

Article

Real-World Insights into Stage I–III Non-Small Cell Lung Cancer in Spain in the Pre-Immunotherapy Era Using AI Techniques: The IntellyLUNG Study

Jesús Corral Jaime ¹, Javier de Castro ² , Aitor Azkarate ³ , Gema García Ledo ^{4,5} , Antonio Calles ⁶, Raquel Marsé ³ , Ana Sofia de Freitas Matos Parreira ³ , Julia Villamayor ², Laura Gutiérrez-Sainz ² , Javier-David Benítez-Fuentes ^{4,5} , Diego Casado Elía ^{4,5,†}, Natalia Gutiérrez ⁶ , Marta Arregui Valles ⁶, Eduard Sarró ⁷, Noelia López ^{8,*} and Savana Research Group ^{7,‡}

¹ Hospital Universitario de Jerez, 11407 Cádiz, Spain

² Hospital Universitario La Paz, Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), Universidad Autónoma de Madrid (UAM), 28046 Madrid, Spain

³ Hospital Universitario Son Espases, 07120 Palma de Mallorca, Spain

⁴ Hospital Universitario HM Sanchinarro, HM Hospitales, Centro Integral Oncológico Clara Campal (HM-CIOCC), 28050 Madrid, Spain

⁵ Instituto de Investigación Sanitaria HM Hospitales, 28050 Madrid, Spain

⁶ Hospital General Universitario Gregorio Marañón, 28007 Madrid, Spain

⁷ Savana Research S.L., 28004 Madrid, Spain

⁸ Medical Affairs Department, Merck Sharp & Dohme (MSD) Spain, 28027 Madrid, Spain

* Correspondence: noelia.lopez@msd.com; Tel.: +34-690656150

† Current address: Hospital General de Villalba, 28400 Collado Villalba, Spain.

‡ Membership of the Savana Research Group is provided in the Acknowledgments.

Abstract

Treatment of non-small cell lung cancer (NSCLC) has been transformed by immunotherapy and targeted therapies. We aimed to characterize clinical features, treatment patterns, and healthcare resource use in patients with early and locally advanced NSCLC before incorporation of these therapies. This retrospective observational study included adults diagnosed with stage I–III NSCLC at four Spanish hospitals between 2014 and 2018, with follow-up until 2021, using artificial intelligence to extract data from electronic health records. A total of 951 patients were included (34.7% stage I, 16.7% stage II, 48.6% stage III), with a median age of 66 years and 31.9% female. Surgery was performed in 78.5% of stage I, 74.8% of stage II, and 35.5% of stage III patients. Among surgical patients, 62.5% received adjuvant chemo- and/or radiotherapy, 20.8% neoadjuvant therapy, and 15.7% both; among non-surgical patients, chemoradiotherapy was the most common treatment (50.4%). Beyond hospitalization, outpatient visits were the most frequently used healthcare resource. These findings provide a historical benchmark of NSCLC care before introduction of immunotherapy and targeted therapies in these settings, highlighting treatment variability and the need for earlier diagnosis, structured treatment pathways, and multidisciplinary management.

Keywords: NSCLC; natural language processing; electronic health records; health resources; Spain



Academic Editors: César Byron Guevara Maldonado, Victoria López and Diego Fernando Riofrío Luzcando

Received: 18 May 2026

Revised: 30 June 2026

Accepted: 2 July 2026

Published: 5 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In Spain, an estimated 34,908 new cases of lung cancer will be diagnosed in 2026, with 66% of these occurring in men [1]. While incidence in men has decreased in recent

years, rates in women have increased, with an estimated incidence in 2026 that is 2.37 times higher than for 2007 [1]. Non-small cell lung cancer (NSCLC) accounts for 80–90% of all lung cancer cases [2]. Early stage (I–II) and locally advanced (III) disease represent a critical window for providing treatment with curative intent, although 5-year survival rates in stage III disease have been low, ranging between 12% and 41% [3]. For early-stage disease, surgery is the foundation of treatment, with lobectomy recommended in most cases [4]. Historically, adjuvant chemotherapy after surgery was recommended to improve survival in patients with stage II–III tumors, while post-surgery radiotherapy (before chemotherapy) was considered for those with incomplete resection [5]. In unresectable stage III disease, the standard approach was concurrent chemoradiotherapy, with sequential chemo- and radiotherapy when concurrent administration was not possible [4].

The arrival of targeted therapies and immunotherapy has dramatically changed the treatment landscape for both resectable and unresectable NSCLC. For patients with tumors harboring targetable oncogenic alterations, current guidelines recommend targeted therapies, and these patients are generally not considered candidates for immune checkpoint inhibitors [6]. In this context, adjuvant osimertinib is now recommended after complete tumor resection in stage IB–IIIA patients with epidermal growth factor receptor (EGFR) mutation-positive tumors [5]. In unresectable EGFR-mutated disease, recent guideline updates now recommend consolidation therapy with osimertinib, with investigation in this area evolving rapidly [7,8]. In resectable NSCLC without actionable genomic alterations, immunotherapy has been integrated into neoadjuvant, perioperative, and adjuvant settings using immune checkpoint inhibitors [9]. In the neoadjuvant setting, nivolumab plus chemotherapy has demonstrated improved event-free survival and overall survival compared with chemotherapy alone [10]. In perioperative approaches, nivolumab, pembrolizumab, and durvalumab have shown improvements in event-free survival and pathological response rates compared with chemotherapy alone, with pembrolizumab also showing significant overall survival benefit [11–15]. In the adjuvant setting, atezolizumab demonstrated disease-free survival benefit and was approved by the European Medicines Agency for patients whose tumors express programmed death ligand (PD-L1) equal to or more than 50%, while pembrolizumab has demonstrated a significant disease-free survival benefit and was approved regardless of tumor PD-L1 expression status in resectable patients after an adjuvant platinum-based chemotherapy [16,17]. In unresectable stage III disease, consolidation therapy with durvalumab is now the standard of care for patients without progression after chemoradiotherapy with curative intent [5].

These advances demand multidisciplinary management strategies, expanding the range of therapeutic options and clinical decision-making pathways [18]. In this context, understanding current practice requires contextualization within prior treatment paradigms. Therefore, this study aimed to describe the real-world clinical characteristics, treatment strategies, and healthcare resource use in patients with stage I, II, or III NSCLC in Spain diagnosed between 2014 and 2018 and followed until 2021. This period preceded the incorporation of immunotherapies and targeted therapies in early-stage and locally advanced disease, although these treatments were already used in advanced stages during part of the study timeframe. To this end, we used artificial intelligence (AI) techniques, including natural language processing (NLP) and machine learning (ML), to extract data from electronic health records (EHRs). This historical perspective provides a baseline for understanding current challenges and evolving priorities in NSCLC management, including early detection strategies and screening, biomarker testing, multidisciplinary approaches, structured patient pathways, and the emerging needs of long-term survivors.

2. Materials and Methods

2.1. Study Design

IntellyLUNG was a multicenter observational study using retrospective AI-powered data extraction to analyze adult patients with early or locally advanced NSCLC who received care at one of the participating study sites. Source data for the study included unstructured and structured information captured in EHRs of patients from four hospitals in the Spanish National Healthcare Network: Hospital Universitari Son Espases (Mallorca), Hospital Universitario La Paz (Madrid), Hospital General Universitario Gregorio Marañón (Madrid), and Grupo Hospitalario HM (Madrid). Participating centers included a combination of secondary and tertiary care hospitals, as well as a large private hospital network. Using NLP and ML, data were extracted from EHRs spanning the study period of 1 January 2014 to 31 December 2021, which included a diagnosis period from 1 January 2014 to 31 December 2018. For all patients, the date of inclusion in the study (index date) was the timepoint within the diagnosis period when NSCLC was identified earliest in their EHRs. A cross-sectional analysis of all patients was performed at their index date, and a longitudinal analysis was conducted over a follow-up period spanning from the index date (a point between January 2014 to December 2018) until their last EHR available, or the end of the study in December 2021 (Figure 1).

