Study Sponsored by AstraZeneca - Produtos Farmacêuticos
Correspondence author: fernanda.estevinho@ulsm.min-saude.pt

CNO 2025

Digital Approach to Evidence Generation - LANCE-PT

Epidemiological and Clinical Patterns of Lung Cancer in Portugal: A Multicentric Retrospective Study using Electronic Health Records

Fernanda Estevinho[1], Rui Dinis[2], Mafalda Costa[1], Marta Vilaça[1], Luisa Narciso[2], Ulisses Brito[3], Rita Luz[4], Cristina Jácome[4], Tânia Caseiro[5], Ângelo Nunes[5], Ângela Dias[6], Ana M. Dias[6], Inês Moital[6], Filipa Bernardo[6]

[1] ULS Matosinhos, [2] ULS Alto Alentejo, [3]ULS Algarve, [4] MTG Research & Development Lab, [5] IQVIA Solutions Portugal, Unipessoal Lda., [6] AstraZeneca Produtos Farmacêuticos, Lda.

INTRODUTION

In Portugal, **Lung Cancer** (LC) is the leading cause of **cancer-related mortality**[1]. Rapid **advancements** in diagnostics, staging, biomarkers and therapeutics emphasizes the **need for continuous generation of real-world evidence** to optimize clinical practice and patient follow-up[2].

OBJECTIVES

Characterize **epidemiological and clinical patterns of LC** in Portugal through Electronic Health Records (EHRs), encompassing information from patient's medical records, laboratory results, and imaging results.

The **first analysis provides preliminary results from** study centre #1 – **Unidade Local de Saúde de Matosinhos** (ULSM).

METHODOLOGY

Observational, multicentre (target - 15 sites), retrospective study involving analysis of data from EHRs of adult patients diagnosed with LC between 01/01/2015 and 31/12/2024.

The analysis incorporated **structured data**, such as diagnostic and therapeutic procedure codes, and use of **Natural Language Processing** (NLP) employed to process **unstructured information** from EHRs (Figure 1).

An "algorithm-goes-to-data" approach ensures patient-level data was not shared during data processing.

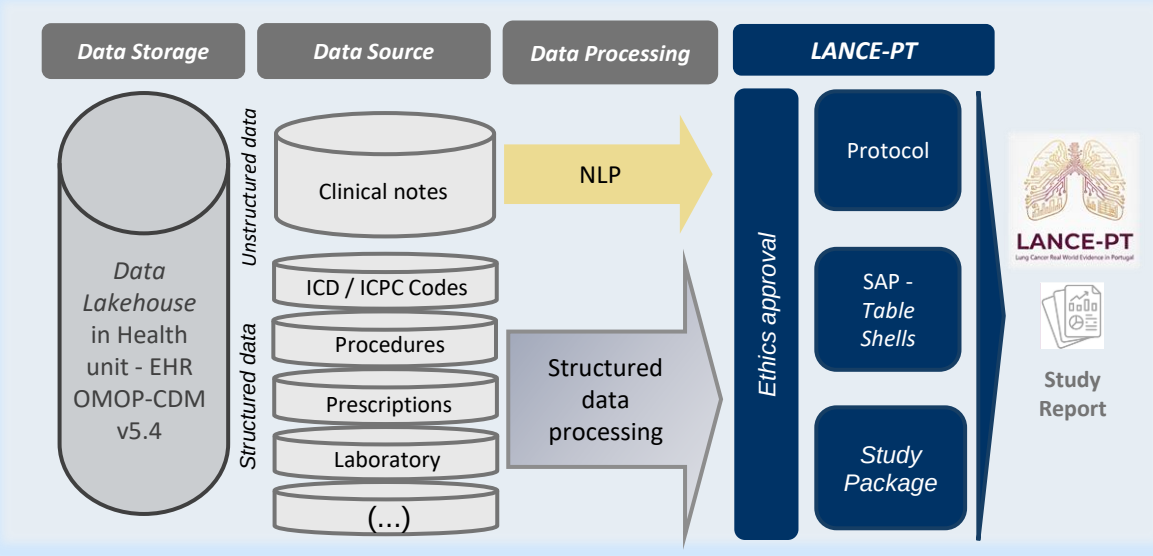

Figure 1 - Data Processing Model (EHR – Electronic Health Records; NLP – Natural Language Processing; SAP-Statistical Analysis Plan)

RESULTS (preliminary results)

This poster presents results of a **preliminary analysis performed in the ULSM Data Lakehouse** (Figure 2 and 3; Table 1 and 2).

