received Date: 19 August 2025

Accepted Date: 22 December 2025

Please cite this article as: Molina R, Trapé J, Garrido A, Salas E, de Ceballos RMDG, Gundín S, Justel AL, de Burgos MG, Rivas JP, Angel Julián Ansón M, Garrido MM, García SL, Marrades R, Atorrasagasti NM, Crespo EL, López-Ramírez C, Herrero FGG, González-Fernández C, Robles VV, Cruz González Cocaño M, Álvarez MR, Carvalho Marta J, García CJ, Vicente RC, Pavón Masa M, de la Maleta Úbeda RS, Bernadich O, Trigueros GE, Orejas A, Carlos Santana Astudillo J, Trapé-Úbeda J, Barco-Sánchez A, CLINICAL APPLICATION OF EXPERT SOFTWARE BASED ON SIX TUMOUR BIOMARKERS TO

STRATIFY THE RISK OF LUNG CANCER IN A PULMONARY RAPID DIAGNOSIS UNIT,
Archivos de Bronconeumología (2026), doi: <https://doi.org/10.1016/j.arbres.2025.12.007>

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier España, S.L.U. on behalf of SEPAR.

AUTHORSHIP FILE

-Type or article: Original Article

CLINICAL APPLICATION OF EXPERT SOFTWARE BASED ON SIX TUMOUR BIOMARKERS TO STRATIFY THE RISK OF LUNG CANCER IN A PULMONARY RAPID DIAGNOSIS UNIT

-List of authors:

Rafael Molina[†]1, Jaume Trapé⁷, Adolfo Garrido², Ernesto Salas⁴, Rosa M^a. Domínguez Gutierrez de Ceballos⁵, Simón Gundín⁶, Antonio León Justel³, Marta García de Burgos⁸, Juana Piedad Rivas⁹, M. Ángel Julián Ansón¹⁰, Manuel Matos Garrido¹¹, Sofía Lorenzo García¹², Ramón Marrades¹³, Nekane Múgica Atorrasagasti¹⁴, Estefanía Luque Crespo¹⁵, Cecilia López-Ramírez¹⁶, Fernando Gustavo Gutiérrez Herrero¹⁷, Carolina González-Fernández⁷, Victoria Villalta Robles⁸, M^a Cruz González Cocaño¹⁹, Mónica Ramos Álvarez¹⁰, José Carvalho Marta¹¹, Clara Jiménez García¹², Raquel Cabezón Vicente², M^a. Pavón Masa¹⁵, Ruth Sáez de la Maleta Úbeda⁶, Oscar Bernadich¹⁸, Graciliano Estrada Trigueros²⁰, Aurora Orejas¹⁹, J. Carlos Santana Astudillo¹⁴, J. Trapé-Úbeda²¹, Antonio Barco-Sánchez³

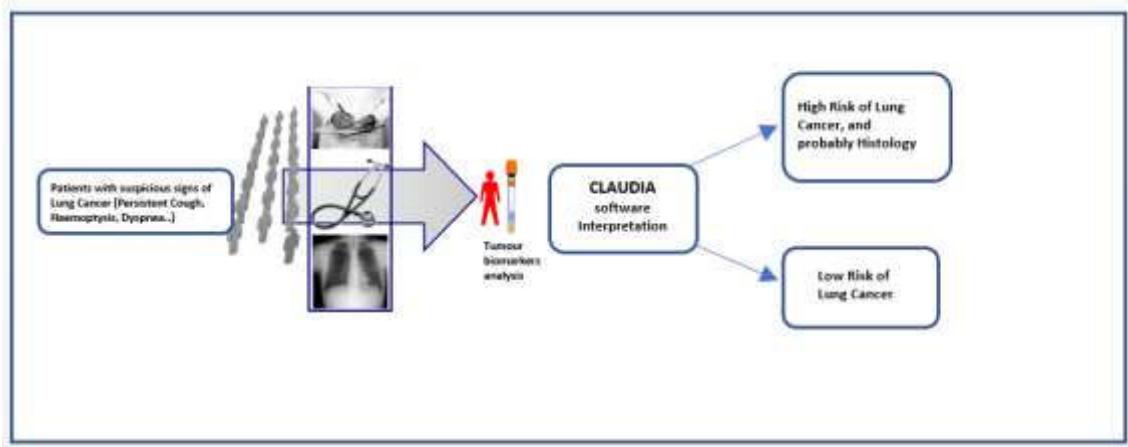
1. Clinical Biochemistry Department (Oncobiology Unit), Centro de Diagnóstico Biomédico (CDB), Hospital Clinic, IDIBAPS, University of Barcelona, Spain
2. Clinical Biochemistry Department, Hospital Donostia, San Sebastian. Spain
3. Clinical Biochemistry Department, University Hospital Virgen Macarena, Seville, Spain
4. Clinical Analysis Department, Hospital San Juan de Dios, Seville, Spain
5. Clinical Analysis Department, University Hospital Virgen del Rocío, Seville, Spain

6. Clinical Analysis Department, Burgos University Hospital, Burgos, Spain
7. Laboratory Medicine, Althaia Xarxa Assistencial Universitària de Manresa, Spain
8. Clinical Analysis Department, General Hospital of Segovia, Segovia, Spain
9. Pneumology Department, Leon University Healthcare Complex, Leon, Spain
10. Clinical Biochemistry Department, Clinical University Hospital Lozano Blesa, Zaragoza, Spain
11. Clinical Pathology Department, Centro Hospitalar Universitario Lisboa Central, Lisbon, Portugal
12. Clinical Analysis Department, University General Hospital of Alicante, Alicante, Spain
13. Respiratory Institute, Hospital Clinic, IDIBAPS, University of Barcelona, Spain
14. Pneumology Department, Hospital Donostia, San Sebastian. Spain
15. Pneumology Department, University Hospital Virgen Macarena, Seville, Spain
16. Pneumology Department, University Hospital Virgen del Rocío, Seville, Spain
17. Pneumology Department, Burgos University Hospital, Burgos, Spain
18. Pneumology Department, Althaia Xarxa Assistencial Universitària de Manresa, Spain
19. Clinical Analysis Department, Leon University Healthcare Complex, Leon, Spain
20. Pneumology Department, General Hospital of Segovia, Segovia, Spain
- 21 General Directorate of Health Planning and Research, Ministry of Health, Generalitat de Catalunya, Spain

-Corresponding author:

Antonio Barco-Sánchez: Clinical Biochemistry Department, University Hospital Virgen Macarena, Seville, Spain.

Graphical abstract



Highlights

- Software improves the effectiveness of biomarkers panel in lung cancer diagnosis.
- CLAUDIA expert software can complement LDCT scans for early detection of Lung Cancer.
- CLAUDIA Expert Software Rapidly Stratifies lung cancer risk in symptomatic Patients.

ABSTRACT

Objectives

Diagnostic tools that stratify lung cancer (LC) risk can help prioritize care for patients at the highest risk and optimize time and procedures to achieve the final diagnosis. We have previously demonstrated that six tumour biomarkers (TBs) - CEA, CYFRA 21.1, CA 15-3, SCC Ag, ProGRP, and NSE - can help assess LC risk. We developed expert software that combines these TBs with clinical and imaging data to estimate LC risk.

Methods.

The diagnostic accuracy of this expert software was evaluated in a multicentre study. We prospectively recruited 2,005 individuals referred to 12 reference hospitals in Spain and Portugal for suspicion of LC. The six TBs were determined and the expert software

was applied to all patients and correlated with the final diagnosis.

Results:

A final diagnosis of LC was made in 1,392 patients. The expert software yielded 87.7% sensitivity, 75.5% specificity, 89.0% positive predictive value and 73.0% negative predictive value. Sensitivity increased with tumour size and extension. The software also provides histological information, correctly predicting cancer in 98.4% of small-cell LC and 93.2% of non-small-cell LC, which correlates with the histological diagnosis of 90% and 91.2%, respectively.

Conclusions:

The expert software developed provides excellent diagnostic accuracy for diagnosing LC. Accordingly, this software can help stratify the risk of LC and prioritize the evaluation of patients at higher risk, optimizing procedures based on risk and knowledge of the most likely histological type, and providing a valuable tool for risk stratification and clinical decision support, particularly in Rapid Diagnostic Units.

Keywords: Lung cancer; Diagnosis; Tumour biomarkers; Expert software; Risk stratification.