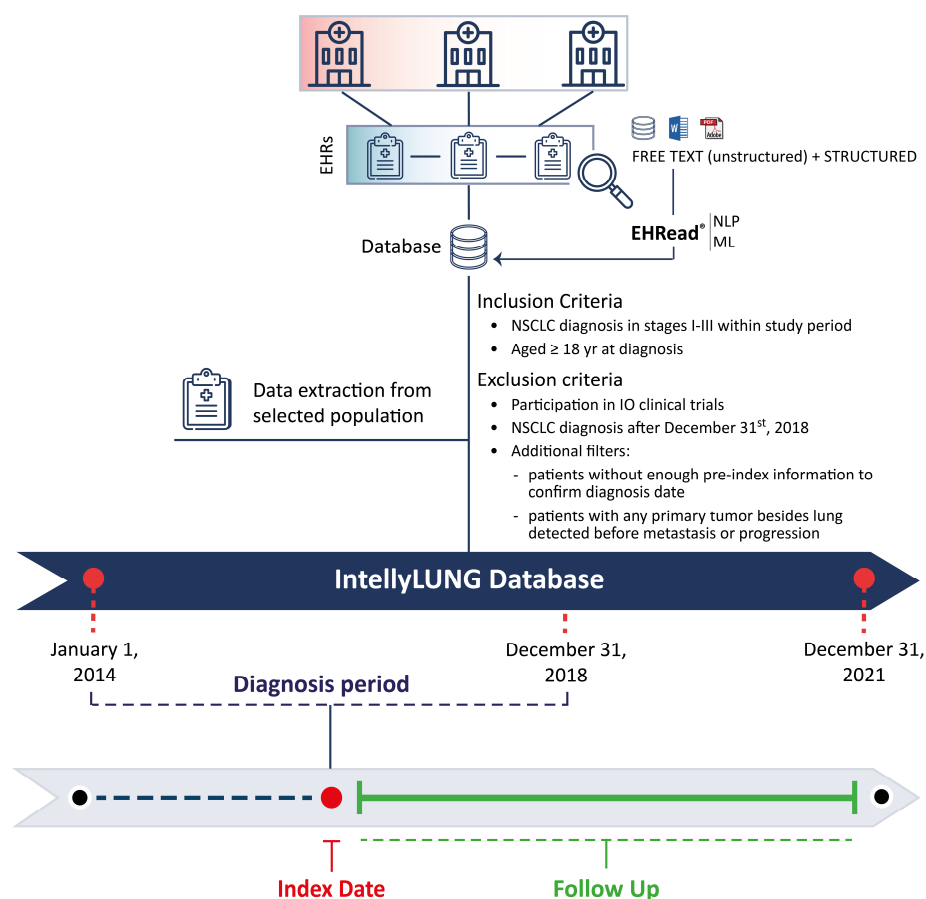


Figure 1. Study design. Structured and unstructured free-text information from EHRs from the 4 participating sites was organized into a study database. Inclusion and exclusion criteria and additional filters were applied to define the study population. Variables were extracted from the database at different time points and analyzed to address the study objectives. EHR, electronic health record; IO, immunotherapy; ML, machine learning; NLP, natural language processing; NSCLC, non-small cell lung cancer.

2.2. Study Population

From the total screening set provided by the participating hospitals, the source population comprised all patients with a clinical diagnosis of NSCLC during the diagnosis period, as determined from information found in EHRs and extracted using NLP. For inclusion in the study population, patients with NSCLC had to have cytologically or histologically confirmed disease of stage I, II, or III, and had to be at least 18 years of age at the time of diagnosis. Patients diagnosed outside of the diagnosis period were excluded from the study population. Patients treated with immunotherapy, either as part of expanded access programs or clinical trials, for early or locally advanced NSCLC were also excluded. After application of the inclusion and exclusion criteria, additional filters were applied to exclude patients for whom the available pre-index information was insufficient to reliably ascertain incident disease status. Specifically, these were patients with few or no clinical records preceding the index date, which limited our ability to distinguish newly diagnosed cases from patients with a prior diagnosis whose earlier disease history was not captured in the available EHR data. Patients with any primary tumor other than lung detected before metastasis or progression were also excluded to avoid misinformation related to other cancers. For analyses, the patients included in the study population were stratified according to NSCLC disease stage at diagnosis and whether or not they underwent surgery.

2.3. Variable Extraction from EHRs

The study variables were selected and validated by clinical experts. Demographic characteristics and healthcare resource use variables were extracted from structured EHR data, while variables related to comorbidities, clinicopathological characteristics (including histological subtype and biomarker profile), surgery type, treatments, imaging tests, and diagnostic procedures were derived from unstructured free text. For each variable assessed at a given time point, the observation nearest to that time point within the predefined reference window was considered. These reference windows were designed to accommodate variations in routine clinical practice across patients, clinicians, and healthcare institutions, thereby maximizing the availability and capture of relevant information from EHRs. Detailed definitions of the reference windows applied to each variable or category are provided in the corresponding table footnotes.

Extraction and analysis of clinical data recorded as unstructured free text in anonymized EHRs, together with their contextual information, was performed using EHRead[®] technology, which leverages NLP and ML techniques to transform free-text data into organized, standardized outputs using SNOMED CT terminology (Medsavana S.L., Madrid, Spain) [19–26]. The source EHRs consisted of all outpatient clinical records, discharge reports, emergency reports, and other available clinical notes and records from each participating site. A predefined list of SNOMED CT terms corresponding to the study variables was developed, incorporating synonyms and acronyms commonly used in clinical practice. The system automatically detected these entities within unstructured clinical narratives, mapped them to standardized SNOMED CT concepts, and enriched them with contextual attributes, including negation and temporality. The extracted entities and their associated metadata (e.g., clinical department, document type, and contextual attributes) were subsequently integrated into a structured database, which served as the basis for deriving and analyzing the study variables. The clinical accuracy of the conceptual definitions and entity mapping was reviewed by medical research experts and oncologists specialized in NLP. The performance of EHRead[®] in accurately extracting key study entities at report level was evaluated using previously published methods [20], with the results summarized in Table S1. All evaluated entities showed F1-scores ranging between 0.80 and 0.98, apart from “stereotactic radiotherapy”, which showed an F1-score of 0.73.

After extraction, variables were reconstructed based on their conceptual definitions by applying dedicated data-wrangling operations to their mapped entities drawing on contextual information provided by EHRead[®]'s NLP models and structured metadata. The NLP models used by EHRead[®] included a set of general models, as well as specific models developed to address the objectives of the study (Supplementary Tables S2 and S3). Disease stage was extracted using two dedicated NLP models (Supplementary Table S3). When multiple classifications were identified, the value closest to the earliest NSCLC mention was selected if separated by >3 months; otherwise, classifications within the same 3-month period were prioritized as follows: specific stage designation, pathological TNM (pTNM), clinical TNM (cTNM), and incomplete or ambiguous TNM.

Further details regarding data source, data processing, data operations to construct relevant variables and filters, and EHRead[®] performance are shown in the Supplementary Materials (Sections S1.1, S1.2, S1.3 and S1.4, respectively).

2.4. Statistical Analysis

Descriptive analyses were conducted on the total study population and stratified by disease stage and surgical treatment. Categorical variables were summarized as frequencies, presenting the number and percentage of patients. Numerical variables were summarized using mean, standard deviation, and median, interquartile range (Q1, Q3). Summary statistics were calculated based on all patients with available values. Department usage rates (visits per 100 patients per year) were calculated as the sum of all observed department visits of each resource divided by the sum of all observed follow-up times. Usage rates are presented as rates per person-year of follow-up. Since anatomic pathology records were received from only 3 of the 4 participating sites, a sensitivity analysis was performed to determine whether the missing pathology data impacted biomarker detection results (Section S1.5 of the Supplementary Materials). Due to the limited number of sites and the need to preserve data anonymity, fully disaggregated center-level analyses are not reported.

3. Results

3.1. Study Population

From a screening set of 4,280,813 patients with 116,612,481 EHRs, 951 adult patients with a known diagnosis of stage I, II, or III NSCLC were included in the study population (Figure 2). A third of the patients in the study population were diagnosed with stage I NSCLC ($n = 330$, 34.7%), and 159 (16.7%) with stage II. Patients diagnosed with locally advanced (stage III) disease comprised nearly half of the study population ($n = 462$, 48.6%, Figure 2). In all patients with stage I, II, or III NSCLC, the median (Q1, Q3) age at diagnosis was 66 years (58, 72), and 303 patients (31.9%) were female (Table 1). More than half were current smokers ($n = 499$, 52.5%), and 134 (14.1%) were former smokers. The most common histological subtype detected in the study population was adenocarcinoma (57.6%), followed by squamous carcinoma (27.9%). The most prevalent comorbidities identified were hypertension (50.9%), chronic obstructive pulmonary disease (42.7%), dyslipidemia (35.3%), diabetes (22.1%), and emphysema (13.7%) (Table 1). Biomarkers were largely untested in the study population (Supplementary Table S4). Of the ten biomarkers examined in the study, only EGFR, anaplastic lymphoma kinase (ALK), and PD-L1 were tested in more than 10% of patients. The most frequently tested was EGFR, which was tested in 316 patients (33.2%), of whom 79 (25.0%) had alterations. ALK was tested in 202 patients (21.2%), of whom 15 (7.4%) had alterations, and tumor proportion score (TPS) for PD-L1 expression was tested in 155 patients (16.3%), of whom 74 (47.7%) had $\geq 1\%$ expression. The proportions of biomarker testing and results detected when considering only the 3 hospitals with

anatomic pathology records included in the study database were similar to those of the overall study population (Supplementary Table S4).

