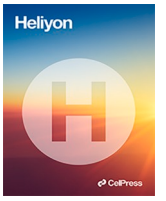
Heliyon 10 (2024) e40147

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

# Treatment patterns and outcomes in patients with non-small cell lung cancer

Florent F. Richy [a,*], Deepa Kumar [b], Pierre Freres [a]

[a] CHU Liege, Belgium
[b] ICON plc global, Greater Philadelphia, USA

ARTICLE INFO

*Keywords:*
Non-small cell lung cancer
NSCLC
Real world evidence
Comparative effectiveness

ABSTRACT

Over the last decade, innovative treatments have enlarged the therapeutic arsenal against NSCLC. Although there is clinical evidence supporting treatment options, controversy surrounds the impact of their combinations and sequence. The goal of this study was to fill this gap using real world data (Symphony Health).

9676 Patients with a diagnosis of non-metastatic (NM) and metastatic (M) NSCLC between July 2016 and December 2016 were included and followed till December 2021. The following data were collected: diagnostic date, age, gender, comorbidities, metastases, treatments (SUrgery, RAdiotherapy, CHemotherapy, IMmunotherapy, and TArgeted therapies), sequence (L1-L4), and death. Treatment patterns (Sankey graphs), and outcomes (Kaplan-Meyer and Cox regressions) were derived.

Amongst NM patients, the most frequently observed treatments were RA, IM, IM and CH in first, second, third and fourth lines respectively. In M patients, the most frequent treatments were CH, IM, IM, and IM in first, second, third and fourth lines respectively. The overall 5-years survival in the cohort was 23.7 % (36 % in NM, 16 % in M). The corresponding median survival time were 1.98 years (3.2 NM, 1.27 M). In NM, best median survival times were recorded for those with only one line SU. For patients who had 2 lines, CH→IM; 3 lines, CH→IM→CH. In M, best median survival times were recorded for those with only one line RA. For patients who had 2 lines CH→IM; 3 lines CH→IM→CH.

These new insights could guide further studies and improve the management of patients with NSCLC.

## 1. Introduction

Lung cancer is the predominant form of neoplasms found in men, featuring an annual incidence of 1.4 million new cases in 2020. In women, its incidence was half that in men (0,7 million cases), following breast (2,2 million cases) and colorectal cancers (0,8 million) worldwide. The approximate number of diagnosed cases was estimated at 2,206,771 (11.4 % of all cancers), while mortality was 1,796,144 (18.0 %) [1]. The mortality is mainly correlated with staging at diagnostic. In practice more than 65 % of lung cancers present with stage 3 or metastases at diagnosis [2]. Non-small cell lung cancer is the most common form of lung cancer, accounting for 85 % of all cases. Three main histological types were defined by the WHO in 2015: adenocarcinoma, squamous cell carcinoma, and

* Corresponding author. University of Liege, Belgium.
*E-mail address:* florent.richy@uliege.be (F.F. Richy).

https://doi.org/10.1016/j.heliyon.2024.e40147
Received 7 July 2023; Received in revised form 4 November 2024; Accepted 4 November 2024
Available online 5 November 2024

large cell carcinoma [3].

The European Society for medical Oncology (ESMO), the American Society for Clinical Oncology and the American Cancer Society, American Association for Cancer Research provide regular updates on their evidence-based recommendations for the treatment of early, locally advanced and advanced NSCLC [4–6]. For stage and II local NSCLC, lobectomy or sleeve/segmentectomy/wedge resection are the standards for eligible patients. For non-eligible patients, radiotherapy is used. For patients at risk of recurrence, systemic adjuvant treatment (chemotherapy, immunotherapy, targeted therapies) can be proposed. For locally advanced patients (stages IIIA and B), radiotherapy is the treatment of choice while surgery is rarely an option. Immunotherapy is often used to lower the risk of relapse. For patients who have stage IV NSCLC, the systemic treatment includes chemotherapy, immunotherapy or targeted therapies, alone or in combination.

These guidelines are developed by multidisciplinary expert panels performing systematic reviews of the evidence through online searches of PubMed, Embase and the Cochrane Library. Although the methodological robustness of the included evidence and the expertise of oncologists involved provide a solid ground, these guidelines suffer intrinsic limitations. First they use aggregated, methodologically different studies to derive averaged estimates. Second, due to the increasing number of treatment options and possible lines of treatment, an overview of all comparative efficacies is almost impossible to establish based on clinical data. Even the most sophisticated network meta-analysis could not control for all biases related to diseases, patients, settings and drugs differences. Finally, the quality of systematic reviews generated and used to create oncology guidelines has been challenged [7]. Indeed, Bierenbaum and colleagues showed that adherence to active cancer treatment clinical practice guidelines recommendations by oncology clinicians is negatively influenced by concern over guidelines content and their last updates, concern about the considered evidence, clinician uncertainty, and organizational and patient factors [8]. These three factors trigger the need for new approaches in the assessment of the comparative effectiveness of therapeutic options in NSCLC. This study was a pilot developed to assess NSCLC treatment patterns and compare their effectiveness's in an ecological fashion using a large-scale claims database.

## 2. Methods

The database used was Symphony Health [9]. This anonymized claims database contains data from 317 million active patients collected by 2.9 million prescribers in the US. It accounts for more than 84 % of the prescriptions dispensed in the US & territories, features a 5,3 years average continuous history for active patients and 17 years of retrospective data.