INTRODUCTION

Lung cancer (LC) is the most prevalent and deadliest form of human cancer worldwide, accounting for 1.8 million deaths annually and 17.8% of all cancer-related fatalities [1]. While in some cases the diagnosis LC is straightforward, it remains challenging in others, particularly when imaging studies reveal indeterminate nodules [2,3]. Several studies have suggested that low-dose computed tomography (LDCT) screening can reduce mortality by LC; however, the results remain controversial [3–8]. Additionally, LDCT is a high-demand resource, leading to delays in early access, and the associated costs, along with the high incidence of indeterminate nodules, make it crucial to prioritize

the use of this study in high-risk patients [4,5,9]. Furthermore, while advanced imaging such as PET-CT offers high diagnostic value, it is an expensive and often less accessible technique in many hospitals, limiting its use as a first-line triage tool.

Circulating tumour biomarkers (TB) are valuable diagnostic tools, particularly when used in combination to enhance sensitivity and specificity [10-13]. However, the optimal TB combination for maximizing accurate diagnosis of LC remains uncertain [10,12-19]. Previous studies by our group have demonstrated that a combination of six TB [carcinoembryonic antigen (CEA), cytokeratin fragment 21-1 (CYFRA 21-1), cancer antigen 15-3 (CA15- 3), squamous cell carcinoma antigen (SCC-Ag), progastrin-releasing peptide (ProGRP), and neuron-specific enolase (NSE)] correlates with the presence of LC and its major histological subtypes, i.e., non-small-cell LC (NSCLC) - adenocarcinoma (ADC) and squamous cell carcinoma (SCC) - and small-cell LC (SCLC)[11,20]. This combined TB model showed significantly greater diagnostic accuracy than a clinical model based solely on tumour size, age, and smoking status [20]. Although some studies indicate the potential benefit of integrating TB serum concentrations with LDCT for optimizing the diagnosis of LC [21], this approach has yet to be fully established. LDCT is a high-demand resource, leading to delays in early access, and the associated costs, along with the high incidence of indeterminate nodules, make it crucial to prioritize the use of this study in high-risk patients [4,5,9].

It is important to acknowledge that TB markers can yield false positives in certain pathophysiological conditions that require differentiation from LC. Incorporating clinical and laboratory variables that identify these conditions can enhance diagnostic accuracy.

To address this challenge, we developed expert software named CLAUDIA (Cancerous Lung Algorithm Useful for DIAgnosis). Using clinical variables, computed tomography (CT) data, and TB concentrations from 5,000 patients, CLAUDIA calculates the risk of lung cancer (LC) and suggests a histological classification.

The study aimed to assess and validate the clinical utility of a TB-based software tool for decision-making in rapid-diagnosis pulmonary units in 12 hospitals in Spain and Portugal.

MATERIAL AND METHODS

Study Design and Participants

This prospective, consecutive study included 2,101 adults presenting signs of LC across 12 hospitals in Spain and Portugal. Patients with prior LC treatment, active malignancies, or renal failure were excluded, while those with non-cancerous conditions were included. The final study population comprised 2,005 individuals. All participants provided informed consent, and data were anonymized (Figure 1).

This study represents the first large-scale, prospective, multicenter external validation of the CLAUDIA algorithm, developed by our group.

The study was approved by the corresponding Ethics Committees (HCB/2017/1060). The study was not registered, so when it was designed, it was not required to be carried out. LC diagnosis followed international guidelines [2] and was confirmed using CT or positron emission tomography scans, and tissue analysis obtained via bronchoscopy, fine-needle aspiration, endobronchial ultrasound, oesophageal ultrasound, or surgical resection. Histology typing was conducted in all patients.

Histological typing and staging of LC

LC subtypes were classified according to the 2015 World Health Organization recommendations [22,23]. Differentiation between SCLC and NSCLC was based on morphological criteria and immunohistochemical markers such as CD56 and synaptophysin [24]. Staging followed international Tumour-Node-Metastasis (TNM)

guidelines [25].

TB measurements

Peripheral blood samples were collected without anticoagulants, centrifuged, and stored at 3–5 °C until analysis. Serum TB concentrations were measured in each laboratory using electrochemiluminescent assays (Elecsys, ROCHE Diagnostics Switzerland). The previously validated upper reference limits (URLs) were: CEA, 5 ng/ml; CYFRA 21-1, 3.3 ng/ml; SCC-Ag, 2 ng/ml; CA 15-3, 35 U/ml; NSE, 25 ng/ml; ProGRP, 65 pg/ml. TB values exceeding these thresholds were classified as “abnormal.”

Expert software

The software is built on algorithms that analyse a comprehensive database of over 5,000 patients from previous studies [11,13,15,20]. It evaluates serum TB concentrations while accounting for biological variability and clinical factors such as pleural effusion, renal insufficiency, smoking, cholestasis, and dermatologic conditions. Different TB cut-off values are applied based on clinical conditions; for instance, in smokers, the CEA threshold is adjusted to 10 ng/mL instead of 5 ng/mL. In patients with renal failure, SCC-Ag is excluded, while in those with hepatopathy, certain TB cut-offs increase by up to 50%.

The software also integrates imaging data, such as nodule size and other characteristics, to enhance diagnostic accuracy. This multi-variable approach improves sensitivity and specificity, significantly reducing false-positive rates.

Based on the input data, the software stratifies patients into risk groups of presenting LC. The categories are as follows: Very high risk (with a probability of more than 95%), high risk (with a probability of LC between 75% and 95%), moderate risk (between 65% and 75%) and low risk (less than 65% probability). For patients classified as moderate risk, the software recommends repeating the TB tests after a period of three to four weeks to refine diagnostic accuracy. This dynamic reassessment method has been shown to improve diagnostic specificity

[10,20,26].

The current study was conducted as a prospective, multicentre external validation of the CLAUDIA software based on the analysis of data from patients who underwent a tumour marker profile for suspected lung cancer. The cohort included a total of 2,005 patients from 12 participating centres (tertiary and regional hospitals) in Spain and Portugal (Supplementary Table S1).

The algorithm is predicated on a rule-based decision model that integrates molecular (biomarker), clinical, and radiological data in order to estimate risk. The methodology described herein facilitates the implementation of the dynamic cut-off thresholds previously delineated for the purpose of personalised risk stratification.

Statistical analysis

Results were expressed as case counts, proportions, medians, and interquartile ranges. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. TB concentrations were compared using parametric (Student's t-test) or non-parametric tests (Wilcoxon, Mann-Whitney, Kruskal-Wallis).

The Net Reclassification Improvement Index (NRI) was used to assess the ability of the software to reclassify LC diagnoses compared to TB analysis alone [27]. NRI quantifies improvements in classification by accounting for true and false positives and negatives.

Net Benefit (NB) analysis [28] was used to compare the diagnostic performance of the combined TB panel against the software. A p-value <0.05 was considered statistically significant. All analyses were performed using SPSS v.25 (IBM Corp.).

To assess robustness and generalizability across the centres (as detailed in Supplementary Table S2), a formal statistical test for heterogeneity in the Sensitivity and Specificity estimates was performed using the Cochran's Q test and the I^2 statistic.

RESULTS

The diagnosis of LC was confirmed in 1,392 (69.4%) patients, while 613 (30.6%) were found to have benign disease. Among the LC cases, 266 were SCLC (19.1%) and 1,126 NSCLC (80.9%); adenocarcinomas were the most frequent subtype (n=635; 56.4%), followed by SCC (n=319; 28.3%) and cancers of indeterminate lineage (n=172; 15.3%) (Figure 1).

In the suspected LC cases, the main finding leading to the diagnostic workup was the presence of radiographic nodules in 49.8% of cases, which were slightly more frequent in non-cancer than in LC patients (56.9% vs. 46.7%). Dyspnoea, haemoptysis, thoracic pain, and persistent cough were present in around 10%; thoracic pain and persistent cough were more frequent in LC, while haemoptysis was more common in non-cancer cases. Constitutional syndrome components occurred in 5% of LC cases versus 1.5% in non-cancer patients (Table 1).

Table 2 compares the clinical and imaging findings and TB concentrations between LC and non-LC patients. Significant differences were observed in gender, smoking habits, cigarette consumption per year, lung nodules (especially those >3 cm), and TB concentrations, all of which were more prevalent in LC patients ($p<0.01$ to 0.001). Smaller nodules were more commonly found in non-LC patients. Within the LC subgroup analysis, NSCLC patients were generally older and had different TB profiles: CEA, CYFRA 21-1, SCC-Ag, and CA 15-3 levels were higher in NSCLC, whereas NSE and ProGRP were elevated in SCLC patients ($p=0.001$). All TB values were higher in NSCLC than in non-LC patients, except for SCC-Ag, which had similar values in SCLC and non-LC patients.