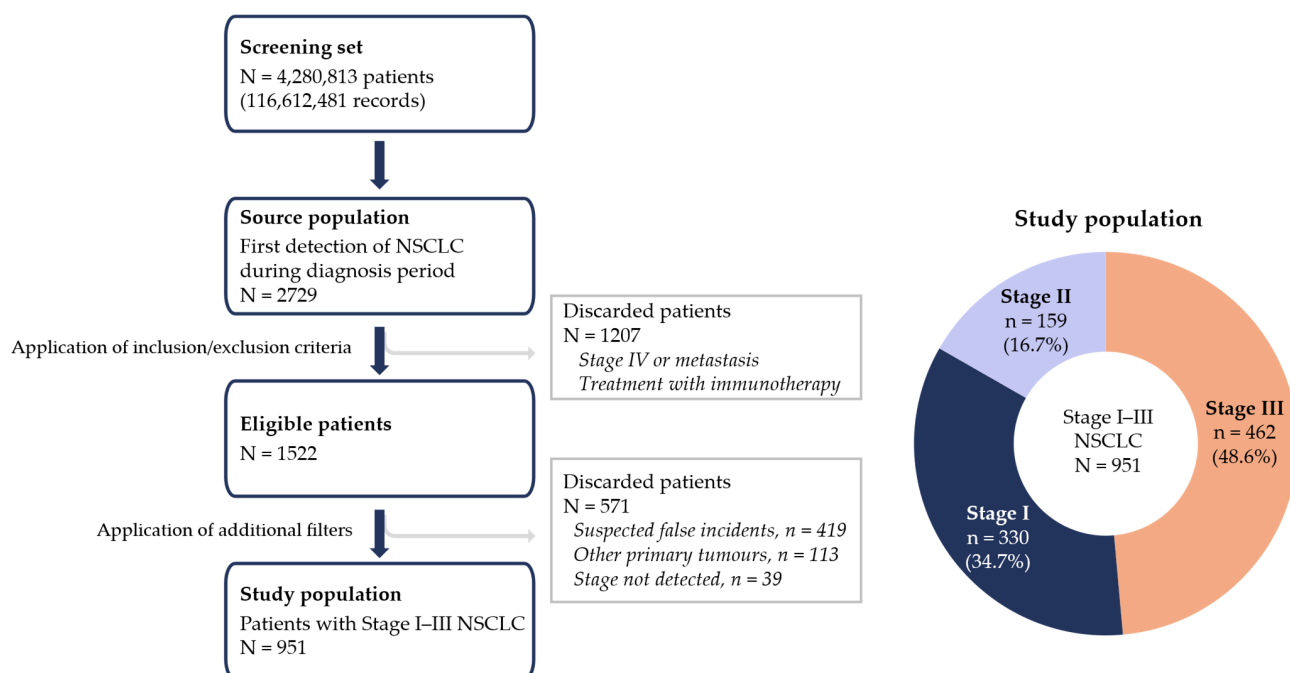


Figure 2. Patient flow chart. From a screening set of 4,280,813 patients, a study population of 951 adult patients with stage I, II, or III NSCLC diagnosed between 2014 and 2018 and followed up until 2021 was selected.

Table 1. Demographics and tumor histological characteristics of patients with NSCLC.

	Stage I n = 330	Stage II n = 159	Stage III n = 462	Total N = 951
Age, median (Q1, Q3)	67 (59, 73)	66 (57, 73)	64 (56, 71)	66 (58, 72)
Females, n (%)	112 (33.9)	50 (31.4)	141 (30.5)	303 (31.9)
Smoking habit, n (%)				
Former	52 (15.8)	29 (18.2)	53 (11.5)	134 (14.1)
Current	165 (50.0)	68 (42.8)	266 (57.6)	499 (52.5)
Never	62 (18.8)	28 (17.6)	61 (13.2)	151 (15.9)
Missing	51 (15.5)	34 (21.4)	82 (17.7)	167 (17.6)
Tumor histology, n (%) [§]				
Adenocarcinoma	213 (64.5)	83 (52.2)	252 (54.5)	548 (57.6)
Squamous carcinoma	67 (20.3)	54 (34.0)	144 (31.2)	265 (27.9)
Missing	47 (14.2)	21 (13.2)	57 (12.3)	125 (13.1)
Comorbidities, n (%) [*]				
Hypertension	178 (53.9)	79 (49.7)	227 (49.1)	484 (50.9)
Chronic obstructive pulmonary disease	152 (46.1)	70 (44.0)	184 (39.8)	406 (42.7)
Dyslipidemia	129 (39.1)	59 (37.1)	148 (32.0)	336 (35.3)
Diabetes	74 (22.4)	29 (18.2)	107 (23.2)	210 (22.1)
Emphysema	43 (13.0)	21 (13.2)	66 (14.3)	130 (13.7)
Ischemic cardiopathy	45 (13.6)	16 (10.1)	52 (11.3)	113 (11.9)
Hypothyroidism	38 (11.5)	18 (11.3)	53 (11.5)	109 (11.5)
Chronic bronchitis	35 (10.6)	16 (10.1)	55 (11.9)	106 (11.1)

Variable values are captured at [First report, End of follow-up], keeping the closest value to index date, for demographic and histological characteristics, and at [First report, Index + 1 year] for comorbidities. [§] Variables with a frequency of less than 5% across all stages and in the overall total are not shown in the table. These include the tumor histologies large cell carcinoma and adenosquamous carcinoma. ^{*} Comorbidities shown are those that were detected in >10% of the study population.

3.2. Surgical and Treatment Approaches

In the study population, 542 patients had undergone thoracic surgery, including 398 (73.4%) who underwent lobectomy, and 64 (11.8%) who underwent pneumonectomy (Table 2). Surgery was performed in 259 of 330 patients with stage I (78.5%) NSCLC, 119 of 159 with stage II (74.8%), and in 164 of 462 (35.5%) with stage III. Of the patients who underwent surgery, 339 (62.5%) received adjuvant chemotherapy and/or radiotherapy treatment, 113 (20.8%) neoadjuvant treatment, and 85 (15.7%) both adjuvant and neoadjuvant treatment (Table 2). Importantly, the recorded stage may reflect downstaging following neoadjuvant treatment in those cases where pathological TNM (pTNM) was prioritized over clinical TNM (cTNM), according to the predefined staging hierarchy (Section S1.3 of Supplementary Materials). Among patients classified as stage I who received neoadjuvant therapy, 12 out of 39 had evidence of a higher earlier clinical stage, and, similarly, among those classified as stage II, 3 out of 16 had a prior stage III. Overall, these 15 patients represented 2.8% of all surgical patients (15/542) and 27.3% of stage I–II surgical patients receiving neoadjuvant therapy (15/55). The potential impact of staging reassignment on neoadjuvant therapy rates is further explored in Supplementary Results and Table S5.

Table 2. Characteristics and clinical management of patients with NSCLC who underwent surgery.

	Stage I n = 259	Stage II n = 119	Stage III n = 164	Total N = 542
Age, median (Q1, Q3)	66 (58, 71)	66 (56, 72)	63 (56, 69)	65 (57, 71)
Females, n (%)	88 (34.0)	40 (33.6)	48 (29.3)	176 (32.5)
Histology, n (%) [§]				
Adenocarcinoma	186 (71.8)	71 (59.7)	105 (64.0)	362 (66.8)
Squamous carcinoma	52 (20.1)	40 (33.6)	51 (31.1)	143 (26.4)
Missing	18 (6.9)	7 (5.9)	5 (3.0)	30 (5.5)
Surgery type, n (%) [§]				
Lobectomy	205 (79.2)	89 (74.8)	104 (63.4)	398 (73.4)
Pneumonectomy	15 (5.8)	18 (15.1)	31 (18.9)	64 (11.8)
Wedge resection	14 (5.4)	2 (1.7)	4 (2.4)	20 (3.7)
Unknown	15 (5.8)	8 (6.7)	22 (13.4)	45 (8.3)
Treatment type, n (%) [§]				
Surgery alone	117 (45.2)	35 (29.4)	23 (14.0)	175 (32.3)
Adjuvant treatment	125 (48.3)	83 (69.7)	131 (79.9)	339 (62.5)
CHT + RDT	44 (17.0)	33 (27.7)	98 (59.8)	175 (32.3)
CHT	37 (14.3)	33 (27.7)	17 (10.4)	87 (16.1)
RDT	26 (10.0)	10 (8.4)	11 (6.7)	47 (8.7)
Unknown *	18 (6.9)	7 (5.9)	5 (3.0)	30 (5.6)
Neoadjuvant treatment	39 (15.1)	16 (13.4)	58 (35.4)	113 (20.8)
CHT + RDT	4 (1.5)	1 (0.8)	12 (7.3)	17 (3.0)
CHT	8 (3.1)	3 (2.5)	11 (6.7)	22 (4.1)
Unknown *	18 (6.9)	8 (6.7)	30 (18.3)	56 (10.3)
Adjuvant and neoadjuvant treatment	22 (8.5)	15 (12.6)	48 (29.3)	85 (15.7)

Variable values are captured at [First report, End of follow-up] for age, sex, and histology; at [Surgery date, Progression] for adjuvant treatments; and at [Index, Surgery date] for neoadjuvant treatments. [§] Variables with a frequency of less than 5% across all stages and in the overall total are not shown in the table. These include the histologies large cell carcinoma and adenosquamous carcinoma, the surgery type segmentectomy, and the neoadjuvant treatment type RDT. * Only mention of ‘adjuvant’ or ‘neoadjuvant’ was detected in free text. CHT, chemotherapy; RDT, radiotherapy; CHT + RDT, chemoradiotherapy. Treatment categories are not mutually exclusive. Patients who received both neoadjuvant and adjuvant therapy are included in both the “Neoadjuvant treatment” and “Adjuvant treatment” rows and are also reported separately under “Adjuvant and neoadjuvant treatment”.

A total of 409 patients in the study population did not undergo surgery (43.0%), of whom 40 (9.8%) received chemotherapy alone, 45 (11.0%) received radiotherapy alone, and 206 (50.4%) received chemoradiotherapy (Table 3). Chemoradiotherapy was detected for 175 patients with stage III disease who did not undergo surgery (58.7%), for 19 (47.5%) with stage II, and for 12 (16.9%) with stage I. Of the 206 patients who received chemoradiotherapy, treatment was administered concomitantly in 147 (71.4%), and sequentially in 51 (24.8%). For 118 patients who did not have surgery (28.9%), no active treatment was detected.

Table 3. Characteristics and clinical management of patients with NSCLC who did not undergo surgery.