Patients aged 18 or more who had an initial diagnosis for NSCLC between January 2016 and June 2016, with no previous Rx for chemotherapy, radiotherapy, immunotherapy, surgery or targeted treatment for cancer were included. The ICD-10 code C34.9 was used. The date the code appeared was considered index date. As there is no specific ICD-10 code for NSCLC, patients receiving etoposide, topotecan or irinotecan were not selected as these therapies are specific to SCLC [10]. NSCLC were categorized as non-metastatic at baseline if they did not receive a C78 code within the first 12 months' post index date, otherwise they were considered progressing. Conversely patients who had a C78 code within 12 months' post index date were considered metastatic at baseline. If a patients received a C78 code after 12 months of diagnosis, he was considered as progressing.

Treatments were categorized as: surgery, radiotherapy, chemotherapy (pemetrexed, paclitaxel,nab-paclitaxel, docetaxel, Cisplatin, Carboplatin, Gemcitabine, Vinorelbine), immunotherapy (nivolumab, pembrolizumab, atezolizumab, avelumab, ipilimumab) and targeted therapies (gefitinib, erlotinib, afatinib, dacomitinib, osimertinib, icotinib, crizotinib, alectinib, brigatinib, lorlatinib, ceritinib, dabrafenib, trametinib, vemurafenib, encorafenib). The following abbreviations were used to code therapies in the analysis datacut: SU for surgery, RA for radiotherapy, CH for chemotherapy, IM for immunotherapy and TA for targeted therapy. Up to four lines of treatment were considered. Outcomes were death (no claims for more than 2 months) and progression to metastatic (C78 code).

Covariates included gender, age, comorbidities (as assessed by Charlson's comorbidity index) [11].

Treatment patterns and their dynamics were modelled using Sankey graphs, a standard model used in science and engineering to represent dynamic flows. Since the 1990s this visual model has been used in life-cycle assessment of products. Sankey diagrams enable to represent several steps along with the information about flows' volumes, and to manage the complexity of the patient paths [12]. We generated separate Sankey diagrams for patients with and without metastases at diagnosis as these are key drivers for treatment patterns and outcomes. In order to assess the complexity of the obtained networks, qualitative analysis was performed based on the treatment dynamics by treatment line while a quantitative approach applied fractal dimension analysis. The fractal dimension is an estimate of a 2D object complexity [13], which in this case is of interest to compare the Sankey graphs generated. Treatment efficiency was assessed by comparison of the median survival time of patients at different lines. To this end, Kaplan-Meyer estimators and their respective median and survival estimates were generated for each treatment pattern at each of the 4 lines of treatment. The analysis was carried out within each of the treatment lines in order to avoid treatment sequence interactions. Accordingly, individual patient exposure was calculated for each corresponding treatment line. In order to limit indication-led bias and the impact of confounding factors, proportional hazards Cox models were also computed, accounting for age, gender and comorbidities including severe obesity. Pairwise comparisons of each treatment option at each line were performed using Tarone-Ware test [14]. Overall survival was estimated at the end of the 5 years observation period accounting for treatment and their sequences using a concatenated analysis. Treatments or treatment sequences administered to less than 20 patients were not considered due to lack of statistical power. Data extraction was performed and quality-checked by a certified programmer (DK) using Hadoop Impala SQL ® while data analysis and interpretation were performed by a trained pharmacoepidemiologist (FR) together with a board-certified PhD oncologist (PF) using Statistica software ®. QC was code-review for data extraction and parallel programming for data analysis.

## 3. Results

A total of 9676 patients with a diagnosis of NSCLC between June and December 2016 were included in the analysis. Table 1 describes their demographics. 49 % were females, the mean (SD) age of the sample was 67,8 (9,6) years old. 3763 (38,4 %) were non-metastatic at treatment initiation while 6033 (61,6) were metastatic. The average Charlson's comorbidity score was 5 (2,6) and 6 % had severe obesity. Non-metastatic patients were significantly younger (67 vs 69 years, $p < 0,01$), had a higher Charlson's comorbidity score (5,9 vs 3,5, $p < 0,01$, when adjusting for the weight of metastatic disease in the score) and had slightly more treatment lines (1,7 vs 1,4, $p < 0,01$) (Table 1).

Amongst non-metastatic patients, 76,2 % had only one line of treatment, while 12,8 %, 6,9 % and 4,1 % 2, 3 and 4 lines, respectively. Amongst metastatic patients, 60,9 % had only one line of treatment, while 20,1 %, 11 % and 8 % 2, 3 and 4 lines, respectively. The later group had a significantly higher number of treatment lines (1,7 vs 1,4, $p < 0,01$) (Tables 2 and 3).

