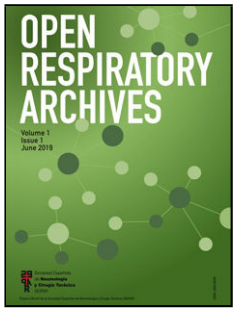
Journal Pre-proof



Synthetic Lung-Cancer Cohorts Generated by a Large Language Model: Epidemiological Validity Assessment

Alvaro Fuentes-Martién MD Julio Mayol MD PhD Barbara Segura Mendez MD PhD Angel Cilleruelo-Ramos MD PhD

PII: S2659-6636(25)00157-2

DOI: https://doi.org/doi:10.1016/j.opresp.2025.100533

Reference: OPRESP 100533

To appear in: *Open Respiratory Archives*

Received Date: 13 September 2025

Accepted Date: 17 December 2025

Please cite this article as: Fuentes-Martién A, Mayol J, Meéndez BS, Cilleruelo-Ramos A, Synthetic Lung-Cancer Cohorts Generated by a Large Language Model: Epidemiological Validity Assessment, *Open Respiratory Archives* (2025), doi: https://doi.org/10.1016/j.opresp.2025.100533

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.

© 2025 Sociedad Española de Neumología y Cirugía Torácica (SEPAR). Published by Elsevier España, S.L.U.

Journal Pre-proof

Short communication

Synthetic Lung-Cancer Cohorts Generated by a Large Language Model: Epidemiological Validity Assessment

Cohortes sintéticas de cáncer de pulmón generadas por inteligencia artificial: evaluación de la validez epidemiológica.

Álvaro Fuentes-Martín, MD
alvarofuentesmartin@gmail.com
ORCID: 0000-0001-7041-4960
Servicio de Cirugía Torácica, Hospital Clínico Universitario de Valladolid.
Universidad de Valladolid

Julio Mayol, MD, PhD
jmayol@ucm.es
ORCID: 0000-0003-2411-6668
Hospital Clínico San Carlos, IdISSC, Universidad Complutense de Madrid

Bárbara Segura Méndez, MD, PhD barbaraseg@hotmail.com
ORCID: 0000-0001-5506-0834
Servicio de Cirugía Cardiaca. Hospital Universitario de Salamanca.

Ángel Cilleruelo-Ramos, MD, PhD ancillera@hotmail.com
ORCID: 0000-0002-1872-0714
Servicio de Cirugía Torácica, Hospital Clínico Universitario de Valladolid. Universidad de Valladolid

Corresponding author

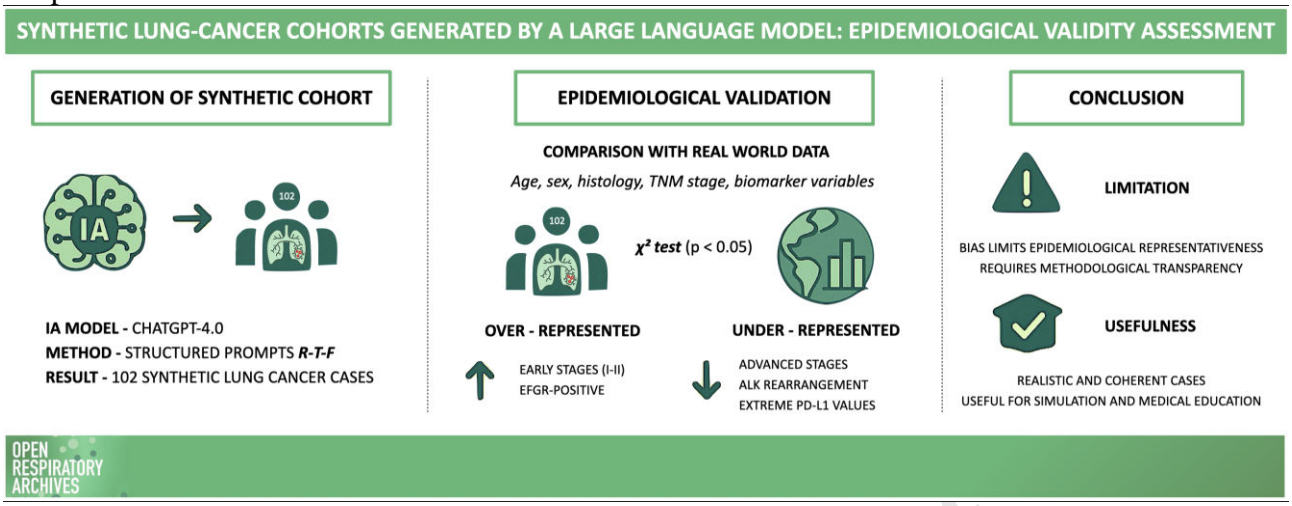
Álvaro Fuentes-Martín
Servicio de Cirugía Torácica, Hospital Clínico Universitario de Valladolid.
Avenida Ramón y Cajal, 47006.
alvarofuentesmartin@gmail.com