	Stage I n = 71	Stage II n = 40	Stage III n = 298	Total N = 409
Age, median (Q1, Q3)	71 (65, 79)	70 (60, 77)	65 (58, 72)	67 (59, 74)
Female, n (%)	24 (33.8)	10 (25.0)	93 (31.2)	127 (31.1)
Histology, n (%) [§]				
Adenocarcinoma	27 (38.0)	12 (30.0)	147 (49.3)	186 (45.5)
Squamous carcinoma	15 (21.1)	14 (35.0)	93 (31.2)	122 (29.8)
Missing	29 (40.8)	14 (35.0)	52 (17.4)	95 (23.2)
Type of treatment, n (%)				
CHT + RDT	12 (16.9)	19 (47.5)	175 (58.7)	206 (50.4)
Concomitant	6 (50.0)	11 (57.9)	130 (74.3)	147 (71.4)
Sequential	3 (25.0)	5 (26.3)	43 (24.6)	51 (24.8)
Unknown	3 (25.0)	3 (15.8)	2 (1.1)	8 (3.9)
CHT alone	8 (11.3)	1 (2.5)	31 (10.4)	40 (9.8)
RDT alone	15 (21.1)	7 (17.5)	23 (7.7)	45 (11.0)
Total CHT (CHT + RDT & CHT)	20 (28.2)	20 (50.0)	206 (69.1)	246 (60.1)
Total RDT (CHT + RDT & RDT)	27 (38.0)	26 (65.0)	198 (66.4)	251 (61.4)
No treatment detected	36 (50.7)	13 (32.5)	69 (23.2)	118 (28.9)

Variable values are captured at [First report, End of follow-up]. Twenty-six patients received neoadjuvant treatment (defined as the mention of neoadjuvant variable) and showed no evidence of having undergone surgery. Therefore, they were classified as patients who did not undergo surgery. [§] Variables with a frequency of less than 5% across all stages and in the overall total are not shown in the table. These include the histologies large cell carcinoma and adenosquamous carcinoma. CHT, chemotherapy; RDT, radiotherapy; CHT + RDT, chemoradiotherapy.

3.3. Healthcare Resource Use

Most patients in the study population were hospitalized at least once during the study period, including 86.4% of patients with stage I NSCLC, 88.7% with stage II, and 79% with stage III disease (Figure 3). Aside from hospitalization, the outpatient clinic was the resource used by the highest percentage of patients in all three disease stage groups (I: 81.2%, II: 73.6% and III: 81.6%), while palliative care was used by the lowest percentage of patients (I: 3.3%, II: 5.7% and III: 7.4%) (Figure 3). Usage rates of healthcare resources varied by treatment group and disease stage (Tables S6). For example, among stage I patients, usage rates for all departments except thoracic surgery (visits/100 patients/year) were numerically higher for those who received multimodal treatment with surgery, chemotherapy, and radiotherapy compared with those who received surgery alone. For patients with stage II NSCLC, usage rates for hospitalization, internal medicine, and palliative care were numerically higher in the surgery-alone group than in the groups that received combined treatment. Among stage III patients, hospitalization, medical oncology, and thoracic surgery usage rates were numerically higher in patients treated with surgery alone than in those who received chemotherapy alone.

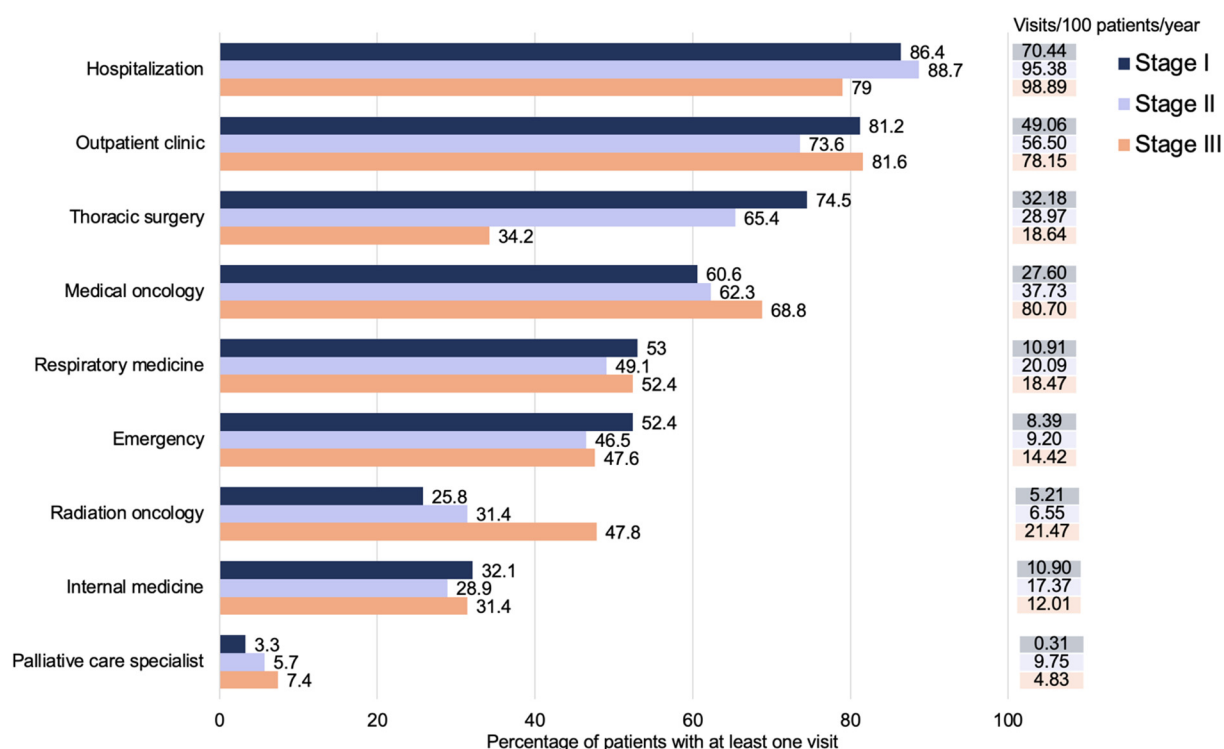


Figure 3. Hospitalization and hospital department visits by patients with NSCLC by disease stage. Colored bars show the percentage of patients hospitalized or with visits to hospital departments. Numbers in the right-hand column show usage rates representing the number of visits per 100 patients per person-year of follow-up.

Regarding imaging tests, computed tomography (CT) was performed on most patients in the study population ($n = 897$, 94.3%), as was positron emission tomography (PET, $n = 788$, 82.9%). Magnetic resonance imaging (MRI) and low-dose CT screening were less used (Table 4). The most common diagnostic procedure performed was biopsy, which was recorded for 716 patients (75.3%), followed by fibrobronchoscopy, recorded for 298 (31.3%) and endobronchial ultrasound (EBUS, 22.2%). There were no notable trends in imaging tests or diagnostic procedures performed when stratified by disease stage (Table 4).

Table 4. Imaging tests and diagnostic procedures in patients with early or locally advanced NSCLC.

	Stage I n = 330	Stage II n = 159	Stage III n = 462	Total N = 951
Imaging test, n (%) [§]				
Computed tomography (CT)	313 (94.8)	146 (91.8)	438 (94.8)	897 (94.3)
Positron emission tomography (PET)	273 (82.7)	136 (85.5)	379 (82.0)	788 (82.9)
Magnetic resonance imaging (MRI)	145 (43.9)	52 (32.7)	199 (43.1)	396 (41.6)
Diagnostic procedure, n (%) [§]				
Biopsy	234 (70.9)	120 (75.5)	362 (78.4)	716 (75.3)
Fibrobronchoscopy	87 (26.4)	48 (30.2)	163 (35.3)	298 (31.3)
Endobronchial ultrasound (EBUS)	45 (13.6)	33 (20.8)	133 (28.8)	211 (22.2)
Mediastinoscopy	9 (2.7)	9 (5.7)	30 (6.5)	48 (5.0)

Variable values are captured at [Index – 4 months, End of follow-up]. Percentages refer to the proportion of patients undergoing each specific and pre-specified procedure within each stage. As patients may have undergone more than one procedure, and additional diagnostic tests may have been performed but were not included, percentages are not mutually exclusive and do not add up to 100%. [§] Variables with a frequency of less than 5% across all stages and in the overall total are not shown in the table. These include the imaging test low-dose CT screening and the diagnostic procedures CT-guided biopsy and endoscopic ultrasound (EUS).

4. Discussion

This study analyzed real-world data extracted from EHRs of adult patients diagnosed with early or locally advanced NSCLC between January 2014 and December 2018, who were followed until December 2021, using NLP- and ML-based methods. It provides a historical benchmark of clinical practice before perioperative immunotherapy and targeted therapies became available in earlier-stage disease management, capturing a pivotal period characterized by heterogeneous treatment approaches and illustrating how management strategies have changed in subsequent years. In contrast to current clinical practice, where treatment selection is increasingly guided by biomarker status and involves multimodal and personalized approaches, the patterns observed here reflect a more limited therapeutic landscape. These findings also highlight the increasing range of therapeutic options now available, and the consequent need for multidisciplinary coordination in decision-making.