Amongst non-metastatic patients, radiotherapy was the most frequently observed treatment in first line (32 %), followed by chemotherapy (26 %), surgery (11 %), immunotherapy (10 %) and a combination of chemotherapy and radiotherapy (9 %). In second line, immunotherapy was the predominant treatment (30 %), followed by radiotherapy (24 %), chemotherapy (20 %), targeted therapy (7 %), and chemo-radiotherapy (6 %). In third line, immunotherapy (32 %) was the most prevalent, followed by chemotherapy (24 %), radiotherapy (15 %), targeted therapy (7 %), and chemo-immunotherapy (6 %). Finally and for the 4 % of patients having a fourth line of treatment, chemotherapy was the practitioners' preferred choice (27 %), followed by immunotherapy (23 %), radiotherapy (16 %), chemotherapy + targeted therapy (8 %) and targeted therapy alone (8 %). Of note, adjuvant and neoadjuvant therapies (CHRA, CHTA, CHIM, CHIMRA, CHIMTA) were rare and their complexity was inversely correlated to their frequencies. In metastatic patients, chemotherapy was the most frequent first line of treatment (30 %), followed by radiotherapy (26 %), chemo-radiotherapy (12 %), and chemotherapy + targeted therapy (7 %). In second line, immunotherapy was the dominant treatment (33 %), followed by chemotherapy (21 %), radiotherapy (17 %), targeted therapy (6 %) and chemotherapy + targeted therapy (5 %). In third line, immunotherapy (29 %) ranked first, followed by chemotherapy (22 %), radiotherapy (19 %), immunotherapy + radiotherapy (7 %), and chemotherapy + targeted therapy (6 %). For the 8 % receiving a fourth line of therapy, immunotherapy (28 %) was followed by chemotherapy (26 %), radiotherapy (15 %), immunotherapy + radiotherapy (7 %), and finally chemotherapy + radiotherapy (6 %). Adjuvant and neoadjuvant therapy schemes were more frequent amongst those patients: chemo + radiotherapy accounted for 16 % of the total across lines, chemotherapy + targeted therapy 10 %. More complex therapeutic schemes (CHIMRA, CHIMTA, CHTARA) were anecdotal (Tables 2 and 3).

The Sankey Diagram gave us a detailed, high-level view of how treatment sequences flows and changes from one line to the other in the two groups of patients. Tracking these revealed many critical insights (Graph 1). In non-metastatic patients receiving RA, the vast majority had no second line. For those who had, a second RA was also the main transition. In those receiving CH, a more important fraction had a second line equally split in between another CH and IM. A minority of patients who had surgery also received a second line RA or SU. The same situation was seen for patients who received IM with a subsequent IM mainly. Interestingly, 25 % of the patients who received CHRA first line were prescribed IM or combined treatment with IM. When analyzing the dynamics between the second and third line, the vast majority of IM, RA and CH patients had no third line while the remaining were almost evenly directed towards IM, CH and RA. Only one third of patients with a third line moved to fourth line, mostly non-combined CH, IM or TA.

In metastatic patients (Graph 2) receiving CH, RA or IM One third to one half of them had no second line. For those who had RA and a second line, the two main pathways were RA, followed by CH and IM. For patients treated with CH, the majority of the second line was IM, followed by CH and RA. From these, half of them had no 3rd line. RA-treated patients in line 2 were directed towards RA 3rd line, patients with IM 2nd line were mainly prescribed IM or CH and those who had CH received IM, followed by CH and RA. The transition between line 3 and 4 for those who had revealed similar patterns. Of note, at each lien transition, a minority of patients received other treatment, namely CHTA, TA, CHRA, IMRA, CHIMRA and CHIMTA by decreasing order.

The two Sankey networks were very different in density. In order to evaluate and compare their respective complexities, the fractal analysis confirmed that the non-metastatic patients network had a lower fractal dimension than the metastatic patients network: 1.42 against 1.56, respectively. That translates the fact that patients with metastases had more heterogeneous treatment patterns by line.

The overall 5-years survival rate was 23.7 % (35,9 % in non-metastatic, 16,3 % in metastatic) (Graphs 3 and 4). The median survival time was 1.98 years (3.2 in non-metastatic, 1,27 in metastatic patients). Survival was negatively impacted by increasing Charlson's comorbidity index, the presence of metastases at index date and, of course, treatment received.

**Table 1**
Demographics.

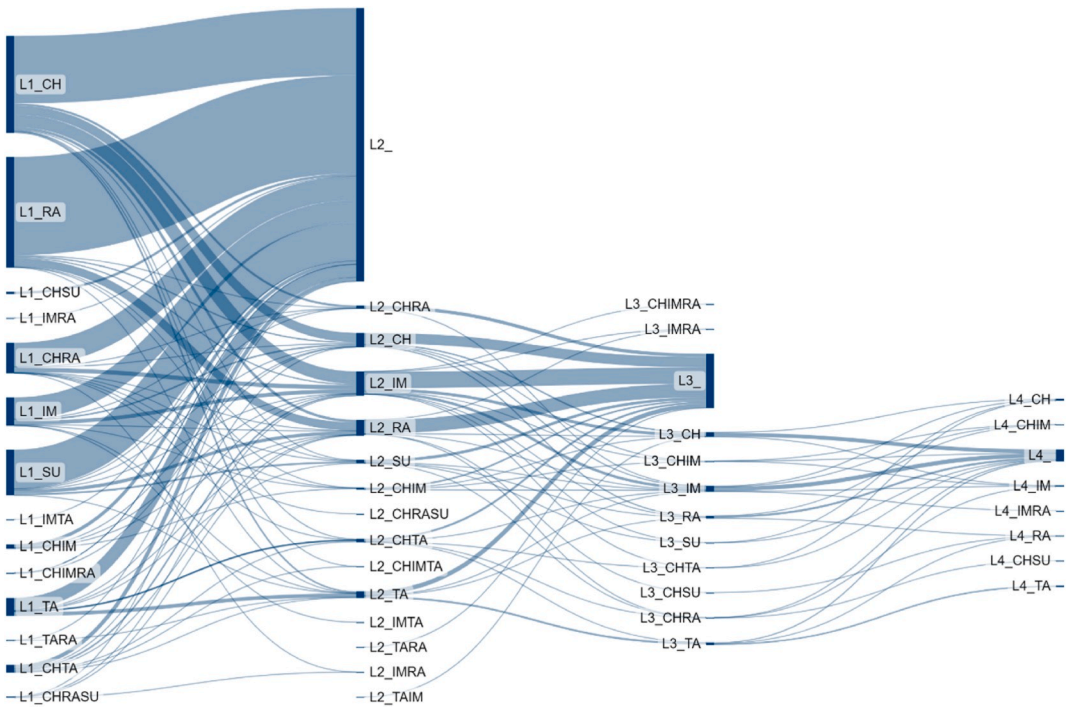
| All patients n = 9676 | | | |
|---|---|---|---|
| Age | 67,8 (9,7) | | |
| Gender (female) | 49 % | | |
| Severe obesity | 6 % | | |
| Charlson's | 5 (2,6) | | |
| | Non-metastatic patients n = 3642 | Metastatic patients n = 6033 | p difference |
| Age | 69 (8,9) | 67 (8,9) | <0,01 |
| Gender (female) | 50 % | 48,2 % | 0,05 |
| Severe obesity | 6 % | 6 % | NS |
| Charlson's | 3,5 (1,7) | 5,9 (2,6) | <0,01 |