ORCID: 0000-0001-7041-4960
Author ID Scopus: 57221197976
ResearcherID: JND-9182-2023

Journal Pre-proof

Journal Pre-proof

Graphical Abstract



Abstract

Large language models (LLMs) are increasingly used in medicine for clinical reasoning and educational simulation. This study assessed the epidemiological plausibility of a synthetic lung-cancer cohort generated by ChatGPT-4.0. A total of 102 virtual cases were created in Spanish using structured prompts including demographic, histologic, and molecular variables. When descriptively compared with international datasets (GLOBOCAN 2020, SEER, and biomarker meta-analyses), the cohort reproduced general disease patterns but showed statistically significant deviations ($p < 0.05$): early-stage disease and EGFR-positive tumors were overrepresented, while advanced stages, ALK rearrangements, and extreme PD-L1 values were underrepresented. These discrepancies likely reflect biases in model training data and the probabilistic nature of generative language models. Despite this quantified generative bias, the utility of these cohorts for non-epidemiological tasks like educational simulation is discussed, provided methodological transparency is maintained.

Abstract

Los modelos de lenguaje de gran escala (LLM) se utilizan cada vez más en medicina para el razonamiento y la simulación clínica. Este estudio evaluó la plausibilidad epidemiológica de una cohorte sintética de cáncer de pulmón generada por ChatGPT-4.0. Se crearon un total de 102 casos sintéticos mediante *prompts* estructurados que incluían variables demográficas, histológicas y moleculares. Al compararla con bases de datos epidemiológicas, la cohorte reprodujo patrones generales de la enfermedad, aunque mostró desviaciones estadísticamente significativas ($p < 0.05$): sobrerrepresentación de estadios iniciales y de EGFR frente a la infrarepresentación de estadios avanzados, reordenamientos ALK y valores extremos de PD-L1. Estas discrepancias reflejan sesgos en el entrenamiento y la naturaleza probabilística de los modelos generativos. A pesar de este sesgo

Journal Pre-proof

generativo cuantificado, se discute la utilidad de estas cohortes para tareas no epidemiológicas como la educación médica, siempre que se mantenga la transparencia metodológica.

Key words

Synthetic Cohorts; Large language models; Thoracic oncology.

Cohortes sintéticas; Modelos de lenguaje de gran escala; Oncología torácica.

Large language models (LLMs) are increasingly applied in medicine, supporting clinical reasoning and educational simulation[1,2]. In oncology, they have been explored as decision-support tools[3,4] and for automated clinical case generation. One of their most promising uses is the creation of synthetic clinical cohorts—realistic yet fictitious datasets that emulate real patient populations while preserving privacy[5]. Despite their potential, the epidemiological representativeness of LLM-generated cohorts remains unvalidated. This study aimed to evaluate the epidemiological plausibility of a synthetic lung-cancer cohort generated by ChatGPT-4.0, comparing key demographic, histologic, and molecular variables against international data.

We conducted a descriptive, exploratory study to assess the internal consistency and external plausibility of data generated by ChatGPT-4.0 (OpenAI). A convenience sample of 102 virtual patients was generated, a size deemed sufficient for an initial exploratory descriptive assessment using Spanish-language prompts structured in a Role–Task–Format framework. Generation occurred in batches of five patients per iteration, reflecting the model's operational text limit. No corrections or parameter adjustments were introduced between batches to preserve methodological consistency.