The stage distribution of the study population was broadly consistent with the Spanish Thoracic Tumors Registry [27] and reflects the typical presentation of early and locally advanced NSCLC, which is most often detected in stage III when prognosis is worse compared with stage I and II disease [4,27]. These findings reinforce the value of diagnosing NSCLC at stages amenable to curative treatment [3]. During the study period, lung cancer screening had not yet been implemented in Spain, and challenges still remain for its implementation [28]. Current evidence suggests that screening criteria such as those of the United States Preventive Services Task Force (USPSTF) [29] may fail to identify more than a third of lung cancer cases in Spain, including over half of cases in women [30], highlighting the need for country-specific eligibility criteria incorporating factors such as sex, age, and smoking habit [30]. In our study population, patients were predominantly older adults with a high prevalence of current or former smoking, and a substantial proportion of women. The increasing incidence of lung cancer in women in Spain over the past two decades has been attributed to the historical peaks in tobacco smoking that occurred later in women than in men [1,31]. Nevertheless, among never-smokers, women are more frequently affected than men [27]. The substantial percentage of missing smoking data in our study highlights incomplete clinical documentation and underscores the need for improved capture of detailed information for refining future screening criteria, particularly given the rising incidence in women and never-smokers [1,32].

Treatment approaches were heterogeneous across disease stages, likely reflecting differences in care practices between centers, patient or physician preferences, and the complex nature of the disease, with patient-specific factors such as comorbidities, fitness or frailty influencing clinical decisions. Given the descriptive nature of this study, guideline adherence or the clinical reasoning underlying treatment selection could not be assessed; thus, observed variability likely reflects real-world practice patterns rather than deviations from standards. Overall, a substantial proportion of patients did not undergo thoracic surgery, while surgery remained the main treatment approach in early-stage disease group. In stage III disease, surgical management was less frequent, reflecting the heterogeneous composition of resectable and unresectable tumors. These findings are consistent with the 2015 Spanish Society of Medical Oncology (SEOM) clinical practice guidelines for the management of NSCLC [33]. They are also in line with population-based data from the Netherlands covering 2008–2018, where surgery remained the most commonly delivered treatment for stage II NSCLC and a major treatment modality for stage I disease, despite evolving practice patterns over time, while multimodal and non-surgical strategies were more commonly used for stage III disease [34]. In our cohort, adjuvant treatment was commonly used, with chemoradiotherapy being the most frequent modality. Neoadjuvant therapy rates should be interpreted with caution, particularly in patients classified as stage I or II. Because pathological TNM was prioritized when available, some patients may have

been classified according to post-treatment pathological stage rather than the initial clinical stage on which treatment decisions were based. Thus, neoadjuvant therapy in stage I–II may in part reflect downstaging and should not be overinterpreted as evidence of treatment selection in patients initially diagnosed with early-stage disease.

Among non-surgical patients, approximately half received chemoradiotherapy, mostly administered concomitantly, in line with contemporary guidelines for unresectable stage III disease [33], although in our study it was also used in selected stage I–II cases. A substantial proportion had no active treatment recorded. This likely reflects a heterogeneous group, including patients receiving best supportive care due to comorbidities or poor performance status, as well as cases of treatment delivered outside participating institutions or not fully captured in the EHR. This is particularly relevant in stage III disease, where not all patients are eligible for standard concurrent chemoradiotherapy. These factors should be considered when interpreting treatment patterns and associated healthcare resource use. Although early-stage NSCLC is generally resectable, a subset of stage I and II patients may not be eligible for surgery due to comorbidities or advanced age. At the time of the study, stereotactic radiotherapy was recommended for such patients [33]; however, radiotherapy alone was detected in only 21.1% of stage I and in 17.5% of stage II, with no information collected on the radiotherapy technique used. Since then, substantial advances in both surgical and radiotherapy techniques have improved patient outcomes [18,35–37]. Minimally invasive and robotic-assisted surgery, together with improved perioperative care, have reduced morbidity and expanded surgical eligibility [18]. Another critical change in surgical strategy since our study has been the introduction of neoadjuvant and perioperative chemoimmunotherapy. In clinical trials, neoadjuvant chemoimmunotherapy is associated with high rates of definitive surgery (80–85% of patients), with R0 resection in most cases [9], and without adversely affecting surgical feasibility or outcomes [11,13,14]. In neoadjuvant and perioperative approaches, immunotherapy has improved event-free survival, with regimens such as neoadjuvant nivolumab plus chemotherapy, and perioperative pembrolizumab demonstrating significant overall survival benefit [10,12–15]. Although PD-L1 expression has been associated with varying benefit from immunotherapy, perioperative pembrolizumab is currently the only approved regimen irrespective of PD-L1 status [9,12–14]. Ongoing optimization of treatment sequencing [18], together with multidisciplinary tumor board evaluation, will be essential to improve outcomes and guide surgical eligibility, treatment decisions, and coordinated care across specialties [38–40].

During the study period, the absence of immunotherapy and targeted therapies for NSCLC was apparent in the low rates of biomarker testing. Among the biomarkers detected, only EGFR, ALK, and PD-L1 were tested in more than 10% of patients, consistent with the emerging role of EGFR, ALK, and PD-L1 inhibitors in advanced disease. These findings should be interpreted cautiously, as anatomical pathology records were available only for a subset of participating sites, and significantly influenced biomarker detection rates. Therefore, observed testing frequencies may partly reflect under-detection related to data availability rather than true clinical practice. Notably, sensitivity analyses excluding patients from sites without pathology data yielded similar results, suggesting limited overall impact. These low testing rates are consistent with the SEOM recommendations in force during the study period (2015) [33], when molecular testing in early-stage disease was not routinely recommended and was primarily restricted to advanced stages. These guidelines were subsequently updated in 2019 [41] and 2022 [42], but continued to focus mainly on molecular testing in advanced-stage disease, with only limited consideration of its use in earlier-stage settings. In this regard, contemporary European guidelines similarly restricted molecular testing mainly to advanced-stage disease during the study period [43]. As immunotherapy and targeted therapies are increasingly used in earlier stages, early

and comprehensive molecular testing becomes more relevant to guide treatment decisions. Despite this, biomarker testing in Spain and elsewhere remains challenging and under-used [44,45], and current SEOM-SEAP clinical guidelines focus primarily on biomarker testing in stage IV disease [42]. These findings support consideration for extending recommendations to earlier stages and strengthened efforts to ensure adequate tissue sampling for optimal testing.

Healthcare resource use in our study was substantial and was consistent with stage distribution and treatment intensity, particularly the higher use of surgery and oncology services in early-stage disease. These findings align with previous Spanish studies identifying surgery, diagnosis, and chemotherapy as major cost drivers in NSCLC [46,47]. As treatment becomes increasingly personalized, healthcare systems will require coordinated planning to support advanced diagnostics, multimodal care, and long-term survivorship needs [38,48–50].

The main strength of this study is the application of NLP and ML methods to a large real-world dataset (951 patients from over 100 million EHRs), enabling comprehensive characterization of NSCLC care. Rigorous validation procedures and multisource data integration enhanced data robustness and allowed for the capture of multidisciplinary care. Nevertheless, several limitations should be acknowledged. First, the study relies on routinely recorded EHR data, which depends on documentation quality and NLP performance. Disease stage was extracted from EHRs based on explicit mentions of stage (I, II, III) near the index date or converted from TNM values within a defined time window. A dedicated NLP model identified the TNM edition when explicitly reported. Otherwise, the 8th edition was assumed for index dates from 2017 onwards, corresponding to its implementation across participating hospitals. Staging was derived using predefined prioritization rules, with pathological stage prioritized when available. Consequently, the reported stage may in some cases reflect post-treatment pathological staging rather than the initial clinical stage on which treatment decisions were based, particularly after neoadjuvant therapy, and should be considered when interpreting treatment patterns. In addition, treatment variables were reconstructed from unstructured clinical text using NLP-based rules, and surgical procedures may have been incompletely captured if performed outside participating hospitals or not recorded in the available data sources. These factors may explain apparent inconsistencies, such as neoadjuvant or adjuvant therapy in stage I disease or neoadjuvant therapy without recorded surgery. Likewise, the absence of recorded treatment does not necessarily indicate absence of treatment in clinical practice and may reflect incomplete data capture, NLP limitations, or appropriate clinical decision-making. In addition, NLP performance varied across treatment-related concepts. The lowest F1-score was observed for “stereotactic radiotherapy”, mainly due to a site-specific acronym not captured in the lexical variants. However, this reflects variability at the level of a specific treatment subtype and does not correspond to the performance of the broader radiotherapy variables, which were derived using multiple complementary criteria. Therefore, although some degree of residual misclassification cannot be excluded, the impact on the main treatment group classification is expected to be limited. Second, data availability was heterogeneous across sites, as radiology, pathology, and oncology pharmacy records were only available from three of the four sites. Third, certain variables, including detailed radiotherapy protocols, chemotherapy cycles, and granular healthcare resource use by treatment approach, were insufficiently captured for robust analysis. Fourth, healthcare resource use was estimated using the number of records available per department, assuming one record per visit; consequently, visit counts may be overestimated when multiple records were generated during a single clinical encounter; however, it is unlikely to affect comparisons between study groups. Fifth, institutional differences in clinical practices, resource availability, and

documentation may have contributed to variability in treatment patterns and healthcare resource use. These factors may also partly explain the moderate inter-annotator agreement (IAA) for the cohort-defining variables NSCLC and stage (0.66 and 0.77, respectively), reflecting the inherent complexity of annotating real-world clinical records, where interpretation of concept mentions may vary depending on whether external annotators based their judgement on clinical (rather than purely terminological) criteria. Finally, death ascertainment relied on hospital EHRs and did not capture out-of-hospital deaths, potentially leading to underestimation of overall mortality. Consequently, survival analyses were not included in the present study. Despite these limitations, several methodological measures, including predefined variable construction rules, expert validation, and sensitivity analyses, were implemented to minimize potential biases, supporting cautious interpretation of the findings within the context of real-world clinical practice.