**Table 2**
Overall survival by lines of treatment.

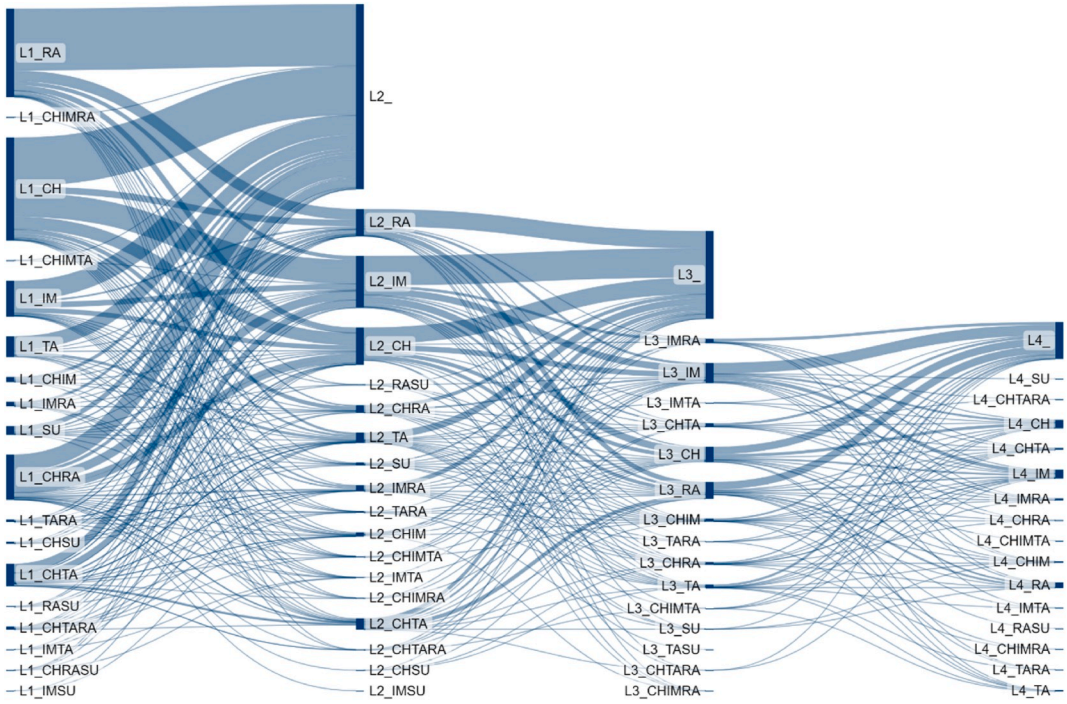
| | Lines of treatment | Deceased % |
|---|---|---|
| Non-metastatic patients | 1 | 62 |
| | 2 | 73 |
| | 3 | 76 |
| | 4 | 72 |
| Metastatic patients | 1 | 85 |
| | 2 | 85 |
| | 3 | 81 |
| | 4 | 73 |