Prompts instructed ChatGPT to create clinically coherent profiles including demographic, oncologic, and molecular variables: age, sex, histologic subtype, TNM stage, and biomarkers (EGFR, ALK, and PD-L1) *(Supplementary Annex 1)*. These variables were extracted manually and compared descriptively with global reference datasets (*GLOBOCAN 2020, SEER*) and biomarker meta-analyses[6-11] *(Figure 1)*. Continuous variables were summarized as mean ± SD and categorical variables as frequencies. Observed cohort frequencies (e.g., stage, biomarkers) were compared against expected population benchmarks (derived from refs 6-11) using Chi-squared (X^2) ($p < 0.05$ was considered statistically significant).

Journal Pre-proof

The synthetic cohort included 102 virtual patients, with a mean age 66.5 ± 6.0 years (range 51–79). In comparison, global oncology registries[7,8] indicate a median diagnostic age of approximately 70 years, suggesting a slightly younger synthetic population. Sex distribution was balanced (51 men, 51 women), a distribution significantly deviating from the expected male predominance (≈65–70%)[7] (X^2=14.24, $p < 0.01$).

Histologic distribution comprised adenocarcinoma 52%, squamous-cell carcinoma 41%, and small-cell carcinoma 7%. No large-cell carcinoma was generated. Although the adenocarcinoma proportion was similar to that observed in population-based studies[6] (≈45–55%), small-cell carcinoma was underrepresented (10–15% expected), and large-cell carcinoma (≈5–7%) was absent, a distribution significantly deviating from population data (X^2=11.82, $p < 0.01$).

Staging analysis showed a predominance of early-stage disease: 65% (Stages I–II), 17% (Stage III), and 18% (Stage IV). Real-world data indicate that ≈40% of non-small-cell lung-cancer (NSCLC) are diagnosed at Stage IV. In our synthetic NSCLC cohort (N=95), only 18% were Stage IV, confirming a significant over-representation of localized stages (X^2=19.34, $p < 0.01$).

Among the 95 simulated NSCLC cases, 45% harbored activating EGFR mutations, a rate significantly higher than Western prevalence (≈15%[9]; X^2=68.24, $p < 0.01$). No ALK rearrangements were identified (0%), a significant deviation from the expected 3–5% prevalence[10] (X^2=3.96, $p < 0.05$). PD-L1 expression was uniformly intermediate (1–49%), with no negative or highly positive cases (≥50%)— a distribution significantly deviating from clinical cohorts[11] (X^2=142.5, $p < 0.01$).

Taken together, the synthetic cohort reproduced general disease patterns but showed notable deviations in key variables, particularly stage distribution and molecular profile, indicating statistically significant generative bias relative to real-world epidemiological distributions.

This exploratory study evaluated the epidemiological consistency of a synthetic lung-cancer cohort generated by ChatGPT-4.0. Although the model produced clinically plausible, coherent profiles, it exhibited systematic biases—most notably a predominance of early-stage disease and EGFR-positive tumors. These deviations likely reflect the probabilistic nature of generative models and the uneven representation of clinical scenarios in training data.

The over-representation of early-stage cases suggests a narrative bias favoring curative, well-structured clinical stories over terminal presentations. The under-representation of small-cell and absence of ALK-positive cases may relate to their lower frequency and visibility in the scientific

Journal Pre-proof

literature. Likewise, the homogeneous PD-L1 pattern indicates that ChatGPT tends to assign intermediate values under uncertainty, reflecting a limitation in quantitative realism.

Such discrepancies align with prior observations of LLM "hallucinations," i.e., systematic deviations from expected patterns due to the statistical weighting of learned text[12]. These findings underscore the need to impose explicit population-level constraints when generating synthetic datasets. Without controlled proportions or post-generation validation, representativeness cannot be assumed.

From an educational standpoint, however, generating structured, realistic clinical cases retains considerable value. Synthetic cohorts can support virtual tumor-board exercises, simulation-based teaching, and preliminary algorithm testing. Their utility lies more in training and hypothesis generation than in precise epidemiological representation.

Reproducibility represents an additional methodological challenge. Generative models evolve continuously, and outputs obtained with one version may not be replicable with subsequent iterations, even when using identical prompts. Clear documentation of the model version, generation date, and prompting framework is therefore essential to maintain methodological transparency and facilitate reproducibility.