5. Conclusions

This study used NLP and ML to analyze real-word data from the EHRs of patients with early and locally advanced NSCLC diagnosed between 2014 and 2018, who were followed until 2021. Our results present a historical snapshot of NSCLC care during a period immediately before integration of immunotherapy and targeted therapy into routine clinical practice for these patients, thus providing insight into how care strategies have evolved and will continue to change in the future. Nearly half of the patients in our study were diagnosed with stage III disease, less than 60% of the study population underwent surgery, biomarkers were largely untested, and most patients were hospitalized at least once. As we move to an era of advanced therapies, our results support care approaches focused on early diagnosis and multidisciplinary collaboration to expand surgical eligibility and determine optimal treatment strategies. In addition, structured patient pathways and coordinated resource planning will be essential for implementing complex workflows and personalized treatment plans to meet the needs of contemporary NSCLC patients and survivors.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/life16071119/s1>, Table S1: Evaluation of EHRead[®] performance for identifying mentions of key variables; Table S2: General NLP models used by EHRead[®]; Table S3: Specific NLP models developed to address study objectives; Table S4: Biomarker profiles by disease stage for patients in the whole study population and in those from the three hospitals that provided anatomical pathology records for the study; Table S5. Impact of staging reassignment on neoadjuvant therapy rates in surgical patients; Table S6: Hospital department usage rates for the study population by treatment group and disease stage.

Author Contributions: Conceptualization: J.C.J., N.L. and Savana Research Group; Methodology: E.S. and Savana Research Group; Software: Savana Research Group; Validation: all authors; Formal analysis: Savana Research Group; Investigation: E.S. and Savana Research Group; Resources: N.L.; data curation: J.d.C., A.A., G.G.L., A.C., R.M., A.S.d.F.M.P., J.V., L.G.-S., J.-D.B.-F., D.C.E., N.G., M.A.V. and Savana Research Group; writing—original draft preparation: E.S., N.L. and Savana Research Group; writing—review and editing: all authors; visualization: E.S. and Savana Research Group; Supervision: J.C.J. and N.L.; project administration: N.L.; funding acquisition: N.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MSD Spain.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Comité de Ética de la Investigación con medicamentos del HOSPITAL UNIVERSITARIO LA PAZ (protocol code PI-5445, date 14 November 2022). The protocol was also reviewed by the corresponding Independent Ethics Committee (IEC), Institutional Review Board

(IRB), and relevant regulatory authorities in accordance with applicable local regulations. As an observational study, it fell under the scope of Royal Decree 957/2020, which governs observational studies with medicinal products for human use in Spain. All methods and analyses followed legal and regulatory requirements and generally accepted research practices described in the latest edition of the Declaration of Helsinki, Good Pharmacoepidemiology Practices, and applicable local regulations. The study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines, with additional consideration of the RECORD (REporting of studies Conducted using Observational Routinely collected Data) extension for studies using routinely collected health data.

Informed Consent Statement: Patient consent was waived because aggregated and anonymized data were analyzed from deidentified EHRs. Data collection and assessment were obtained as part of routine clinical activity and performed in a blinded manner.

Data Availability Statement: The original contributions presented in the study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank Savana Research Group members involved in the study design, development, and analysis: Miren Taberna, Judith Marín-Corral, David Casadevall, Natalia Polo, Paulo Luz, Luisa Martínez, Sebastian Menke, Ignacio Salcedo, Guillermo Arguello, Natalia Iglesias and Adrián Ceja. Medical writing assistance for preparation of this manuscript was provided by Julia Cope.

Conflicts of Interest: MSD Spain was involved in the study design, collection, analysis, and interpretation of data, as well as in the writing of the report and the decision to submit the article for publication. Data extraction and analysis were performed using the EHRead® technology (Medsavana S.L.), and Savana Research S.L. employees contributed to study design, data processing, analysis and writing of the report. Jesús Corral Jaime declares support for the present manuscript from MSD. Javier de Castro declares grants or contracts paid to his institution from AstraZeneca, Bristol Myers Squibb, MSD, and Roche; consulting fees from AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, GlaxoSmithKline, Janssen-Cilag, Lilly, MSD, Novartis, Pfizer, Roche, Takeda, and Sanofi; and honoraria from AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, MSD, Novartis, Roche, Pfizer, Takeda, Sanofi, and Bayer. Aitor Azkarate declares consulting fees from Roche and Boehringer Ingelheim; honoraria from Roche, Boehringer Ingelheim, Lilly, MSD, Bristol Myers Squibb, Amgen, Pfizer, and Regeneron; and support for meetings or travel from Roche, Johnson and Johnson, and Pfizer. Gema María García-Ledo declares funding for the investigation and writing of the present study from MSD; honoraria from AstraZeneca, Bristol Myers Squibb, and Roche Farma; support for meetings or travel from Janssen; and holding stock or stock options in Abbvie. Antonio Calles declares consulting fees from AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly and Company, MSD, Novartis, Pfizer, Roche/Genentech, Takeda, Amgen, Summit Therapeutics, PharmaMar, BeiGene, Janssen, MSD, and Daiichi Sankyo; honoraria from AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, MSD, Novartis, Pfizer, Roche; and support for meetings or travel from Boehringer Ingelheim, Bristol Myers Squibb, MSD, Roche, and Pfizer. Raquel Marsé declares honoraria from Roche, Pfizer, AstraZeneca, Regeneron, and Takeda; and support for meetings or travel from Roche, PharmaMar, Pfizer, and Takeda. Ana Sofia de Freitas Matos Parreira declares payment for a presentation from AstraZeneca; and travel support from MSD, Pfizer, and Novartis. Laura Gutiérrez-Sainz declares support for the present manuscript from MSD; honoraria from Bristol Myers Squibb, Pfizer, Roche, Takeda, Amgen, and AstraZeneca; support for meetings or travel from Roche, Pfizer, AstraZeneca, Takeda, and Janssen; and participation on a data safety monitoring board or advisory board from Roche. Natalia Gutiérrez declares honoraria from MSD and Tactics; and support for meetings or travel from Janssen, Pfizer, and MSD. Marta Arregui Valles declares honoraria from Lilly, AstraZeneca, and Roche, and support for meetings or travel from Lilly, AstraZeneca, and Roche. Eduard Sarró is an employee of Savana Research S.L. Noelia López is an employee of MSD, Spain. Savana Research Group declares that Savana Research Group members are employees of Medsavana S.L or Savana Research S.L. Julia Villamayor, Javier

David Benítez, and Diego Casado Elía declare that they have no conflicts of interest related to the present study.

Abbreviations

The following abbreviations are used in this manuscript:

ALK	Anaplastic lymphoma kinase
BRAF	B-RAF proto-oncogene
EHR	Electronic health record
EGFR	Epidermal growth factor receptor
KRAS	Kirsten rat sarcoma viral oncogene homolog
MET	Mesenchymal–epithelial transition factor
ML	Machine learning
NLP	Natural language processing
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic tyrosine receptor kinase
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
RET	Rearranged during transfection
ROS-1	c-Ros oncogene-1
SNOMED CT	Systematized nomenclature of medicine clinical terms

References

1. Sociedad Española de Oncología Médica. *Las Cifras del Cáncer en España 2026, Depósito Legal: M-3132-2026*; Sociedad Española de Oncología Médica: Madrid, Spain, 2026; ISBN 978-84-09-82151-8.
2. Planchard, D.; Popat, S.; Kerr, K.; Novello, S.; Smit, E.F.; Faivre-Finn, C.; Mok, T.S.; Reck, M.; Van Schil, P.E.; Hellmann, M.D.; et al. Metastatic Non-Small Cell Lung Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* **2018**, *29*, iv192–iv237. [[CrossRef](#)] [[PubMed](#)]
3. Goldstraw, P.; Chansky, K.; Crowley, J.; Rami-Porta, R.; Asamura, H.; Eberhardt, W.E.E.; Nicholson, A.G.; Groome, P.; Mitchell, A.; Bolejack, V.; et al. The IASLC 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. [[CrossRef](#)] [[PubMed](#)]
4. Postmus, P.E.; Kerr, K.M.; Oudkerk, M.; Senan, S.; Waller, D.A.; Vansteenkiste, J.; Escriu, C.; Peters, S. Early and Locally Advanced Non-Small-Cell Lung Cancer (NSCLC): ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.* **2017**, *28*, iv1–iv21. [[CrossRef](#)] [[PubMed](#)]
5. Remon, J.; Soria, J.-C.; Peters, S. Early and Locally Advanced Non-Small-Cell Lung Cancer: An Update of the ESMO Clinical Practice Guidelines Focusing on Diagnosis, Staging, Systemic and Local Therapy. *Ann. Oncol.* **2021**, *32*, 1637–1642. [[CrossRef](#)] [[PubMed](#)]
6. Passaro, A.; Leighl, N.; Blackhall, F.; Popat, S.; Kerr, K.; Ahn, M.J.; Arcila, M.E.; Arrieta, O.; Planchard, D.; De Marinis, F.; et al. ESMO Expert Consensus Statements on the Management of EGFR Mutant Non-Small-Cell Lung Cancer. *Ann. Oncol.* **2022**, *33*, 466–487. [[CrossRef](#)] [[PubMed](#)]
7. Daly, M.E.; Singh, N.; Ismaila, N.; for the Management of Stage III NSCLC Guideline Expert Panel; Daly, M.E.; Singh, N.; Ismaila, N.; Antonoff, M.B.; Arenberg, D.A.; Bradley, J.; et al. Management of Stage III Non-Small Cell Lung Cancer: ASCO Guideline Rapid Recommendation Update. *J. Clin. Oncol.* **2024**, *42*, 3058–3060. [[CrossRef](#)] [[PubMed](#)]
8. Delcuratolo, M.D.; Crespi, V.; Saba, G.; Mogavero, A.; Napoli, V.M.; Garbo, E.; Cani, M.; Ungaro, A.; Reale, M.L.; Merlini, A.; et al. The Evolving Landscape of Stage III Unresectable Non-Small Cell Lung Cancer “between Lights and Shadows”. *Cancer Treat. Rev.* **2025**, *135*, 102918. [[CrossRef](#)] [[PubMed](#)]
9. Phillips, W.J.; Jackson, A.; Kidane, B.; Lim, G.; Navani, V.; Wheatley-Price, P. Immunotherapy for Early-Stage Non-Small Cell Lung Cancer: A Practical Guide of Current Controversies. *Clin. Lung Cancer* **2025**, *26*, 179–190. [[CrossRef](#)] [[PubMed](#)]
10. Forde, P.M.; Spicer, J.; Lu, S.; Provencio, M.; Mitsudomi, T.; Awad, M.M.; Felip, E.; Broderick, S.R.; Brahmer, J.R.; Swanson, S.J.; et al. Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. *N. Engl. J. Med.* **2022**, *386*, 1973–1985. [[CrossRef](#)] [[PubMed](#)]
11. Provencio, M.; Nadal, E.; González-Larriba, J.L.; Martínez-Martí, A.; Bernabé, R.; Bosch-Barrera, J.; Casal-Rubio, J.; Calvo, V.; Insa, A.; Ponce, S.; et al. Perioperative Nivolumab and Chemotherapy in Stage III Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2023**, *389*, 504–513. [[CrossRef](#)] [[PubMed](#)]