**Table 3**
median survivals in the concatenated analysis by treatment pattern, N > 50, by ascending median survival.

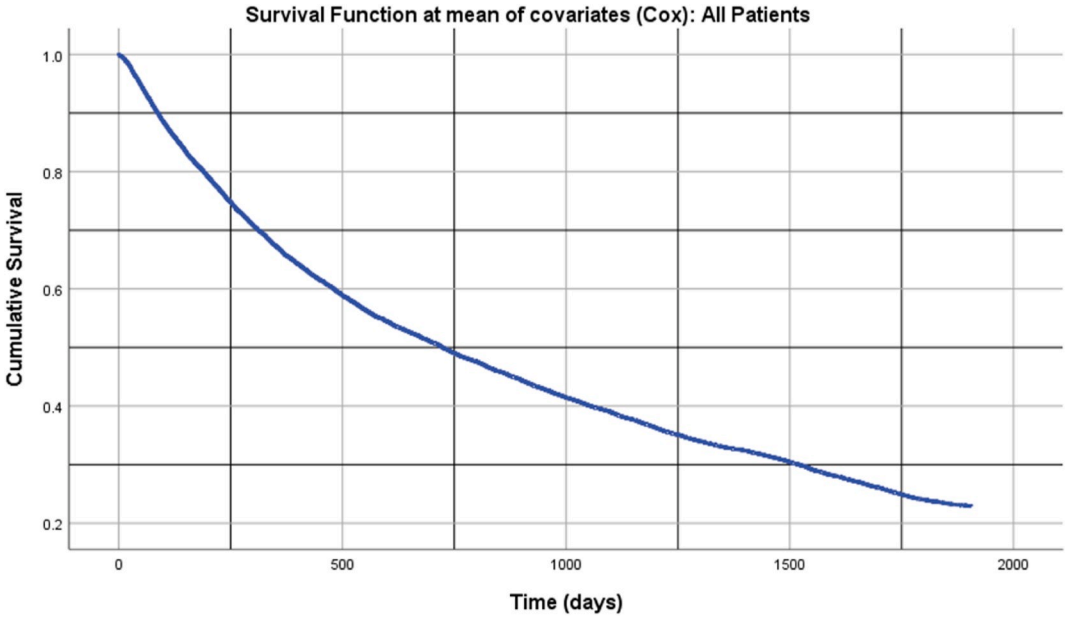
| Non-metastatic patients | | | | Metastatic patients | | |
|---|---|---|---|---|---|---|
| | Treatment sequence | % | Median survival (days) | Treatment sequence | % | Median survival (days) |
| 1 Line of treatment | I | 8 | 325 | C | 20 | 207 |
| | C | 23 | 422 | CT | 4 | 213 |
| | T | 4 | 515 | R | 26 | 236 |
| | CR | 9 | 528 | CR | 7 | 258 |
| | R | 34 | 700 | I | 8 | 407 |
| | S | 14 | 1794 | CI | 1 | 416 |
| | | | | T | 6 | 544 |
| | | | | S | 2 | 1156 |
| 2 Lines of treatment | C→I | 3 | 337 | I→R | 1 | 193 |
| | C→C | 2 | 744 | CT→I | 2 | 358 |
| | R→R | 2 | 1353 | C→I | 6 | 374 |
| | | | | CR→I | 2 | 384 |
| | | | | I→I | 8 | 407 |
| | | | | C→C | 3 | 474 |
| | | | | R→R | 3 | 588 |
| | | | | C→R | 2 | 856 |