Another limitation concerns the batch-generation process. Because the cohort was created in groups of five patients—an operational restriction of ChatGPT-4.0—sequential generation may have introduced distributional bias. Although no systematic drift was visually observed, future research should quantify inter-batch variability to confirm data stability.

Despite these limitations, LLM-generated cohorts demonstrate the feasibility of producing complex, internally coherent clinical datasets without compromising confidentiality. With appropriate methodological safeguards and validation procedures, this approach could enhance educational realism and foster innovation in thoracic-oncology training.

The synthetic lung-cancer cohort generated with ChatGPT-4.0 reproduced general disease patterns but showed statistically significant deviations in stage distribution and biomarker prevalence. This quantified generative bias limits its epidemiological representativeness, the findings illustrate both the promise and the boundaries of LLM-based data generation. Responsible implementation requires methodological transparency, explicit acknowledgment of biases, and quantitative validation. With continued refinement, large language models could become valuable

Journal Pre-proof

complementary tools for simulation, medical education, and early-phase research in thoracic oncology.

Authors' Contributions

All authors contributed substantially to the design of the study, data analysis, manuscript drafting, and critical revision of its content. All authors have read and approved the final version of the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Statement: This study did not involve real patients or human subjects. Instead, it was based entirely on synthetic data generated through an artificial intelligence model (ChatGPT-4.0), and no identifiable or confidential patient information was used. Therefore, the requirement for informed consent was waived. Nevertheless, the study protocol was reviewed and approved by the Clinical Research Ethics Committee of our institution (Reference: PI–25–146–C), and the project was conducted in accordance with the ethical principles of the Declaration of Helsinki (2013 revision). The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Declaration of Generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the authors used the Generative Pre-trained Transformer 4 (ChatGPT-4) not only for grammar review and translation, but also for the structured generation of a synthetic cohort of 102 virtual patients with lung cancer, based on predefined clinical, molecular, and psychosocial parameters. This simulated dataset was used for research purposes within the framework of this study. After using this tool, the authors reviewed, validated, and edited the output as necessary, and take full responsibility for the content of the publication.

Data Availability Statement

All data used in this study were synthetically generated using the ChatGPT-4.0 model (OpenAI) and do not correspond to real individuals. The full methodology used for data generation is described in the Methods

Journal Pre-proof

section. Examples of the synthetic cases are provided in the supplementary material. No real patient data were accessed or used in this study.

REFERENCES

1. Zitu MM, Le TD, Duong T, Haddadan S, Garcia M, Amorrortu R, et al. Large language models in cancer: potentials, risks, and safeguards. BJR Artif Intell. 2024;2(1):ubae019. doi:10.1093/bjrai/ubae019
2. Fuentes-Martín Á, Cilleruelo-Ramos Á, Segura-Méndez B, Mayol J. Can an artificial intelligence model pass an examination for medical specialists? Arch Bronconeumol. 2023;59(8):534-536. doi:10.1016/j.arbres.2023.03.017
3. Zabaleta J, Aguinagalde B, Lopez I, Fernandez-Monge A, Lizarbe JA, Mainer M, et al. Utility of Artificial Intelligence for Decision Making in Thoracic Multidisciplinary Tumor Boards. J Clin Med. 2025 Jan 10;14(2):399. doi: 10.3390/jcm14020399. PMID: 39860405; PMCID: PMC11765867.
4. Nardone V, Marmorino F, Germani MM, Cichowska-Cwalińska N, Menditti VS, Gallo P, et al. The Role of Artificial Intelligence on Tumor Boards: Perspectives from Surgeons, Medical Oncologists and Radiation Oncologists. Curr Oncol. 2024 Aug 27;31(9):4984-5007. doi: 10.3390/curroncol31090369. PMID: 39329997; PMCID: PMC11431448.
5. Jin Q, Wang Z, Floudas CS, Chen F, Gong C, Bracken-Clarke D, et al. Matching patients to clinical trials with large language models. Nat Commun. 2024 Nov 18;15(1):9074. doi: 10.1038/s41467-024-53081-z. PMID: 39557832; PMCID: PMC11574183.6. Zhang Y, Vaccarella S, Morgan E, et al. Global variations in lung cancer incidence by histological subtype in 2020: a population-based study. Lancet Oncol. 2023;24(11):1206-1218. doi:10.1016/S1470-2045(23)00444-8
7. SEER Cancer Stat Facts: Lung and Bronchus Cancer. National Cancer Institute. Accessed April 2, 2025. https://seer.cancer.gov/statfacts/html/lungb.html
8. Subramanian J, Morgensztern D, Goodgame B, Baggstrom MQ, Gao F, Piccirillo J, et al. Distinctive characteristics of non-small cell lung cancer (NSCLC) in the young: a surveillance, epidemiology, and end results (SEER) analysis. J Thorac Oncol. 2010 Jan;5(1):23-8. doi: 10.1097/JTO.0b013e3181c41e8d. PMID: 19934774.9. Rosell R, Moran T, Queralt C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med. 2009;361(10):958-967. doi:10.1056/NEJMoa0904554
10. Cognigni V, Pecci F, Lupi A, Pinterpe G, De Filippis C, Felicetti C, et al. The Landscape of ALK-Rearranged Non-Small Cell Lung Cancer: A Comprehensive Review of Clinicopathologic,