12. Cascone, T.; Awad, M.M.; Spicer, J.D.; He, J.; Lu, S.; Sepesi, B.; Tanaka, F.; Taube, J.M.; Cornelissen, R.; Havel, L.; et al. Perioperative Nivolumab in Resectable Lung Cancer. *N. Engl. J. Med.* **2024**, *390*, 1756–1769. [[CrossRef](#)] [[PubMed](#)]
13. Wakelee, H.; Liberman, M.; Kato, T.; Tsuboi, M.; Lee, S.-H.; Gao, S.; Chen, K.-N.; Dooks, C.; Majem, M.; Eigendorff, E.; et al. Perioperative Pembrolizumab for Early-Stage Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2023**, *389*, 491–503. [[CrossRef](#)] [[PubMed](#)]
14. Heymach, J.V.; Harpole, D.; Mitsudomi, T.; Taube, J.M.; Galffy, G.; Hochmair, M.; Winder, T.; Zukov, R.; Garbaos, G.; Gao, S.; et al. Perioperative Durvalumab for Resectable Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2023**, *389*, 1672–1684. [[CrossRef](#)] [[PubMed](#)]
15. Spicer, J.D.; Garassino, M.C.; Wakelee, H.; Liberman, M.; Kato, T.; Tsuboi, M.; Lee, S.-H.; Chen, K.-N.; Dooks, C.; Majem, M.; et al. Neoadjuvant Pembrolizumab plus Chemotherapy Followed by Adjuvant Pembrolizumab Compared with Neoadjuvant Chemotherapy Alone in Patients with Early-Stage Non-Small-Cell Lung Cancer (KEYNOTE-671): A Randomised, Double-Blind, Placebo-Controlled, Phase 3 Trial. *Lancet* **2024**, *404*, 1240–1252. [[CrossRef](#)] [[PubMed](#)]
16. Felip, E.; Altorki, N.; Zhou, C.; Csósz, T.; Vynnychenko, I.; Goloborodko, O.; Luft, A.; Akopov, A.; Martinez-Marti, A.; Kenmotsu, H.; et al. Adjuvant Atezolizumab after Adjuvant Chemotherapy in Resected Stage IB–IIIA Non-Small-Cell Lung Cancer (IMpower010): A Randomised, Multicentre, Open-Label, Phase 3 Trial. *Lancet* **2021**, *398*, 1344–1357. [[CrossRef](#)] [[PubMed](#)]
17. O'Brien, M.; Paz-Ares, L.; Marreaud, S.; Dafni, U.; Oselin, K.; Havel, L.; Esteban, E.; Isla, D.; Martinez-Marti, A.; Faehling, M.; et al. Pembrolizumab versus Placebo as Adjuvant Therapy for Completely Resected Stage IB–IIIA Non-Small-Cell Lung Cancer (PEARLS/KEYNOTE-091): An Interim Analysis of a Randomised, Triple-Blind, Phase 3 Trial. *Lancet Oncol.* **2022**, *23*, 1274–1286. [[CrossRef](#)] [[PubMed](#)]
18. Asmara, O.D.; Hardavella, G.; Ramella, S.; Petersen, R.H.; Tietzova, I.; Boerma, E.C.; Tenda, E.D.; Bouterfas, A.; Heuvelmans, M.A.; Van Geffen, W.H. Stage III NSCLC Treatment Options: Too Many Choices. *Breathe* **2024**, *20*, 240047. [[CrossRef](#)] [[PubMed](#)]
19. Izquierdo, J.L.; Morena, D.; González, Y.; Paredero, J.M.; Pérez, B.; Graziani, D.; Gutiérrez, M.; Rodríguez, J.M. Clinical Management of COPD in a Real-World Setting. A Big Data Analysis. *Arch. Bronconeumol.* **2021**, *57*, 94–100. [[CrossRef](#)]
20. Canales, L.; Menke, S.; Marchesseau, S.; D'Agostino, A.; Del Rio-Bermudez, C.; Taberna, M.; Tello, J. Assessing the Performance of Clinical Natural Language Processing Systems: Development of an Evaluation Methodology. *JMIR Med. Inform.* **2021**, *9*, e20492. [[CrossRef](#)] [[PubMed](#)]
21. Argüello-González, G.; Aquino-Esperanza, J.; Salvador, D.; Bretón-Romero, R.; Del Río-Bermudez, C.; Tello, J.; Menke, S. Negation Recognition in Clinical Natural Language Processing Using a Combination of the NegEx Algorithm and a Convolutional Neural Network. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 216. [[CrossRef](#)] [[PubMed](#)]
22. Muñoz Martín, A.J.; Lecumberri, R.; Souto, J.C.; Obispo, B.; Sanchez, A.; Aparicio, J.; Aguayo, C.; Gutierrez, D.; García Palomo, A.; Benavent, D.; et al. Correction: Prediction Model for Major Bleeding in Anticoagulated Patients with Cancer-Associated Venous Thromboembolism Using Machine Learning and Natural Language Processing. *Clin. Transl. Oncol.* **2025**, *27*, 1857. [[CrossRef](#)] [[PubMed](#)]
23. Muñoz, A.J.; Souto, J.C.; Lecumberri, R.; Obispo, B.; Sanchez, A.; Aparicio, J.; Aguayo, C.; Gutierrez, D.; Palomo, A.G.; Fanjul, V.; et al. Development of a Predictive Model of Venous Thromboembolism Recurrence in Anticoagulated Cancer Patients Using Machine Learning. *Thromb. Res.* **2023**, *228*, 181–188. [[CrossRef](#)] [[PubMed](#)]
24. Blanco-Carrasco, A.J.; Merino-Torres, J.F.; Rubio Almanza, M.; Canovas Molina, G.; Brito-Sanfiel, M.A.; Barajas Galindo, D.E.; Cuellar Olmedo, L.A.; Mauricio, D.; Tofé Povedano, S.; Balsa Barro, J.A.; et al. Characterizing the Clinical Profile and Prevalence of People with Diabetes Attended in the Hospital Setting by Using Unstructured Healthcare Data and Natural Language Processing: The Diabetic@ Study. *Diabetes Res. Clin. Pract.* **2025**, *226*, 112214. [[CrossRef](#)] [[PubMed](#)]
25. Navarro-González, J.F.; Perez De Isla, L.; Cánovas Molina, G.; Brito-Sanfiel, M.Á.; Barajas Galindo, D.E.; Cuellar Olmedo, L.Á.; Mauricio, D.; Tofé Povedano, S.; Balsa Barro, J.A.; Rubio Almanza, M.; et al. Predicting Chronic Kidney Disease in Type 2 Diabetes Using Natural Language Processing on Healthcare Data. *Kidney Dis.* **2025**, *12*, 18–28. [[CrossRef](#)] [[PubMed](#)]
26. Navarro-González, J.F.; Pérez De Isla, L.; Cánovas Molina, G.; Brito-Sanfiel, M.Á.; Barajas Galindo, D.E.; Cuellar Olmedo, L.Á.; Mauricio, D.; Tofé Povedano, S.; Balsa Barro, J.A.; Rubio Almanza, M.; et al. Natural Language Processing of Unstructured Healthcare Data for Predicting Heart Failure in Individuals with Type 2 Diabetes. *J. Clin. Med.* **2026**, *15*, 3287. [[CrossRef](#)] [[PubMed](#)]
27. Provencio, M.; Carcereny, E.; Rodríguez-Abreu, D.; López-Castro, R.; Guirado, M.; Camps, C.; Bosch-Barrera, J.; García-Campelo, R.; Ortega-Granados, A.L.; González-Larriba, J.L.; et al. Lung Cancer in Spain: Information from the Thoracic Tumors Registry (TTR Study). *Transl. Lung Cancer Res.* **2019**, *8*, 461. [[CrossRef](#)] [[PubMed](#)]
28. Ruano-Ravina, A.; Provencio-Pulla, M.; Casan Clarà, P. Lung Cancer Screening with Low-Dose Computerised Tomography. Thoughts on Its Use in Spain. *Med. Clin. (Engl. Ed.)* **2016**, *147*, 366–370. [[CrossRef](#)]
29. US Preventive Services Task Force; Krist, A.H.; Davidson, K.W.; Mangione, C.M.; Barry, M.J.; Cabana, M.; Caughey, A.B.; Davis, E.M.; Donahue, K.E.; Doubeni, C.A.; et al. Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **2021**, *325*, 962–970. [[CrossRef](#)] [[PubMed](#)]