Table 3 summarizes the diagnostic sensitivity, specificity, NPV, and PPV of TB for predicting the risk of LC, both individually and in combination. The individual diagnostic sensitivity ranged from 19.4% (NSE) to 59.8% (CYFRA 21-1), while specificity was notably higher, varying from 89.6% (CYFRA 21-1) to 99.3% (NSE). This variation is

observed because LC is not a single disease but consists of multiple histological subtypes with distinct behaviours, expression patterns and treatment responses. For example, NSE is predominantly elevated in SCLC, which accounts for approximately 20% of LC cases. Consequently, the sensitivity of NSE is low across all LC subtypes but remains highly specific for SCLC when elevated. By incorporating a comprehensive tumour marker panel, the software accounts for these variations, enabling accurate classification of LC subtypes based on expression patterns. Combined TB assessments, whether defined as a ≥ 1 abnormal TB marker or through software-based analysis, substantially improved diagnostic sensitivity (89.4% and 87.7%, respectively). The software demonstrated a higher specificity (75.5%) compared to ≥ 1 abnormal TB (63.9%) and exhibited the best NPV (73.03%) and PPV (89.06%).

The NRI analysis of the software led to an overall reclassification rate of 9.8% compared to the elevated ≥ 1 TB rule. Among 113 reclassified cases, 104 were changed from positive to negative (7.09%), while 9 shifted from negative to positive. Consequently, the software improved NRI by approximately 10% in distinguishing LC from non-LC cases.

The NB difference between the elevated ≥ 1 TB approach and the software was 0.07%, meaning that for every 14 cases analysed, one additional true cancer case was identified without increasing the false positives.

The prevalence and size of lung nodules were significantly greater in LC patients (Table 4). Across all nodule size categories (<1 cm, 1–3 cm, and >3 cm), TB levels were significantly higher in LC patients (Figure 3B). Figure 2A illustrates the relationship between TB levels and the histological subtypes of LC. NSE and ProGRP showed greater sensitivity in detecting SCLC, whereas CYFRA 21-1 and CEA demonstrated higher diagnostic utility in NSCLC. Of note, NSE and ProGRP concentrations were comparatively lower in NSCLC. Furthermore, CYFRA 21-1 concentrations were markedly elevated in SCC, while CEA levels were significantly increased in ADC.

Figure 2B reveals that, within NSCLC, the sensitivity of TB tends to rise proportionally with increasing tumour burden or more extensive dissemination.

Figure 3A presents concordant findings for SCLC, further supporting the diagnostic relevance of NSE and ProGRP in this histological subtype.

Figure 3B depicts the estimated probability of malignancy stratified according to nodule size and TB positivity, demonstrating that both increased nodule dimensions and TB positivity are associated with a significantly higher likelihood of cancer.

Table 5 summarizes the concordance between algorithm-predicted classifications and the definitive histological diagnoses. The algorithm achieved high concordance for SCLC (90.0%), NSCLC (91.2%), and its major subtypes, including ADC (79.4%) and SCC (63.4%). However, the performance was less robust in cases categorized as unspecified NSCLC (uNSCLC), as well as in the moderate-risk groups, which yielded concordance rates ranging from 8.7% to 38.0%. These findings demonstrate that the algorithm accurately distinguishes among the most common tumour types, while less well-defined or indeterminate categories remain a challenge. The heterogeneity analysis of the diagnostic accuracy across the 12 participating centres revealed mixed results. We found homogeneity in Specificity between centers ($p=0.081$; $I^2=39.0\%$), indicating the algorithm's stable capacity to correctly identify non-malignancy. Conversely, significant statistical heterogeneity was detected for Sensitivity ($p=0.005$; $I^2=59.6\%$). We tested the hypothesis that this variability was driven by differences in case-mix. When centers were classified into two clinical subgroups based on the percentage of early-stage patients (those with $>25\%$ and those with $<20\%$) see table S3, homogeneity was successfully demonstrated in both Sensitivity and Specificity within each subgroup, confirming that the observed heterogeneity was explained by differences in patient populations (table S2).

DISCUSSION

Recent bibliometric analyses provide insights into the growing interest in TBs for the diagnosis of LC. A review of 990 publications from 2000 to 2022 highlights the sustained research focus on TBs and underscores the unmet need for expert software to analyse these biomarkers [29]. Our study addresses this gap by validating a six-TB panel (CEA, CA15-3, SCC, CYFRA 21-1, NSE, and ProGRP) in conjunction with expert software. Previous research has demonstrated that this panel delivers high sensitivity (88.5%) and specificity (82%) in single determinations, with further improvements through serial testing [20]. In our multicentre study, the panel exhibited a similar sensitivity (87.7%), confirming its robustness and applicability in larger, diverse patient populations. Notably, this present study represents the first large-scale multicentre evaluation of these TBs with the software, as most previous investigations have been limited to smaller, single centre cohorts.

Our findings demonstrate that the CLAUDIA software achieves strong diagnostic performance (Sensitivity 89%, Specificity 73%), which is highly competitive with established methods described in the diagnostic literature, such as LDCT [4,5,7,30,31] and nucleic acid-based liquid biopsies [9,32]. Since our cohort has a high disease prevalence due to its focus on patients already undergoing diagnostic workup for highly suspicious clinical and radiological findings, the algorithm's performance requires further validation in the context of a low-prevalence screening setting before its utility can be fully established as a supplement to current screening procedures.

LC comprises various histological subtypes associated with distinct TB expression patterns. For instance, SCLC predominantly expresses NSE and ProGRP, while NSCLC subtypes exhibit higher levels of other TBs (Figures 2A, 2B, 3A). This diversity underscores the need for a comprehensive TB panel capable of detecting all LC subtypes. As shown in figures 2B, 3A and table 3, there is an association between the TB expression

pattern and the histological subtype, which enables the software to suggest a particular histology based on the TB concentrations measured. Furthermore, tumour stage plays a crucial role in diagnosis as early-stage LC- characterized by smaller nodules- poses a greater diagnostic challenge than advanced disease. Previous studies, including our own, have demonstrated that abnormal TB levels significantly increase cancer risk across all nodule sizes. Specifically, patients with nodules smaller than one centimetre exhibit a fivefold increased risk if TB levels are abnormal, while those with nodules larger than three centimetres and abnormal TB levels have a >95% risk of LC [11,13,15,20,33]. Figure 3B illustrates the strong association between nodule size and TB positivity in the probability of presenting cancer.

While essential, static imaging techniques such as CT scans have limitations in assessing the dynamic nature of tumour growth and behaviour. This challenge is particularly evident in aggressive tumours, which can progress rapidly over short timeframes. The software addresses this issue by integrating static imaging data (e.g., nodule size, shape, and presence) with dynamic biomarker data from TB concentrations. This combination of the two types of data significantly enhances diagnostic accuracy and facilitates earlier detection of LC.

Unlike artificial intelligence models [34], our approach utilises predefined cut-off values for different pathologies based on empirical data. It applies expert-driven rules to personalize these cut-offs, leveraging a robust, validated database of over 5,000 patients, including healthy individuals, those with benign pathologies, and LC patients [11,13,20]. By proactively addressing potential sources of error, the software ensures a high level of reliability. It minimizes false positives, even in complex cases involving confounding factors such as pleural effusion or other non-cancerous conditions. In cases of diagnostic uncertainty, repeating measurements after 3-4 weeks significantly improves diagnostic

accuracy, reaching a high specificity as already shown in previous studies by our group [20].

In our study, the software demonstrated notable improvements in diagnostic performance compared to TBs only. Specificity increased from 63.9% to 75.5%, PPV from 84.9% to 89.1%, and sensitivity was maintained at 87.7%. The NRI analysis revealed a 10% improvement in the accuracy of classification compared to using TB alone, which, according to some authors, could have a greater clinical impact than a 10% increase in the area under the curve in some analyses such as the risk-prediction [35], highlighting the clinical relevance of incorporating the expert software into diagnostic workflows. Our results demonstrate that the algorithm performs reliably for the primary histologic subtypes of LC, particularly SCLC and NSCLC, reaching concordance rates of 90.0% and 91.2%, respectively. Subtype-level discrimination was also strong for ADC and SCC, supporting the potential utility of the algorithm in routine diagnostic workflows. Nevertheless, accuracy substantially decreased in uNSCLC and in specimens classified as moderate risk, indicating that indeterminate or overlapping features may reduce algorithm precision in these categories. This limitation underscores the need for further refinement of the algorithm, potentially through training with a larger and more diverse dataset or by integrating additional diagnostic modalities. Ultimately, improving algorithmic performance in these challenging cases is critical for its broader implementation in clinical practice. These analytical features render this working model a valuable asset for Rapid Diagnosis Units. It facilitates swift stratification of patients based on their risk of developing LC, enabling the expeditious identification of patients who should be given high priority for treatment. Moreover, the ability to suggest the histological subtype of a tumour may provide critical guidance in determining the suitability of a patient for surgical intervention. In cases in which SCLC is indicated,

concordance with the definitive histopathological diagnosis exceeds 90%. Early identification of the potential histological subtype also facilitates more targeted diagnostic strategies, such as obtaining sufficient tissue for next-generation sequencing, particularly in patients with tumours located in anatomically challenging regions.