Journal Pre-proof

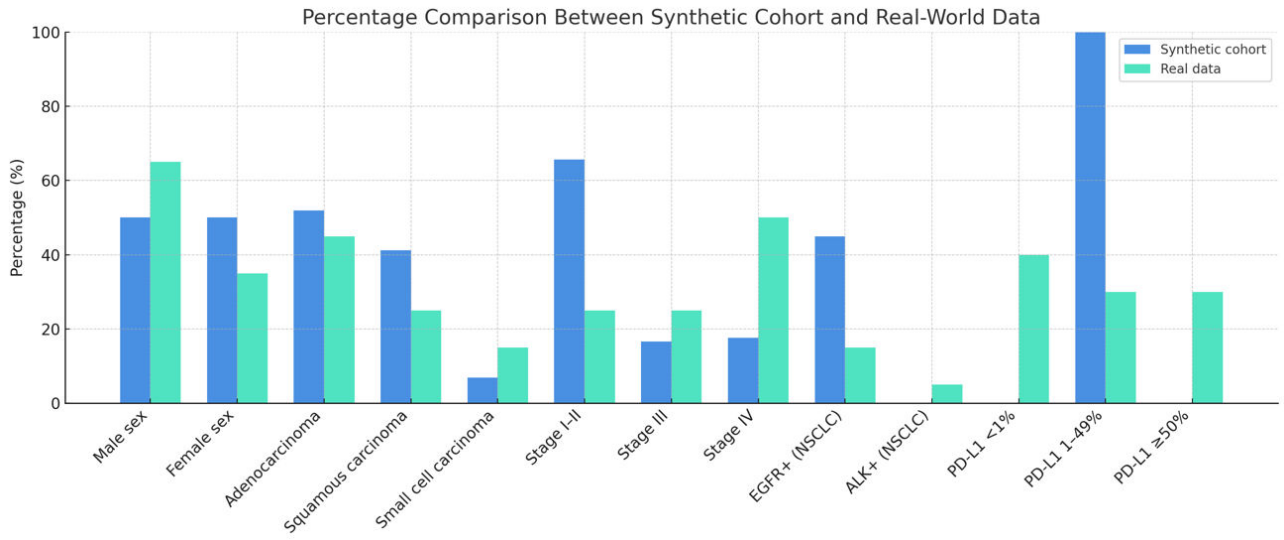
Genomic Characteristics, and Therapeutic Perspectives. Cancers (Basel). 2022 Sep 29;14(19):4765. doi: 10.3390/cancers14194765. PMID: 36230686; PMCID: PMC9563286.

11. Hwang DM, Albaqer T, Santiago RC, Weiss J, Tanguay J, Cabanero M, et al. Prevalence and Heterogeneity of PD-L1 Expression by 22C3 Assay in Routine Population-Based and Reflexive Clinical Testing in Lung Cancer. J Thorac Oncol. 2021 Sep;16(9):1490-1500. doi: 10.1016/j.jtho.2021.03.028. Epub 2021 Apr 26. PMID: 33915250.
12. McKenna N, Li T, Cheng L, Hosseini MJ, Johnson M, Steedman M. Sources of hallucination by Large Language Models on inference tasks. In The 17th Conference of the European Chapter of the Association for Computational Linguistics: Findings of EACL 2023. Stroudsburg, PA: Association for Computational Linguistics (ACL). 2023. p. 2758-2774 doi: 10.18653/v1/2023.findings-emnlp.182

Journal Pre-proof

Journal Pre-proof

Figure 1. Percentage comparison between the synthetic cohort and real-world data.