**Graph 1.** Sankey graph for non-metastatic patients.




**Graph 2.** Sankey graph for metastatic patients.

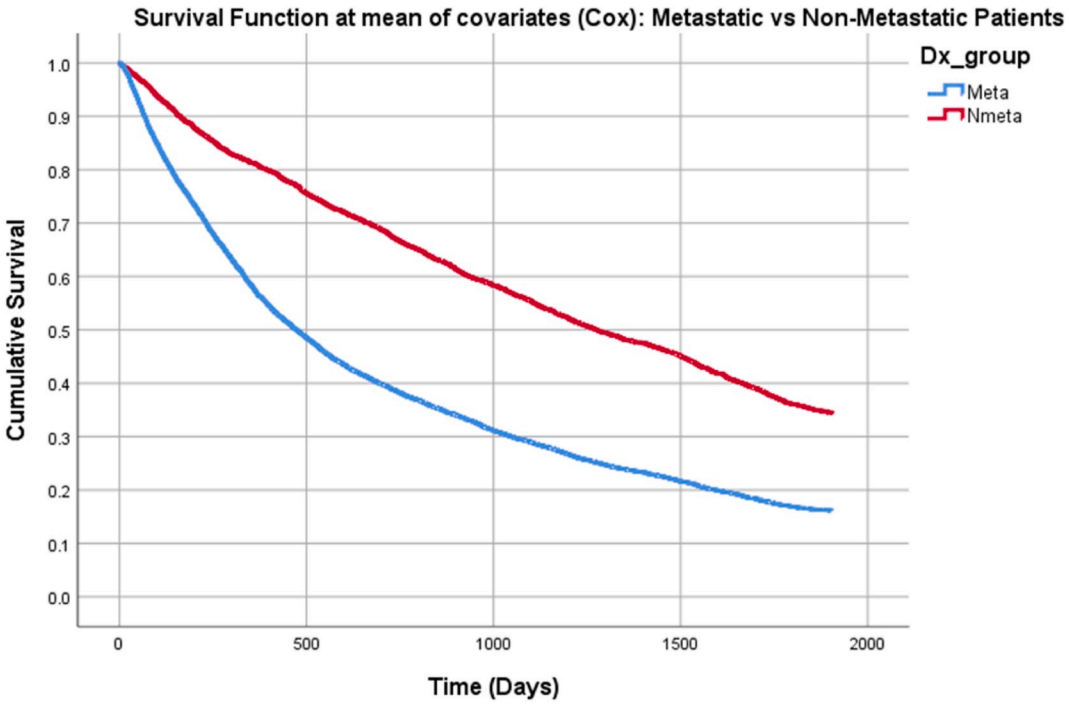



**Graph 3.** Global survival.

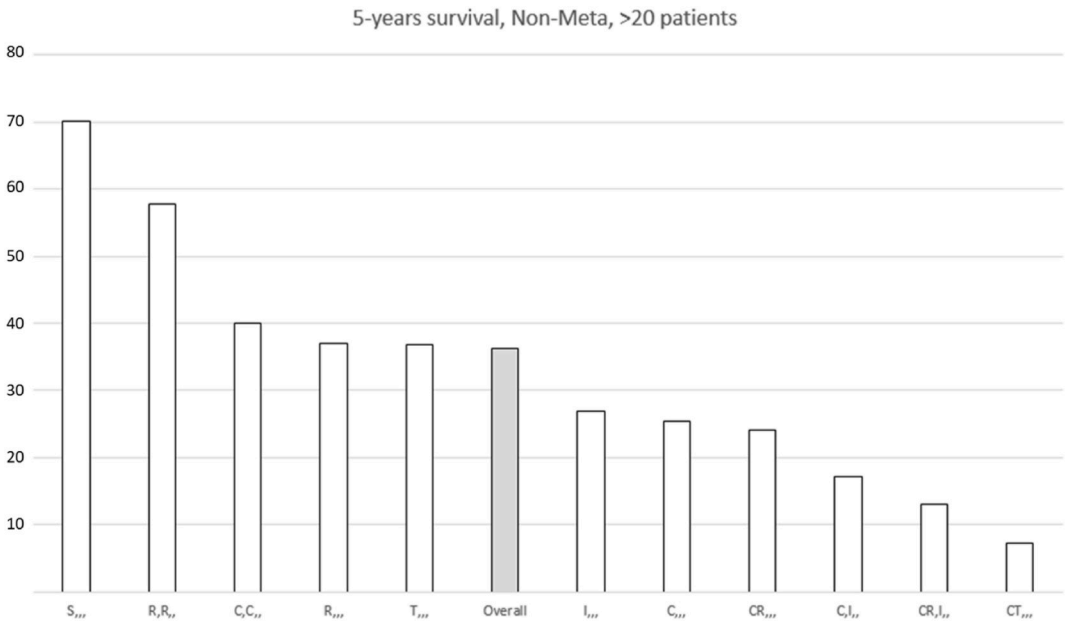
The total exposures in non-metastatic patients were 6778.45, 681.19, 178.08 and 82,29 patient-years in lines 1, 2, 3 and 4. Respective mortalities was 60 %, 48 %, 47 % and 63 % (Graph 5).

For non-metastatic patients having one line of treatment, the top 3 most frequent treatments and their related median survival and OS at 5 years were: SU(-, 69 %), RA (950, 31 %), TA (1300, 38 %). For those who had 2 lines: CH,IM (325, 18 %), CH, CH (850, 40 %), CH + RA,IM (660, 16 %). For those who had 3 lines: CH,IM,CH (650, 0), CH, IM, IM (900, 0), TA,TA,TA (1450, 40).For those with 4 lines, too few patients were present to compute frequencies and survival.

The total exposures in metastatic patients were 7195.17, 3075.39, 1154.82 and 856.42 patient-years in lines 1, 2, 3 and 4. The total



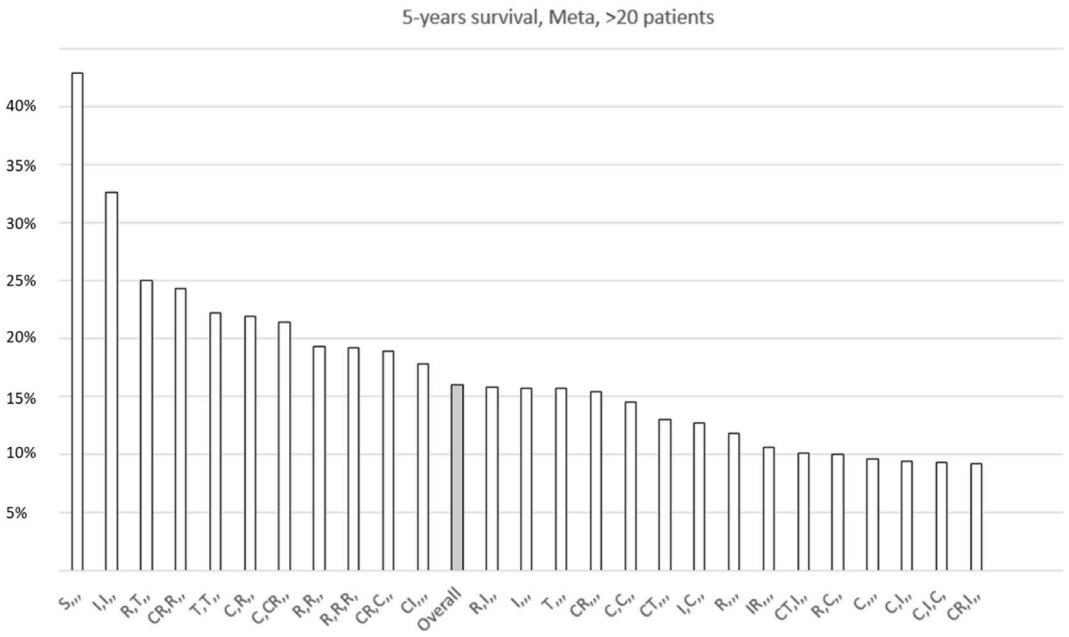

**Graph 4.** Survival: metastatic vs non-metastatic patients.




**Graph 5.** 5-years survival, non-metastatic patients, >20 patients per pattern.

mortality over the 5 years follow-up period was 83.7 %.