The specificity achieved by the expert software in this study (75.5%) was slightly lower than the 82% reported in previous studies [20]. This discrepancy may be attributed to the multicentre nature of the trial, which involved routine clinical conditions across 12 hospitals and laboratories. Nevertheless, our findings in previous studies reinforce the potential of serial TB determinations in reducing false positives, emphasizing the importance of dynamic TB assessments in real-world clinical settings [20,36,37].

Overall, our findings highlight the potential of the six-TB panel and expert software as valuable complements to LDCT in early LC detection. The strong correlation between TB results and nodule size suggests that patients with small nodules and abnormal TB levels should be prioritized as high-risk groups. Compared to established screening methods for other cancers, such as faecal occult blood testing for colorectal cancer or prostate-specific antigen testing for prostate cancer, our TB panel and software demonstrate superior diagnostic performance, with advantages that include being non-invasive, cost-effective, widely available, easy to repeat in a short time and capable of achieving high specificity when combined with serial testing.

We conclude that this study confirms the effectiveness of a six-serum TB panel (CEA, CA 15-3, SCC- Ag, CYFRA 21-1, NSE, and ProGRP) combined with the CLAUDIA expert software for diagnosing LC. The CLAUDIA expert software further enhances diagnostic accuracy, achieving the highest sensitivity-to-specificity ratio (87.7% and 75.5%), as well as the best PPV (89%), and NPV (73.03%) among currently available

diagnostic tests for LC. Moreover, the correlation between TB results and nodule size, along with a sensitivity of 70% in early stages, suggests that the TB lung panel and CLAUDIA expert software could serve as valuable complementary tools to LDCT scans for early detection of LC. The aim of the CLAUDIA algorithm was not to replace established imaging modalities but to serve as a simple, fast, (providing results in four hours) standardized, and affordable tool that supports rapid clinical decision-making. The algorithm provides results within four hours and is designed to be applicable in any healthcare setting, regardless of local imaging resources.

We acknowledge that the observed prevalence of lung cancer in our study cohort (69.4%) is high. This was expected, as the population comprised patients referred to Rapid Diagnostic Units due to an existing clinical or radiological suspicion of malignancy, resulting in a naturally high pre-test probability. This high-risk context is precisely where clinicians need the most support in stratifying indeterminate cases and act promptly.

Importantly, CLAUDIA goes beyond simple binary risk stratification by predicting the most probable histological subtype in high-risk cases (Table 5). This crucial additional information allows clinicians to anticipate subsequent diagnostic steps or complementary molecular testing, significantly improving patient management.

The most relevant finding in our meta-analysis is the significant heterogeneity observed in Sensitivity, which contrasts with the stable performance observed in Specificity. Our subsequent subgroup analysis, classifying centres by the proportion of early-stage patients, effectively resolved this heterogeneity. This confirms that the variation in Sensitivity is predominantly a function of the pre-test probability (i.e., the disease stage distribution) rather than fundamental differences in the core performance of the CLAUDIA algorithm or variations in laboratory methodology. This dependency on case-

mix is a known phenomenon in diagnostic tests, and the restoration of homogeneity within defined subgroups strengthens the argument for the test's consistency when applied to similar patient populations.

Further research is needed to validate these findings and assess their clinical significance.

Strengths and limitations

We carried out the validation study in 12 different centres with minimal prior training with the software. Despite the differences among the centres, the results have been very satisfactory. We believe that with more extensive prior training, the results could have significantly improved. However, this also highlights the intuitive and user-friendly strengths of this software.

The primary strength of this work is its prospective, multicentre external validation design. The evaluation across 12 hospitals with varying complexity and protocols assures the reproducibility and robustness of the CLAUDIA algorithm, a fact formally confirmed by the statistical homogeneity of Specificity across all centres. This provides strong evidence that the tool is reliable regardless of the specific healthcare setting. The study will allow us to evaluate the capacity of tumour markers in combination with imaging methods for the diagnosis of patients arriving at the rapid diagnosis unit under unified criteria. A key strength of this study is the successful explanation of initial statistical heterogeneity through a robust subgroup analysis based on clinical criteria (proportion of early-stage patients). The demonstrated homogeneity of both Sensitivity and Specificity within the defined clinical subgroups confirms the algorithm's reliable and generalizable performance when applied to patient cohorts with similar disease stage distributions. Furthermore, the overall homogeneity of Specificity across all 12 centers strengthens confidence in the test's ability to minimize false positives, a critical factor for screening

applications.

The study's limitations include the lack of registration in a public database like ClinicalTrials.gov, a practice we acknowledge as advisable for the prospective component of the study. The primary weakness lies in the initial finding of significant heterogeneity in the overall Sensitivity when all 12 centers are analyzed together. Although this heterogeneity was statistically explained by the case-mix differences, it underscores the need for standardized patient selection criteria in future studies to minimize variability. Another limitation is the dependence on aggregated data (TP/FP/TN/FN), which prevents us from performing more detailed individual patient data (IPD) meta-regression to identify other potential predictors of performance variation beyond the disease stage. We also acknowledge that our study did not include a direct comparison against other established risk prediction models, such as the Mayo Clinic model. Future validation studies should aim to benchmark our model directly against these tools to better ascertain its comparative effectiveness.

FINAL DECLARATION SECTION

-Funding of the research:

“This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors”.

Contribution of authors:

R. Molina, A. Barco and J Trapé: study concept and design, statistical analysis, and manuscript drafting. They had full data and took responsibility for the integrity of the data analysis which they performed.

All authors collaborated in the recruitment of patients, data curation and submission of results for analysis.

-Conflict of interest of every author:

“The authors declare not to have any conflicts of interest that may be considered to influence directly or indirectly the content of the manuscript”.

-Artificial intelligence involvement:

“The authors declare that the material has not been partial or totally produced with the help of any artificial intelligence software or tool”.

Acknowledgements:

The authors are especially grateful and want to pay tribute to Dr. Rafael Molina, who designed and led the study and sadly passed before its end. We have all learned from his enthusiasm, drive, and personal values. He has been a leader in handling TB and was our teacher and friend. This study was his idea and would not have been possible without his encouragement. We will continue his teachings, improving and developing greater TB management algorithms to improve the diagnosis of cancer. We hope that this work serves as a tribute to him.

We would like to express special gratitude to Luis de Cabo for his selfless help in revising

the manuscript and the enthusiasm he conveyed to us.

We are indebted to the study participants. All the professionals from the different hospitals collaborated with professionalism and enthusiasm. This study has allowed collaboration between the Clinical Laboratory and the Pneumology Departments of the different hospitals, improving diagnosis and patient care.

Ethics in publishing

1. Does your research involve experimentation on animals?:

No

2. Does your study include human subjects?:

Yes

If yes; please provide name of the ethical committee approving these experiments and the registration number. :

Etics Comitee of Hospital Clinic of Barcelona register number HCB/2017/1060.