30. Candal-Pedreira, C.; Ruano-Ravina, A.; Calvo De Juan, V.; Cobo, M.; Trigo, J.M.; Carcereny, E.; Cucurull, M.; López Castro, R.; Solís García, E.; Sánchez-Gastaldo, A.; et al. Addressing Lung Cancer Screening Eligibility in Spain Using 2013 and 2021 US Preventive Service Task Force Criteria: Cross-Sectional Study. *ERJ Open Res.* **2023**, *9*, 00468–2023. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Remon, J.; Reguart, N.; García-Campelo, R.; Conde, E.; Lucena, C.-M.; Persiva, O.; Navarro-Martin, A.; Rami-Porta, R. Lung Cancer in Spain. *J. Thorac. Oncol.* **2021**, *16*, 197–204. [\[CrossRef\]](#) [\[PubMed\]](#)
32. LoPiccolo, J.; Gusev, A.; Christiani, D.C.; Jänne, P.A. Lung Cancer in Patients Who Have Never Smoked—An Emerging Disease. *Nat. Rev. Clin. Oncol.* **2024**, *21*, 121–146. [\[CrossRef\]](#) [\[PubMed\]](#)
33. García-Campelo, R.; Bernabé, R.; Cobo, M.; Corral, J.; Coves, J.; Dómine, M.; Nadal, E.; Rodríguez-Abreu, D.; Viñolas, N.; Massuti, B. SEOM Clinical Guidelines for the Treatment of Non-Small Cell Lung Cancer (NSCLC) 2015. *Clin. Transl. Oncol.* **2015**, *17*, 1020–1029. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Evers, J.; De Jaeger, K.; Hendriks, L.E.L.; Van Der Sangen, M.; Terhaard, C.; Siesling, S.; De Ruyscher, D.; Struikmans, H.; Aarts, M.J. Trends and Variations in Treatment of Stage I–III Non-Small Cell Lung Cancer from 2008 to 2018: A Nationwide Population-Based Study from the Netherlands. *Lung Cancer* **2021**, *155*, 103–113. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Shirai, K.; Aoki, S.; Endo, M.; Takahashi, Y.; Fukuda, Y.; Akahane, K.; Musha, A.; Sato, H.; Wakatsuki, M.; Ishikawa, H.; et al. Recent Developments in the Field of Radiotherapy for the Management of Lung Cancer. *Jpn. J. Radiol.* **2025**, *43*, 186–199. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Chang, J.Y.; Mehran, R.J.; Feng, L.; Verma, V.; Liao, Z.; Welsh, J.W.; Lin, S.H.; O'Reilly, M.S.; Jeter, M.D.; Balter, P.A.; et al. Stereotactic Ablative Radiotherapy for Operable Stage I Non-Small-Cell Lung Cancer (Revised STARS): Long-Term Results of a Single-Arm, Prospective Trial with Prespecified Comparison to Surgery. *Lancet Oncol.* **2021**, *22*, 1448–1457. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Soto-Cambres, A.M.; Farré, N. Is SBRT the Optimal First-Line Treatment for Operable Early-Stage NSCLC in Elderly Patients? *Clin. Transl. Oncol.* **2025**, *27*, 3570–3579. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Popat, S.; Navani, N.; Kerr, K.M.; Smit, E.F.; Batchelor, T.J.P.; Van Schil, P.; Senan, S.; McDonald, F. Navigating Diagnostic and Treatment Decisions in Non-Small Cell Lung Cancer: Expert Commentary on the Multidisciplinary Team Approach. *Oncologist* **2021**, *26*, e306–e315. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Ospina, A.V.; Bolufer Nadal, S.; Campo-Cañaveral De La Cruz, J.L.; González Larriba, J.L.; Macía Vidueira, I.; Massutí Sureda, B.; Nadal, E.; Tranco, F.H.; Álvarez Kindelán, A.; Del Barco Morillo, E.; et al. Multidisciplinary Approach for Locally Advanced Non-Small Cell Lung Cancer (NSCLC): 2023 Expert Consensus of the Spanish Lung Cancer Group GECP. *Clin. Transl. Oncol.* **2024**, *26*, 1647–1663. [\[CrossRef\]](#) [\[PubMed\]](#)
40. De Castro, G.; Souza, F.H.; Lima, J.; Bernardi, L.P.; Teixeira, C.H.A.; Prado, G.F. Does Multidisciplinary Team Management Improve Clinical Outcomes in NSCLC? A Systematic Review with Meta-Analysis. *JTO Clin. Res. Rep.* **2023**, *4*, 100580. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Garrido, P.; Conde, E.; De Castro, J.; Gómez-Román, J.J.; Felip, E.; Pijuan, L.; Isla, D.; Sanz, J.; Paz-Ares, L.; López-Ríos, F. Updated Guidelines for Predictive Biomarker Testing in Advanced Non-Small-Cell Lung Cancer: A National Consensus of the Spanish Society of Pathology and the Spanish Society of Medical Oncology. *Clin. Transl. Oncol.* **2020**, *22*, 989–1003. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Isla, D.; Lozano, M.D.; Paz-Ares, L.; Salas, C.; De Castro, J.; Conde, E.; Felip, E.; Gómez-Román, J.; Garrido, P.; Enguita, A.B. New Update to the Guidelines on Testing Predictive Biomarkers in Non-Small-Cell Lung Cancer: A National Consensus of the Spanish Society of Pathology and the Spanish Society of Medical Oncology. *Clin. Transl. Oncol.* **2023**, *25*, 1252–1267. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Zemanova, M.; Pirker, R.; Petruzelka, L.; Zbožíková, Z.; Jovanovic, D.; Rajer, M.; Bogos, K.; Purkalne, G.; Ceriman, V.; Chaudhary, S.; et al. Care of Patients with Non-Small-Cell Lung Cancer Stage III—The Central European Real-World Experience. *Radiol. Oncol.* **2020**, *54*, 209–220. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Mestre-Ferrándiz, J.; Franch Camino, B.; Hidalgo, Á.; Del Llano Núñez-Cortés, A.; Del Llano Señarís, J.E.; Lumbreras, B.; Beas Pedraza, D.; Nuño-Solinís, R.; Paz-Ares, L.; Ramón Y Cajal, S.; et al. Expert-Based Collaborative Analysis of the Situation and Prospects of Biomarker Test Implementation in Oncology in Spain. *Clin. Transl. Oncol.* **2024**, *26*, 985–990. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Fox, A.H.; Osarogiagbon, R.U.; Farjah, F.; Jett, J.R.; Johnson, B.E.; Rivera, M.P.; Smith, R.A.; Wistuba, I.I.; Silvestri, G.A. The American Cancer Society National Lung Cancer Roundtable Strategic Plan: Advancing Comprehensive Biomarker Testing in Non-Small Cell Lung Cancer. *Cancer* **2024**, *130*, 4188–4199. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Corral, J.; Espinàs, J.A.; Cots, F.; Pareja, L.; Solà, J.; Font, R.; Borràs, J.M. Estimation of Lung Cancer Diagnosis and Treatment Costs Based on a Patient-Level Analysis in Catalonia (Spain). *BMC Health Serv. Res.* **2015**, *15*, 70. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Abal Arca, J.; Blanco Ramos, M.Á.; De La Infanta, R.G.; Pérez López, C.; González Pérez, L.; Lamela López, J. Coste hospitalario del diagnóstico del cáncer de pulmón. *Arch. Bronconeumol.* **2006**, *42*, 569–574. [\[CrossRef\]](#)
48. De Castro, J.; Alonso-Fernández, P.; Castrodeza, J.J.; Gayete, Á.; Hernando, F.; Martínez-Olmos, J.; Massuti, B.; Paz-Ares, L.; Sisó-Almirall, A.; Vicente, D.; et al. Monitoring of the Oncological Process for Lung Cancer in Spain: An Expert Consensus Report. *Clin. Transl. Oncol.* **2025**, *27*, 3867–3876. [\[CrossRef\]](#) [\[PubMed\]](#)

49. Hsu, M.L.; Guo, M.Z.; Olson, S.; Eaton, C.; Boulanger, M.; Turner, M.; Miller, M.E.; Nguyen, A.; Szczepanek, K.; Shenolikar, R.; et al. Lung Cancer Survivorship: Physical, Social, Emotional, and Medical Needs of NSCLC Survivors. *J. Natl. Compr. Cancer Netw.* **2024**, *22*. [[CrossRef](#)] [[PubMed](#)]
50. Xiang, Y.; Liu, X.; Wang, Y.; Zheng, D.; Meng, Q.; Jiang, L.; Yang, S.; Zhang, S.; Zhang, X.; Liu, Y.; et al. Mechanisms of Resistance to Targeted Therapy and Immunotherapy in Non-Small Cell Lung Cancer: Promising Strategies to Overcoming Challenges. *Front. Immunol.* **2024**, *15*, 1366260. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.