For metastatic patients having one line of treatment the top 3 most frequent treatments and their related median survival and OS at 5 years were: RA (300, 17 %), CH (225, 16 %), and I (400, 17 %). For patients who had 2 lines: CH, IM (350, 9 %); CH,CH (450, 16 %); CH + RA, IM (350, 10 %). For patients who received three lines of therapy: CH, IM, CH (550,10 %); CH, CH, IM (500, 10 %); CH, IM, IM (750, 11 %). For those who had 4 lines of therapy, there were too few patients and too many sequences to compute frequencies and survival (see Graph 6).




**Graph 6.** 5-years survival, metastatic patients, >20 patients per pattern.

## 4. Discussion

Real world data are becoming increasingly important in the evaluation of modern therapeutics. They allow for a pragmatic assessment of treatment patterns found in a variety of conditions and are particularly insightful in oncology where multiple lines of different treatments can be offered to the patient. Whereas meta-analyses of RCT remain the gold standard in terms of evidence, observational data is a precious guide in generating new hypotheses and confronting experimental against pragmatic data.

NSCLC is of obscure prognosis, even for early stages (50 % of the patients will have a relapse within 5 years) [15]. Most recurrences occur during that time period, as only 2.5 % of patients present recurrence after 10 years [16]. Local treatments (radiotherapy, surgery) are the treatments of choice, although few individuals qualify for them, especially in patients with metastases. While platinum-based chemotherapy regimens have long been the standard treatment option for patients with advanced NSCLC, the past 2 decades have witnessed a profound transformation of the treatment paradigm for patients with lung cancer. There is a controversy surrounding the therapeutic benefit of specific treatment sequences over treatment combinations [17]. Carroll and colleagues have analyzed the treatment patterns and outcomes of NSCLC in patients before and after (2016) the introduction of immuno-oncologic therapies. Their conclusion was that despite the increase in median survival, there was a continuing need for improved therapeutic approaches based on innovative combinations. Our data, spanning from 2016 to 2021 corroborate these findings. Despite a lack of statistical power for some subgroups, metastatic patients who received a combination of chemotherapy and immunotherapy on first line had longer median survival than patients who received immunotherapy alone. Paradoxically, a more than two-fold difference was observed between patients who had immuno followed by radiotherapy as compared to those who received chemo followed by radiotherapy. Also, combined approaches with radiotherapies as last treatment more than probably reflect a limitation of our methodology to discriminate between combined (less than 1 month between claims) and sequential (more than one month) therapies. Our analysis could also suggest that patients with more treatment lines had better outcomes. This is an artefact linked to the confounding effects of age, gender, and comorbidities. When adjusting for these covariates and using a Cox proportional hazards model, the 5 years median survival rates were the highest for patients with one line of treatment while equivalent for patients with 2–4 lines for non-metastatic patients. Of interest, the situation was opposite for metastatic patients as median survival was clearly correlated to the number of lines of therapy. A possible confounding effect could be that Charslon's comorbidity index (CCI) was defined at index date. As CCI fluctuates over time the impact of this assumption is that patients with better health status have access to more lines of therapies, especially in those with metastases.

Our work has shown that the most frequent treatments in non-metastatic patients by lines of patients were RA, IM, IM and CH. When contrasting these frequencies to median term survival, only RA in first line was coherent. The sequences offering the highest outcomes observed were RA, CH, IM and IM. Within metastatic patients there was no correlation between the frequency of treatments (RA, TA, IMRA, TA) and the observed efficacy in lines 1 to 4. The combinations and sequences leading to the highest outcomes were: CH + IM, CH + RA, CH, CH + IM. Although being purely observational and not comparative, these findings outline the need for additional clinical and pragmatic research to optimize treatment at a patient-level. For instance, Simeone and colleagues have shown that carboplatin plus paclitaxel was the most common first-line therapy (18.6 %), and nivolumab was the most common second- (31.0 %) and third-line (38.4 %) therapy; 26.7 % of all patients being untreated. Median OS from initial metastatic diagnosis was 11.1

months (95 % CI: 10.8–11.5). Second-line immunotherapy extended OS by over 3 months versus second-line chemotherapy [18]. These data are very comparable to our findings in metastatic patients, except that Simeone and colleagues did not account for radiotherapy which is the most common first step in the therapeutic engagement. The median OS of the sequence chemotherapy first line followed by immunotherapy was almost identical in the two studies: 337 (+3 months) versus 374 (+4 months) days in Simeone et al. and our study, respectively. This pattern was not found in our sample of non-metastatic patients: 422 days for 1 line chemotherapy versus 337 days for patients with chemotherapy followed by immunotherapy. It goes without saying that baseline characteristics of these patients explain this apparently paradoxical situation, as demonstrated by comparing the baseline Charlson comorbidity scales: chemotherapy: 3.27 (1.63) and chemotherapy followed by immunotherapy 5.47 (2.55), $p < 0.05$.