If yes; please confirm authors compliance with all relevant ethical regulations. :

Yes

If yes; please confirm that written consent has been obtained from all patients. :

Yes

3. Does your study include a clinical trial?:

No

4. Are all data shown in the figures and tables also shown in the text of the Results section and discussed in the Conclusions?:

Yes

REFERENCES

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49. <https://doi.org/10.3322/CAAC.21660>
- [2] Gagl M, Powell CA, Schraufnagel DE, Schonfeld N, Rabe K, Hill NS, et al. An official American Thoracic Society/European respiratory society statement: The role of the pulmonologist in the diagnosis and management of lung cancer. *Am J*

Respir Crit Care Med 2013;188:503-7. <https://doi.org/10.1164/rccm.201307-1269ST>

[3] McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369:910-9. <https://doi.org/10.1056/NEJMoa1214726>

[4] Crosbie PA, Balata H, Evison M, Atack M, Bayliss-Brideaux V, Colligan D, et al. Second round results from the Manchester a “Lung Health Check” community-based targeted lung cancer screening pilot. Thorax 2019;74:700-4. <https://doi.org/10.1136/thoraxjnl-2018-212547>

[5] Crosbie PA, Balata H, Evison M, Atack M, Bayliss-Brideaux V, Colligan D, et al. Implementing lung cancer screening: Baseline results from a community-based “Lung Health Check” pilot in deprived areas of Manchester. Thorax 2019;74:405-9. <https://doi.org/10.1136/thoraxjnl-2017-211377>

[6] Wilson DO, Weissfeld JL, Fuhrman CR, Fisher SN, Balogh P, Landreneau RJ, et al. The Pittsburgh lung screening study (PLuSS): Outcomes within 3 years of a first computed tomography scan. Am J Respir Crit Care Med 2008;178:956-61. <https://doi.org/10.1164/rccm.200802-336OC>

[7] Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: A systematic review to update the U.S. preventive services task force recommendation. Ann Intern Med 2013;159:411-20. <https://doi.org/10.7326/0003-4819-159-6-201309170-00690>

[8] The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395-409. <https://doi.org/10.1056/NEJMoa1102873>

[9] Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model

incorporating lung function: Development and validation in the UK biobank Prospective cohort study. *J Clin Oncol* 2017;35:861-9.
<https://doi.org/10.1200/JCO.2016.69.2467>

[10] Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brünner N, Chan DW, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. *Clin Chem* 2008;54:e1-29.
<https://doi.org/10.1373/clinchem.2008.105601>

[11] Molina R, Auge JM, Escudero JM, Marrades R, Viñolas N, Carcereny E, et al. Mucins CA 125, CA 19.9, CA 15.3 and TAG-72.3 as tumor markers in patients with lung cancer: Comparison with CYFRA 21-1, CEA, SCC and NSE. *Tumor Biol* 2009;29:371-80. <https://doi.org/10.1159/000181180>

[12] Wu L, Chang W, Zhao J, Yu Y, Tan X, Su T, et al. Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clin Cancer Res* 2010;16:3760-8. <https://doi.org/10.1158/1078-0432.CCR-10-0193>

[13] Molina R, Augé JM, Bosch X, Escudero JM, Viñolas N, Marrades R, et al. Usefulness of serum tumor markers, including progastrin-releasing peptide, in patients with lung cancer: Correlation with histology. *Tumor Biol* 2009;30:121-9. <https://doi.org/10.1159/000224628>

[14] Korse CM, Holdenrieder S, yi Zhi X, Zhang X, Qiu L, Geistanger A, et al. Multicenter evaluation of a new progastrin-releasing peptide (ProGRP) Immunoassay across Europe and China. *Clin Chim Acta* 2015;438:388-95.
<https://doi.org/10.1016/j.cca.2014.09.015>

[15] Molina R, Filella X, Augé JM, Fuentes R, Bover I, Rifa J, et al. Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell

lung cancer as an aid in histological diagnosis and prognosis: Comparison with the main clinical and pathological prognostic factors. *Tumor Biol* 2003;24:209-18. <https://doi.org/10.1159/000074432>

[16] Shiels MS, Pfeiffer RM, Hildesheim A, Engels EA, Kemp TJ, Park JH, et al. Circulating inflammation markers and prospective risk for lung cancer. *J Natl Cancer Inst* 2013;105:1871-80. <https://doi.org/10.1093/jnci/djt309>

[17] Qiu J, Choi G, Li L, Wang H, Pitteri SJ, Pereira-Faca SR, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J Clin Oncol* 2008;26:5060-6. <https://doi.org/10.1200/JCO.2008.16.2388>

[18] Sin D, Tammemagi CM, Lam S, Barnett MJ, Duan X, Tam A, et al. Pro-surfactant protein B as a biomarker for lung cancer prediction. *J Clin Oncol* 2013;31:4536-43. <https://doi.org/10.1200/JCO.2013.50.6105>

[19] Yee J, Sadar MD, Sin DD, Kuzyk M, Xing L, Kondra J, et al. Connective tissue-activating peptide III: A novel blood biomarker for early lung cancer detection. *J Clin Oncol* 2009;27:2787-92. <https://doi.org/10.1200/JCO.2008.19.4233>

[20] Molina R, Marrades RM, Augé JM, Escudero JM, Viñolas N, Reguart N, et al. Assessment of a combined panel of six serum tumor markers for lung cancer. *Am J Respir Crit Care Med* 2016;193:427-37. <https://doi.org/10.1164/rccm.201404-0603OC>

[21] Triphuridet N, Vidhyarkorn S, Worakitsitisoran A, Sricharunrat T, Teerayathanakul N, Auewarakul C, et al. Screening values of carcinoembryonic antigen and cytokeratin 19 fragment for lung cancer in combination with low-dose computed tomography in high-risk populations: Initial and 2-year screening outcomes. *Lung Cancer* 2018;122:243-8. <https://doi.org/10.1016/J.LUNGCAN.2018.05.012>

[22] Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* 2015;10:1243-60. <https://doi.org/10.1097/JTO.0000000000000630>

[23] Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J Thorac Oncol* 2022;17:362-87. <https://doi.org/10.1016/J.JTHO.2021.11.003>

[24] Kobayashi S, Tsuta K, Sekine S, Yoshida A, Sasaki B, Shibuki Y, et al. Pulmonary neuroendocrine tumors with nuclear inclusion. *Pathol Res Pract* 2013;209:574-7. <https://doi.org/10.1016/J.PR.P.2013.06.016>

[25] Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The International Association for the Study of Lung Cancer, Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11:39-51. <https://doi.org/10.1016/J.JTHO.2015.09.009>

[26] Molina R, Barak V, Van Dalen A, Duffy MJ, Einarsson R, Gion M, et al. Tumor markers in breast cancer- European Group on Tumor Markers recommendations. *Tumour Biol* 2005;26:281-93. <https://doi.org/10.1159/000089260>

[27] Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72. <https://doi.org/10.1002/SIM.2929>

[28] Peirce CS. The numerical measure of the success of predictions. *Science* 1884;4:453-4. <https://doi.org/10.1126/SCIENCE.NS-4.93.453-A>

[29] Ke SP, Chen SM, Jiang Y, Gong HX, Yu JL, Li X, et al. Bibliometric and visualized analysis of applying tumor markers in lung cancer diagnosis from 2000 to 2022. *Cancer Innov* 2023;2:265-82. <https://doi.org/10.1002/CAI2.74>

[30] The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *N Engl J Med* 2011;365:395-409. <https://doi.org/10.1056/nejmoa1102873>

[31] Shen H. Low-dose CT for lung cancer screening: opportunities and challenges. *Front Med* 2018;12:116-21. <https://doi.org/10.1007/S11684-017-0600-1>

[32] Dao J, Conway PJ, Subramani B, Meyyappan D, Russell S, Mahadevan D. Using cfDNA and ctDNA as Oncologic Markers: A Path to Clinical Validation. *Int J Mol Sci* 2023;24:13219. <https://doi.org/10.3390/IJMS241713219>

[33] Molina R, Filella X, Torres MD, Ballesta AM, Mengual P, Cases A, et al. SCC antigen measured in malignant and nonmalignant diseases. *Clin Chem* 1990;36:251-4. <https://doi.org/10.1093/clinchem/36.2.251>

[34] Hatem R, Simmons B, Thornton JE. A Call to Address AI “Hallucinations” and How Healthcare Professionals Can Mitigate Their Risks. *Cureus* 2023;15:e44720. <https://doi.org/10.7759/CUREUS.44720>

[35] Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-35. <https://doi.org/10.1161/CIRCULATIONAHA.106.672402>

[36] Callister MEJ, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British Thoracic Society Pulmonary Nodule Guideline Development Group, on behalf of the British Thoracic Society Standards of Care Committee. *Thorax* 2015;70 Suppl 2:ii1-54. <https://doi.org/10.1136/THORAXJNL-2015-207168>

[37] Escudero JM, Auge JM, Filella X, Torne A, Pahisa J, Molina R. Comparison of serum human epididymis protein 4 with cancer antigen 125 as a tumor marker in

patients with malignant and nonmalignant diseases. Clin Chem 2011;57:1534-44. <https://doi.org/10.1373/clinchem.2010.157073>

FIGURE LEGENDS

Figure 1. Composition of the whole study population and diagnostic subgroups.