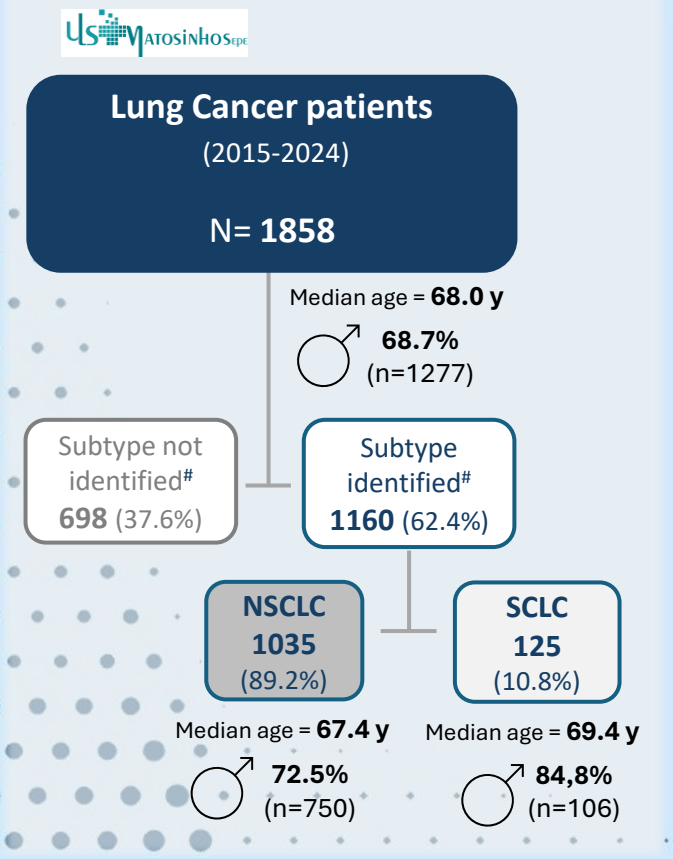

Figure 2 – Lung Cancer cohort - 2015-2024 diagnosis (ICD-9: 162; ICD-10: C34; ICPC-2: R84 with any registry in ULSM) and distribution per subtypes;# Free text search; SCLC- Small Cell Lung Cancer; NSCLC- Non-Small-Cell Lung Cancer

Table 1 – Habits and Clinical characteristics

n(%)	Lung Cancer	SCLC	NSCLC
Smoking habits			
Smoking former	194 (10.4%)	20 (16.0%)	126 (12.2%)
Smoking current	556 (29.9%)	**55 (44.0%)**	290 (28.0%)
Smoking never	**583 (31.4%)**	17 (13.6%)	**291 (28.1%)**
Not classified	**525 (28.3%)**	33 (26.4%)	**328 (31.7%)**
Alcohol abuse			
Alcohol former	43 (2.3%)	≤ 5 (masked)	21 (2.0%)
Alcohol current	506 (27.2%)	**42 (33.6%)**	**288 (27.8%)**
Alcohol never	**572 (30.8%)**	26 (29.8%)	264 (25.5%)
Not classified	737 (39.1%)	53 (42.0%)	462 (44.6%)
Comorbidities			
Obesity	120 (6.5%)	9 (7.2%)	64 (6.2%)
Hypertension	**1 032 (55.5%)**	**76 (60.8%)**	**577 (55.8%)**
Dyslipidemia	**1 621(87.2%)**	**113 (90.4%)**	**950 (91.8%)**
DM type I	46 (2.5%)	0 (0.0%)	22 (2.1%)
DM type II	879 (47.3%)	66 (52.8%)	535 (51.7%)

Table 2 – NSCLC Tumour Characteristics

n(%)	Lung Cancer	SCLC	NSCLC
Grade			
Grade 1	354 (19.1%)	14 (11.2%)	219 (21.2%)
Grade 2	220 (11.8%)	16 (12.8%)	139 (13.4%)
Grade 3	451 (24.3%)	22 (17.6%)	312 (30.1%)
Grade 4	155 (8.3%)	17 (13.6%)	116 (11.2%)
Not classified	678 (36.5%)	56 (44.8%)	249 (24.1%)
Histological characterization NSCLC[&]			
Large cell	-	-	≤ 5 (masked)
Squamous Cell	-	-	273 (26.4%)
Adenocarcinoma	-	-	**805 (77.8%)**
TNM classification NSCLC (n=422; 40.7% of 1035) [$]			
Stage I	-	-	101 (23.9%)
Stage II	-	-	23 (5.5%)
Stage IIIa	-	-	84 (19.0%)
Stage IIIb	-	-	62 (14.7%)
Stage IIIc	-	-	22 (5.2%)
Stage IV	-	-	**130 (30.8%)**

[&] includes patients with both types; [$]Free text search using only TNM terms

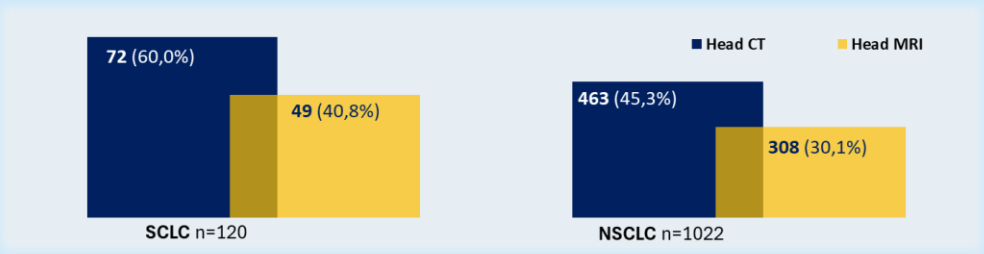

Figure 3– Imaging Procedures per LC subtype (CT - Computed tomography; MRI - Magnetic Resonance Imaging; searched by codes – structure data)

CONCLUSIONS

The systematic analysis of EHRs using **advanced NLP techniques** represents a **new paradigm generation of evidence in LC** and allows a standardized analysis of a **large number of lung cancer** patients;

These findings results will foster **an informed discussion on large-scale LC cohorts and digital methodologies.**

References: 1- RON, I. P. Administração Central do Sistema de Saúde. 2023; 2 -Meyer ML, et al. Annals of Oncology. 2024

Acknowledgements: 1) Research team members, of each participating hospital, not included in the authorship of this poster; 2) IQVIA and MTG team members with contribution to this work and not included in the authorship; ULSM Department of Information Technologies members for conducting this data processing.

Disclaimers: Dias A; Moital I; Bernardo F are AstraZeneca's employees; Dias AM is IQVIA's employee providing service on behalf of AstraZeneca