Our study has limitations. Our Symphony Health analysis datacut is built on the basis of claims data. Although very specific, this coding system does not allow to account for medical records data, notably in terms of staging at baseline. We had to make the assumption that non-metastatic patients were stage I-IIIA without a code for metastatic disease during the first 12 months of follow-up. Although this reflects clinical practice there might be a few patients miscoded due to the intervals between oncology visits. Similarly, comorbidities and their severity are under-reported due to the nature of the database. For the sake of clarity and due to the very high volume of data we did not include patient who developed metastasis during the observation period. Indeed, progression is itself a confounder for treatment patterns and survival, and including those patients would have diluted the conclusions, as also noted by Simeone and colleagues[14]. A dedicated research on this subgroup of patients is ongoing. Additionally, the analyses of OS by treatment sequence, and OS from the initial NSCLC diagnosis date, included immortal time bias for some patients that could be a confounder in the classification between non-metastatic patients and metastatic at baseline.

## 5. Conclusion

This study described the treatment patterns of US patients with metastatic or non-metastatic NSCLC. It revealed that in majority, the frequency of specific treatments was associated to the highest median survivals for non-metastatic patients. However, in metastatic patients these two variables were uncorrelated. On top of individual treatment sequence, an important number of treatment combinations, irrelevant to the treatment lines were also observed in this latter group of patients.

This research outlines the need for interfacing patient, real-world and clinical data to maximize the treatment effectiveness down to the individual person who suffers NSCLC. The authors wish to pay due respect to patients who contribute building tomorrow's evidence by sharing their data.

### Ethical aspects

The present observational research has been conducted on the basis of retrospective, pre-existing, anonymized, claims data collected in the context of usual medical practice. Identification of the patients was impossible as source data did not contain any sensitive related variable. The protocol of the study and the present publication have been reviewed by the Ethics Committee of the University of Liege Academic Hospital (Belgium) (File 2024/330, see next page) who issued a positive advice provided that the study respected the local regulations on human person experimentation (May 7, 2004 law), GDPR and HIPAA.

### Adverse events reporting

This study was retrospective, non-imposed and focused solely on treatment patterns and efficacy. It did not include any adverse event and thus did not qualify as a PASS. Accordingly, adverse event reporting was neither required nor possible by regulations and design, respectively.

### Accordingly

- The data associated with this study has not been deposited in a publicly available repository.
- The authors do not have permission to share source data.

### CRediT authorship contribution statement

**Florent F. Richy:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Conceptualization. **Deepa Kumar:** Writing – review & editing, Formal analysis, Data curation. **Pierre Freres:** Writing – original draft, Investigation.

### Data availability statement

This study was not an agreement with a private research solution company. It was part of an internal research and development program conducted by ICON plc.

The datacut used to perform this study is a subset of the commercial Symphony Health database, an ICON plc company.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

FR, DK and PF have reviewed and agreed on the final version of the manuscript. FR and DK were involved in the study design, analyses and interpretation of the data and writing of the manuscript while PF was involved in the interpretation of the data and writing of the manuscript.

## References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (2021) 209–249.
[2] R. Meza, C. Meernik, J. Jeon, et al., Lung cancer incidence trends by gender, race and histology in the United States, 1973-2010, PLoS One 10 (2015) e0121323.
[3] W.D. Travis, E. Brambilla, A.G. Nicholson, et al., The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification, J. Thorac. Oncol. 10 (2015) 1243–1260.
[4] www.esmo.org.
[5] www.asco.org.
[6] www.cancer.org.
[7] F. Vigna-Taglianti, P. Vineis, A. Liberati, et al., Quality of systematic reviews used in guidelines for oncology practice, Ann. Oncol. 17 (2002) 691–701.
[8] M. Bierbaum, F. Rapport, G. Arnolda, et al., Clinicians' attitudes and perceived barriers and facilitators to cancer treatment clinical practice guideline adherence: a systematic review of qualitative and quantitative literature, Implement. Sci. 15 (2020) 39.
[9] https://www.iconplc.com/solutions/technologies/symphony-health.
[10] J.R. Jett, S.E. Schild, K.A. Kesler, et al., Treatment of small cell lung cancer diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines, Chest 143 (5_suppl) (2013) e400S–e419S.
[11] M.E. Charlson, T.P. Szatrowski, J. Peterson, et al., Validation of a combined comorbidity index, J. Clin. Epidemiol. 47 (11) (1994) 1245–1251.
[12] A. Lamer, G. Laurent, S. Pelayo, et al., Exploring patient path through Sankey diagram: a proof of concept, IOS 270 (2020) 218–2022.
[13] K. Falconer, Fractal Geometry, Wiley, 2003.
[14] Hazra & Gogtay, Biostatistics series module 9: survival analysis, Indian J. Dermatol. 62 (3) (2017 May-Jun) 251–257.
[15] S. Carnio, S. Novello, M. Papotti, et al., Prognostic and predictive biomarkers in early stage non small-cell lung cancer: tumor based approaches including gene signatures, Transl. Lung Cancer Res. 2 (2013) 372–381.
[16] D. Sonada, Y. Matsuura, M. Nakao, et al., Cancer Manag. Res. 11 (2019) 6765–6774.
[17] R. Carroll, M. Bortolini, A. Calleja, et al., Trends in treatment patterns and survival outcomes in advanced non-small cell lung cancer: a Canadian population-based real-world analysis, BMC Cancer 22 (1) (2022 Mar 10) 255.
[18] J.C. Simeone, B.L. Nordstrom, K. Patel, et al., Treatment patterns and overall survival in metastatic non-small-cell lung cancer in a real-world, US setting, Future Oncol. 15 (30) (2019 Oct) 3491–3502.