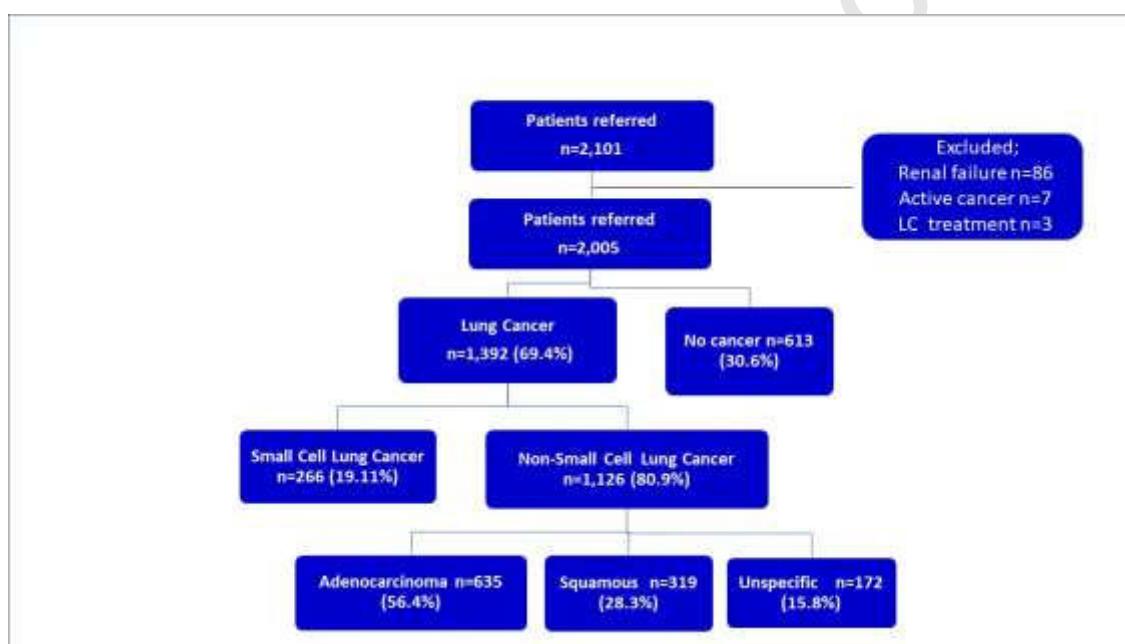
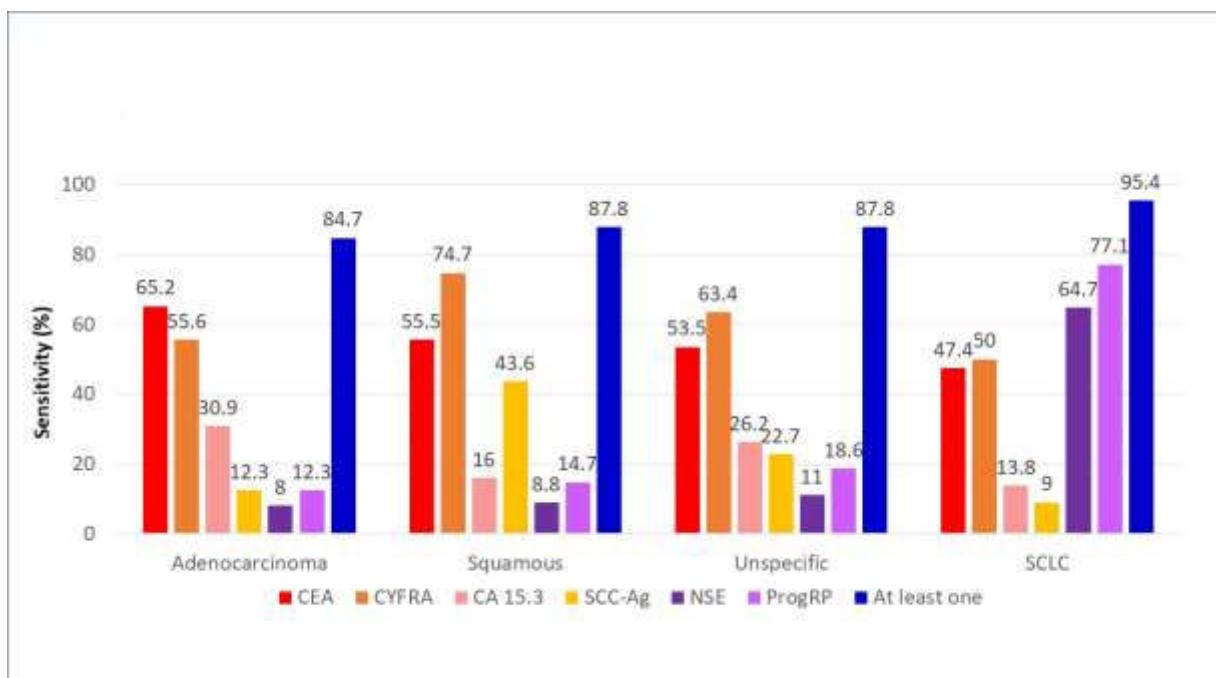


Figure 2.

2A. The sensitivity of the tumour biomarkers in relation to histology.



2B. Serum tumour biomarker sensitivity is subdivided according to tumour stage in NSCLC.

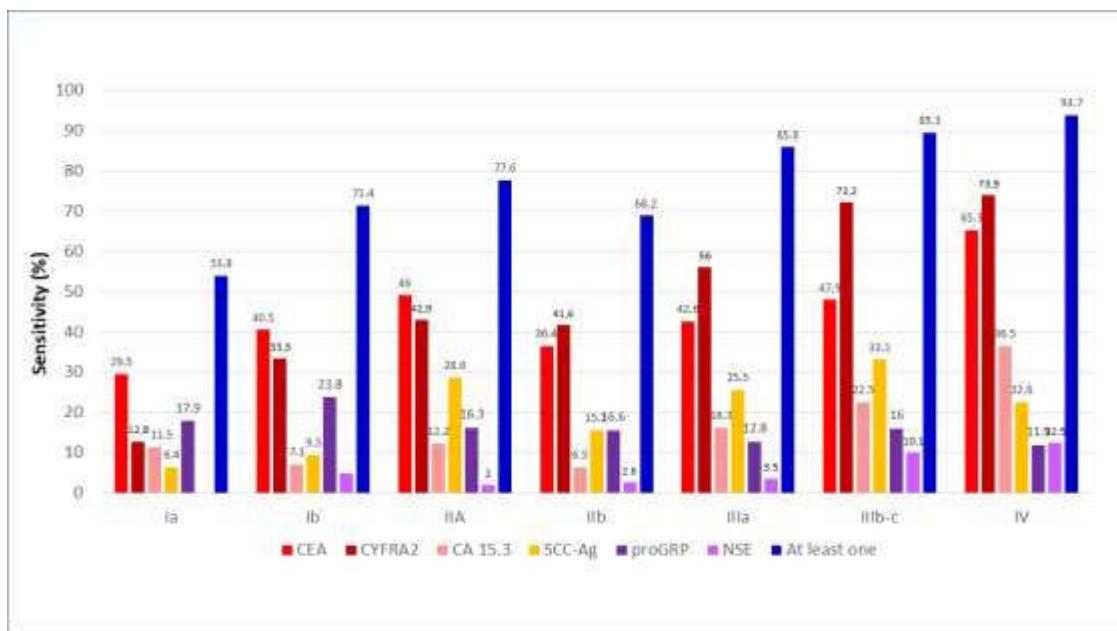
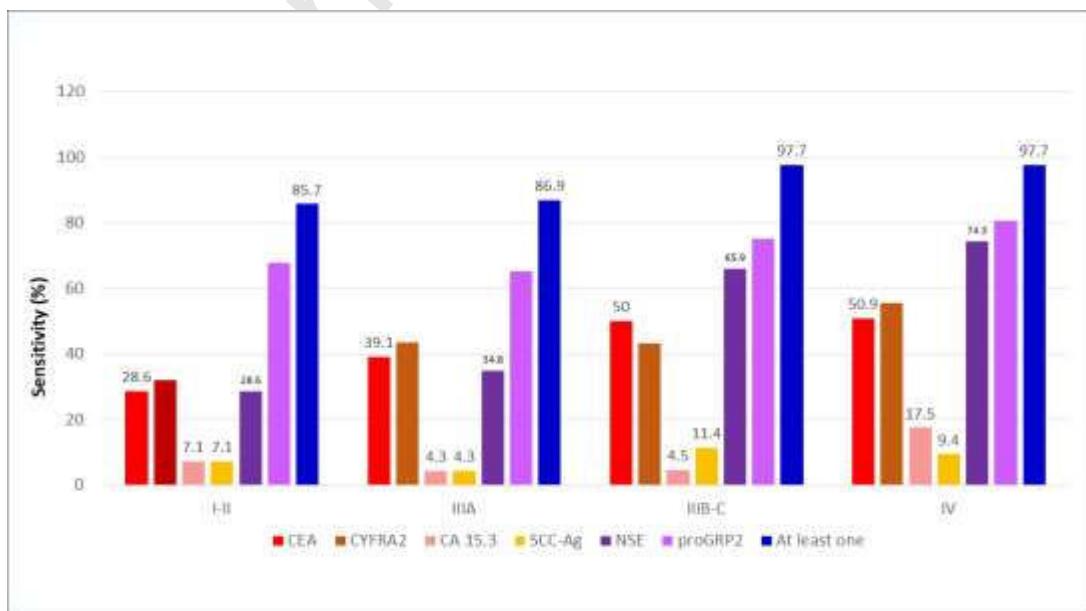


Figure 3.