Journal Pre-proof

Journal Pre-proof

Supplementary Annex 1. Structured examples of virtual patients generated using ChatGPT-4.0 through iterative prompting based on the Role–Task–Format (RTF) scheme. These examples include data fields relevant to surgical decision-making, including tumor location, imaging findings, pulmonary function, and resectability assessment, making them suitable for training scenarios in thoracic oncology.

Patient 1

1. Identification
    - Patient ID: VTC001
    - Age: 68 years
    - Sex: Male
2. Past Medical History
    - Prior malignancy: None
    - Smoking history: 45 pack-years; former smoker (quit 4 years ago)
    - Occupational exposure: 15 years of silica in mining
    - Comorbidities: Hypertension; mild COPD; Type 2 diabetes mellitus
3. Clinical Presentation
    - Symptoms: Persistent cough, mild dyspnea, 4 kg weight loss over 2 months
    - Duration: 5 months
4. Imaging Studies
    - Chest X-ray: Nodular mass in left upper lobe
    - CT chest/abdomen: 3.5 cm peripheral lesion in left upper lobe, abutting main bronchus without overt invasion
    - PET-CT: SUV_{max} 10.2 in primary nodule; mediastinal lymph node (4R) SUV_{max} 4.8
    - MRI: Not performed
5. Diagnostic Procedures
    - Bronchoscopy with biopsy: Performed; sample obtained
    - Transthoracic needle biopsy: Performed; confirmed malignancy
    - EBUS: Positive nodal involvement in stations 4R and 7
    - Thoracentesis: Not applicable
6. Histopathology
    - Histologic subtype: Pulmonary adenocarcinoma
    - Grade: Moderately differentiated
    - IHC: TTF-1 positive; Napsin A positive
7. TNM Staging (8th ed.)
    - T: T2a (>3 cm but ≤5 cm)
    - N: N2 (mediastinal lymph node metastases)
    - M: M0 (no distant metastases)
    - Overall stage: IIIA
8. Molecular and Biomarker Testing
    - EGFR: Negative
    - ALK: Negative
    - ROS1: Negative
    - KRAS G12C: Positive
    - PD-L1 TPS: 20%

Journal Pre-proof

9. Functional Assessment
    - Pulmonary function: FEV_1 66% predicted; FVC 70% predicted; DLCO 58% predicted
    - Six-minute walk: 380 m (SpO_2 96 → 94%)
    - Cardiac evaluation: Unremarkable
    - Treatment fitness: Candidate for chemotherapy
10. Psychosocial Factors & Patient Preferences
    - Emotional status: Moderate anxiety
    - Social support: Adequate family support
    - Preference: Favors curative-intent therapies
11. Prior Therapies & Response
    - None
12. Treatment Toxicity & Tolerance
    - Not applicable
13. Life Expectancy & Performance Status
    - Karnofsky Performance Status: 80%
    - ECOG: 1
14. Frailty/Vulnerability Index
    - Charlson Comorbidity Index: 4

Patient 11

1. Identification
    - Patient ID: VTC011
    - Age: 74 years
    - Sex: Male
2. Past Medical History
    - Prior cancer: Prostate cancer treated 10 years ago, currently in remission
    - Smoking history: 50 pack-years; former smoker (quit 3 years ago)
    - Occupational exposure: 20 years of asbestos in construction
    - Comorbidities: Hypertension; chronic renal insufficiency; COPD
3. Clinical Presentation
    - Symptoms: Persistent cough, progressive dyspnea, 6 kg weight loss over 4 months
    - Duration: 5 months
4. Imaging Studies
    - Chest X-ray: Mass in right upper lobe
    - CT chest/abdomen: 4.3 cm peripheral lesion right upper lobe, abutting main bronchus without invasion
    - PET-CT: SUV_{max} 11.2 in primary nodule; mediastinal lymph nodes SUV_{max} 5.1
    - MRI: Not performed
5. Diagnostic Procedures
    - Bronchoscopy with biopsy: Performed; sample obtained
    - Transthoracic needle biopsy: Performed; confirmed malignancy
    - EBUS: Not performed
    - Thoracentesis: Not applicable
    - Mediastinoscopy: Not performed
6. Histopathology
    - Histologic subtype: Squamous cell carcinoma