3A. Serum tumour biomarker sensitivity is subdivided according to tumour stage in SCLC.



3B. Probability of lung cancer according to serum tumour biomarker levels and nodule size (CT scan).

Adenocarcinoma ADC, Squamous cell carcinoma SCC, Unspecific non-small-cell lung cancer uNSCLC, and SCLC= Small-cell lung cancer

CA15.3 = carbohydrate antigen 15.3; CEA = carcinoembryonic antigen; CYFRA 21-1 = cytokeratin-19 fragment, NSE = neuron-specific enolase; ProGRP = pro-gastrin-releasing peptide; SCC-Ag = squamous cell carcinoma-associated antigen. For further explanations, see the text.

OR: Odd ratio; cm: centimetre

TB: tumour biomarker; T1: Initial analysis; T2: Second analysis 3-4 weeks later; LC: Patients with final diagnosis of lung cancer; No LC: Patients with final diagnosis of no lung cancer.

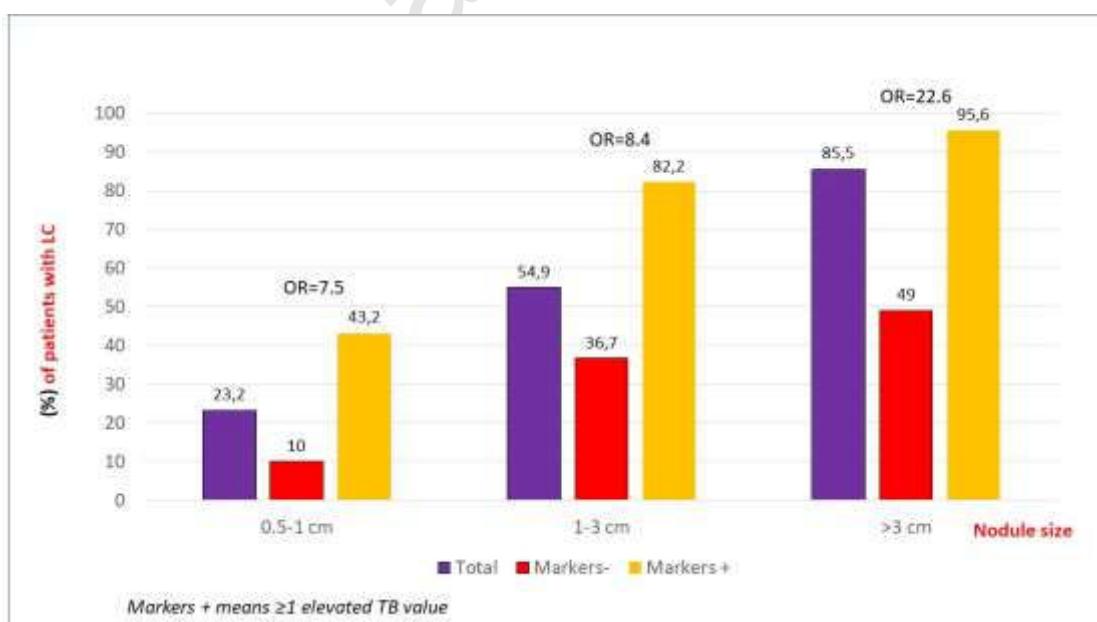


Table 1. The main symptoms of patients referred for suspicious signs of LC and the proportion of LC diagnosed according to the symptoms.

Table 2. Clinical characteristics and tumour biomarker values in all participants (n, percentage, or median [interquartile range]).

SCLC: small-cell lung cancer; NSCLC: non-small-cell lung cancer; cm: centimetre
NS: non-significant. * $p=0.01$ and ** $p<0.0001$ vs. no cancer; * patients without metastases.

Table 3. Sensitivity, specificity, negative predictive values (NPV), and positive predictive values (PPV) of each TB tumour biomarker investigated (upper panel), as well as their evaluation combined with the use of the CLAUDIA algorithm.

Table 4. Tumour biomarker values (median [interquartile range]) stratified by nodule size and type (benign vs. cancer). To avoid potential bias due to the presence of metastasis, patients with stage IV lung cancer were excluded from this analysis. For further explanations, see the text. (2005 patients).

Table 5. Confusion matrix comparing algorithm-suggested histological classifications with definitive diagnoses in lung cancer specimens. Absolute frequencies are shown for each diagnostic category (no cancer, SCLC, uNSCLC, ADC, SCC), along with the percentage of patients with cancer (% patients with cancer) and concordance rates for each group.

Table 1. Main symptoms of patients referred for suspicious signs of lung cancer (LC) and proportion (%) of LC diagnosed according to the symptom.

Main Symptom	Total patients	LC confirmed	
Radiographic nodules	1000	49.8%	651 46.7%
Dyspnoea	219	10.9%	156 11.2%
Haemoptysis	205	10.2%	131 9.4%
Thoracic pain	200	9.9%	154 11%
Persistent cough	197	9.8%	146 10.4%
Constitutional symptoms	79	3.9%	70 5.0%
Persistent Fever	21	1.0%	14 1.0%
Dysphonia	17	0.8%	16 1.1%
Other symptoms	67	3.3%	54 3.8%

Table 2. Clinical characteristics and tumour biomarker values in all participants (n, percentage or median [interquartile range]).

	No cancer (n=613)	p valu es	Lun g ca nc er (n =1, 39 2)		NSCLC (n=1,126)	p values	S CL C (n= 266)
Females, %	37.8%	<0.01	29.7%		29.5%	NS	30.6%
Age, yrs.	64 [56-74]	NS	66 [60-73]		67 [60-74]	0.01	65 [59-71]
Current smokers, %	37.7%	<0.00 1	48.9%		46%	<0.001	61.3%
Former smokers, %	29.4%		34.2%		35.4%		28.9%
Never smoked, %	33%		16.9%		18.6%		9.8%
Pack-yrs.	39 [23-50]	<0.001	45 [30-60]		45 [30-60]	<0.001	49 [30-62]
Presence of nodule	349 (56.9%)	<0.001	651(46.8%)		556 (49.4%)	NS	95 (35.7%)
<1 cm	88(25.2%)	<0.00 1	26(4.1%)		24 4.4%	NS	2 (2.1%)
1-3 cm	194(54.6%)		235(36.2%)		203 (56.5%)		32 (33.6%)
>3 cm	67(19.1%)		390(59.9%)		329 (59.1%)		61 (64.2%)
CEA (ng/ml)	2.2 [1.4-3.4]	<0.001	5.5 [2.6-22.6]**		6.1 [2.6-24.5] **	0.001	4.8[2.2-12.8]*
CYFRA 21-1 (ng/ml)	1.8 [1.3-2.6]	<0.001	4[2.4-8]**		4.4 [2.5-8.7]**	0.001	3.4[2.3-5.3]**
SCC-Ag (ng/ml)	1.1 [0.8-1.5]	<0.001	1.2 [0.8-2.2]**		1.3 [0.9-2.4]**	0.001	1[0.7 -1.5]
CA 15.3 (U/ml)	15 [10-22]	<0.001	21 [13.8-34]**		22 [14-36]**	0.001	19 [12-26.5]
NSE (ng/ml)	12 [10-14.3]	<0.001	14.4 [11.7-		13.5 [11-18.7]**	0.001	39 [20-86]**

			22]**				
ProGRP (pg./ml)	38 [27-49]	<0.001	43.2 [30.6-67]**		39.9 [28.7-54]**	0.001	453 [69.5-1,727]**

SCLC: small-cell lung cancer; NSCLC: non-small-cell lung cancer; cm: centimetre

NS: non-significant. * p=0.01 and **p<0.0001 vs. no cancer; * patients without metastases

Table 3. Sensitivity, specificity, negative predictive values (NPV) and positive predictive values (PPV) of each tumour biomarker investigated (upper panel), as well as their combined evaluation and r with the use of the CLAUDIA algorithm.

	Sensitivity 1,223/1,392	Specificity 463/613	NPV	PPV
INDIVIDUAL ASSESSMENT				
CEA, ng/ml	52.5% 731	91.7 %562	46%	93.5%
CYFRA 21-1, ng/ml	59.8% 832	89.6%549	49.5%	92.8%
SCC-Ag, ng/ml	20.1% 280	96.4% 591	34.7%	92.7%
CA 15-3 U/ml	23.5% 327	98.5% 604	36.2%	97.3%
NSE, ng/ml	19.4% 270	99.3% 609	35.2%	98.5%
ProGRP pg/ml	26% 362	91.2%559	35.2%	87.1%
COMBINED ASSESSMENT				
≥1 abnormal TM value (six tumour markers)	89.44% 1245	63.95% 392	72.3%	84.9%
Algorithm	87.7% 1223	75.5% 463	73.03%	89.06%

Table 4. Tumour biomarker values (median [interquartile range]) stratified by nodule size and type (benign vs. cancer). To avoid the potential bias due to the presence of metastasis, patients with stage IV lung cancer were excluded from this analysis. For further explanations, see text (2005 patients).

	Nodule size < 1 cm			Nodule size 1-3 cm			Nodule size >3 cm		
	Be ni g n n	p value	Ca n ce r n	Be nig n n=	p value	Ca nce r n=	Be ni g n n	p val ue	Ca nc er n=
				19		235			39

	= 8 8		= 2 6	4			= 67		0
CEA ng/ml	2.4 [1.3-3.8]	NS	3.1 [2.1-7.5]	2.2 [1.2-3.3]	0.001	3.7 [2.2-7.9]	2.2 [1.4-3.8]	0.001	4.6 [2.5-12.1]
CYFRA 21-1 ng/ml	1.7 [1.3-2.4]	0.001	2.9 [1.6-4.3]	1.9 [1.4-2.7]	0.001	2.8 [1.6-3.8]	1.7 [1.3-2.3]	0.001	3.8 [2.3-7.1]
SCC-Ag ng/ml	1.1 [0.9-1.4]	NS	1.2 [0.8-1.6]	1.1 [0.8-1.5]	0.054	1.2 [0.9-1.7]	1.1 [0.9-1.5]	0.001	1.6 [1-3]
CA 15-3 U/ml	14 [9.6-19.8]	0.012	21.4 [12.9-31.8]	15 [10-21]	0.01	16.2 [11.5-25]	13.5 [9-20.8]	0.001	19[13-27]
NSE ng/ml	12 [10-14]	NS	12 [9.2-19]	12 [10-14.9]	0.03	12 [10.5-18]	12 [10.1-14.8]	0.001	14.1 [11.6-20]
ProGRP pg/ml	39 [29.3-51]	0.076	49 [25.8-66]	35 [26-46.8]	0.009	43 [31.5-61.1]	33 [23.2-44.3]	0.001	42.8 [30-66]

NS: non-significant.

Table 5. Confusion matrix comparing algorithm-suggested histological classifications with definitive final diagnosis. Absolute frequencies are shown for each diagnostic category (no cancer, SCLC, uNSCLC, ADC, SCC), along with the percentage of patients with cancer (% patients with cancer) and concordance rates for each group.

		Histology						Total	Suggested vs diagnosed		Suggested vs diagnosed *uNSCLC and ADC	
		No cancer	SC LC	NSC LC	uNSC LC	ADC	SC C		Concordance with cancer	Concordance with histology	%patients with cancer	Concordance
Classification algorithm	no cancer	464	13	158	24	99	35	635	26,9	73.1%		
	SCLC	4	224	21	4	11	6	249	98,4	90.0%		
	NSCLC	65	19	869	131	486	252	953	93,2	91.2%		
	uNSCLC	16	6	131	30	66	35	153	89,5	19.6%		
	ADC	28	11	488	72	372	44	527	94,7	70.6%	93.5%	79,41176 471
	SCC	21	2	250	29	48	173	273	92,3	63.4%		

Moderate risk NSLS C	56	7	52	10	20	22	11 5	51.3%	8.7%		
moderate risk ADC	23	2	25	3	19	3	50	54.0%	38.0%		
moderate risk SCC	1	1	1	0	0	1	3	66.7%	33.3%		
Total	613	266	1126	172	63 5	31 9	20 05				

SCLC Small cell Lung cancer; NSCLC no-small cell Lung cancer uNSCLC unclassified NSCLC; ADC adenocarcinoma; SCC Squamous cel carcinoma

*indicate patients suggested by algorithm as unclassified NSCLC and ADC and diagnosed as uNSCLC or ADC when algorithm suggest ADC or uNSCLC the prediction of cancer is 93.5% and the concordance 79.4%

Table S1

Sensitivity, specificity, PNV and PPV and according to centre

Hospital	N	Sensitivity	Specificity	NVP	PPV	Accuracy
1	613	84%	81%	72%	90%	83%
2	94	90%	77%	82%	85%	84%
3	46	90%	71%	56%	95%	87%
4	351	82%	68%	64%	84%	77%
5	155	88%	87%	44%	98%	88%
6	320	94%	78%	80%	93%	90%
7	160	92%	67%	84%	81%	82%
8	5	83%	100%	67%	100%	87%
9	96	87%	78%	58%	94%	85%
10	20	94%	75%	75%	93%	90%
11	105	81%	98%	81%	98%	89%
12	36	88%	80%	72%	92%	86%
All	2005	88%	76%	73%	89%	84%

1. Hospital Clinic De Barcelona; 2 Althaia Xarxa Assistencial Universitaria de Manresa; 3 León University Healthcare Complex 4 Hospital Universitario de Donostia; 5 Hospital Universitario Virgen del Rocío, 6 Hospital Universitario Virgen de la Macarena; 7 Hospital San Juan de Dios de Sevilla, 8 University General Hospital of Alicante; 9 Hospital Universitario de Burgos; 10 Centro Hospitalar Universitario Lisboa Central; 11 Hospital general de Segovia; 12 University Hospital Lozano Blesa de Zaragoza

Table S2**Homogeneity between different all centres.**

	Q Statistic	df	p-value	I ²
Sensitivity	27.25	11	0.005	59.6%
Specificity	18.06	11	0.081	39.0%
Homogeneity between different centres with less than 20% of early stage				
	Q Statistic	df	p-value	I ²
Sensitivity	7.79	7	0.352	10.2%
Specificity	12.01	7	0.100	41.7%
Homogeneity between different centres with more than 25% of early stage				
	Q Statistic	df	p-value	I ²
Sensitivity	1.96	3	0.581	0.0%
Specificity	4.79	3	0.188	37.4%

Heterogeneity analysis was performed using Cochran's Q test for all comparisons. The Q statistic is the Chi-squared value, with df representing the degrees of freedom (n-1). The I² statistic quantifies the percentage of total variation that is due to true heterogeneity between centres rather than chance. Centres were classified into clinical subgroups based on the proportion of patients with early-stage disease (IA, IB, IIA, IIB).

Table S3**Distribution of patients according to stage among the centres**

Hospital	% early stage (IA, IB, IIA, IIB)	% advanced disease (IIIA, IIIB, IV)
1	25.0	75
2	15.4	84.6
3	12.8	87.2
4	26.4	73.6
5	15.0	85
6	12.0	88.0
7	13.4	86.6
8	0.0	100
9	17.9	12.1
10	37.0	63.0
11	13.1	86.1

12	30,8	69,2
----	------	------

1. Hospital Clinic De Barcelona; 2 Althaia Xarxa Assistencial Universitaria de Manresa; 3 León University Healthcare Complex 4 Hospital Universitario de Donostia; 5 Hospital Universitario Virgen del Rocío, 6 Hospital Universitario Virgen de la Macarena; 7 Hospital San Juan de Dios de Sevilla, 8 University General Hospital of Alicante; 9 Hospital Universitario de Burgos; 10 Centro Hospitalar Universitario Lisboa Central; 11 Hospital general de Segovia; 12 University Hospital Lozano Blesa de Zaragoza