Journal Pre-proof

- Grade: Moderately differentiated
- IHC: p40 positive; CK5/6 positive

7. TNM Staging (8th ed.)
   - T: T2b (>5 cm but ≤7 cm)
   - N: N1 (ipsilateral hilar lymph node metastases)
   - M: M0
   - Overall stage: IIB
8. Molecular and Biomarker Testing
   - EGFR: Negative
   - ALK: Negative
   - KRAS: Negative
   - PD-L1 TPS: 40%
9. Functional Assessment
   - Pulmonary function: FEV_1 55% predicted; FVC 65% predicted; DLCO 60% predicted
   - Six-minute walk: 280 m (SpO_2 94 → 90%)
   - Cardiac evaluation: Stable hypertension; chronic renal disease
   - Treatment fitness: Candidate for surgery with precautions
10. Psychosocial Factors & Patient Preferences
    - Emotional status: Moderate anxiety
    - Social support: Limited family support
    - Preference: Prefers curative-intent therapy if feasible
11. Prior Therapies & Response
    - Prior radiotherapy for prostate cancer; complete remission
12. Treatment Toxicity & Tolerance
    - Mild post-radiation fatigue
13. Life Expectancy & Performance Status
    - Karnofsky: 70%
    - ECOG: 1
14. Frailty/Vulnerability Index
    - Charlson Comorbidity Index: 5

Patient 28

1. Identification
   - Patient ID: VTC028
   - Age: 62 years
   - Sex: Female
2. Past Medical History
   - Prior malignancy: None
   - Smoking history: 15 pack-years; former smoker (quit 12 years ago)
   - Occupational exposure: None relevant
   - Comorbidities: Hypertension; Type 2 diabetes mellitus
3. Clinical Presentation
   - Symptoms: Persistent dry cough, mild dyspnea, 3 kg weight loss over 3 months
   - Duration: 4 months
4. Imaging Studies
   - Chest X-ray: Lesion in right middle lobe

Journal Pre-proof

- CT chest/abdomen: 2.8 cm peripheral lesion right middle lobe, no critical-structure invasion
- PET-CT: SUV_{max} 7.9 in primary nodule; no other uptake
- MRI: Not performed

5. Diagnostic Procedures
   - Bronchoscopy: Not performed
   - Transthoracic needle biopsy: Performed; confirmed malignancy
   - EBUS: Not applicable
   - Thoracentesis: Not applicable
   - Mediastinoscopy: Not performed
6. Histopathology
   - Histologic subtype: Pulmonary adenocarcinoma
   - Grade: Well differentiated
   - IHC: TTF-1 positive; Napsin A positive
7. TNM Staging (8th ed.)
   - T: T1c (>2 cm but ≤3 cm)
   - N: N0
   - M: M0
   - Overall stage: IA
8. Molecular and Biomarker Testing
   - EGFR: Positive
   - ALK: Negative
   - KRAS: Negative
   - PD-L1 TPS: 15%
9. Functional Assessment
   - Pulmonary function: FEV_1 80% predicted; FVC 82% predicted; DLCO 75% predicted
   - Six-minute walk: 350 m (SpO_2 97 → 95%)
   - Cardiac evaluation: Controlled hypertension; otherwise unremarkable
   - Treatment fitness: Candidate for surgery
10. Psychosocial Factors & Patient Preferences
    - Emotional status: Stable
    - Social support: Good family support
    - Preference: Prefers less invasive options if effective
11. Prior Therapies & Response
    - None
12. Treatment Toxicity & Tolerance
    - Not applicable
13. Life Expectancy & Performance Status
    - Karnofsky: 85%
    - ECOG: 1
14. Frailty/Vulnerability Index
    - Charlson Comorbidity Index: 2

Journal Pre-proof