Machine Learning-Based Analysis of Treatment Sequences Typology in Advanced Non-Small-Cell Lung Cancer Long-Term Survivors Treated With Nivolumab

Christos Chouaïd, Valentine Grumberg, Alexandre Batisse, Romain Corre, Matteo Giaj Levra, Anne-Françoise Gaudin, Martin Prodel, Joannie Lortet-Tieulent, Jean-Baptiste Assié, Francois-Emery Cotté

ARTIFICIAL INTELLIGENCE

original reports abstract

Machine Learning–Based Analysis of Treatment Sequences Typology in Advanced Non–Small-Cell Lung Cancer Long-Term Survivors Treated With Nivolumab

Christos Chouaïd, MD[1]; Valentine Grumberg, PharmD[2]; Alexandre Batisse, MScEng[3]; Romain Corre, MD[4]; Matteo Giaj Levra, MD[5]; Anne-Françoise Gaudin, PharmD[2]; Martin Prodel, PhD[3]; Joannie Lortet-Tieulent, PhD[3]; Jean-Baptiste Assié, MD[1,6]; and Francois-Emery Cotté, PhD[2]

PURPOSE Immune checkpoint inhibitors substantially changed advanced non–small-cell lung cancer (aNSCLC) management and can lead to long-term survival. The aims of this study were (1) to use a machine learning method to establish a typology of treatment sequences on patients with aNSCLC who were alive 2 years after initiating a treatment with anti–programmed death-ligand 1 monoclonal antibody nivolumab and (2) to describe the patients' characteristics according to the typology of treatment sequences.

MATERIALS AND METHODS This retrospective observational study was based on data from the comprehensive French hospital discharge database for all patients with lung cancer with at least one line of platinum-based chemotherapy, starting nivolumab between January 1, 2015, and December 31, 2016, and alive 2 years after nivolumab treatment initiation. Patients were followed until December 31, 2018. A typology of most common treatment sequences was established using hierarchical clustering with time sequence analysis.

RESULTS Two thousand two hundred twelve study patients were, on average, 63.0 years old, 69.9% of them were men, and 61.9% had a nonsquamous cell carcinoma. During the 2 years after nivolumab treatment initiation, clusters of patients with four basic types of treatment sequences were identified: (1) almost continuous nivolumab treatment (44% of patients); (2) nivolumab most of the time followed by a treatment-free interval or a chemotherapy (15% of patients); and a short or medium nivolumab treatment, followed by (3) a long systemic treatment-free interval (17% of patients) or (4) a long chemotherapy (23% of patients).

CONCLUSION This machine learning approach enabled the identification of a typology of four representative treatment sequences observed in long-term survival. It was noted that most long-term survivors were treated with nivolumab for well over 1 year.

JCO Clin Cancer Inform 6:e2100108. © 2022 by American Society of Clinical Oncology

ASSOCIATED CONTENT

Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on December 15, 2021 and published at ascopubs.org/journal/cci on February 3, 2022: DOI https://doi.org/10.1200/CCI.21.00108

INTRODUCTION

Immune checkpoint inhibitors (ICIs) substantially changed advanced non–small-cell lung cancer (aNSCLC) management in the second-line setting and more recently, in the first-line setting.[1,2] Evidence that ICI is extending survival and that long-term survival can be achieved is accumulating, in both clinical trials and real-world setting. Although no consensus yet exists, several studies have established the definition of long-term survival in aNSCLC at more than 2 years from the time of ICI administration.[3] For instance, in a second-line setting, survival benefit has been demonstrated in phase III randomized clinical trials for nivolumab, atezolizumab, and pembrolizumab compared with docetaxel.[1] The clinical outcomes obtained in clinical trials for selected patients were later confirmed in the real-world setting.[3-5]

Real-world studies also explored nivolumab outcomes for subgroups of patients (elderly patients with brain metastases or renal impairment) and demonstrated overall survival (OS) benefit.[4]. In patients treated with nivolumab in second-line setting or later, almost half receive another treatment line after nivolumab[6] and a substantial proportion is retreated with nivolumab after a treatment-free interval or chemotherapy.[7] Despite the increasing importance of long-term data on ICI, only a few publications focus on long-term survivors.[8-10] The observed treatment sequences of these long-term survivors can provide

CONTEXT

Key Objective

Because cancer treatment sequences in the real-world setting are complex and variable, it is hard to see the big picture when thousands of patients are involved. Using the French national hospital discharge database, this study applies a machine learning approach to determine a typology of treatment sequences in more than 2,200 patients with advanced non–small-cell lung cancer treated with an immunotherapy (ie, nivolumab) and are alive 2 years after initiating this treatment.

Knowledge Generated

Four treatment sequences were identified in these long-term survivors, with different characteristics. Most of these patients were continuously or almost continuously treated with nivolumab for 2 years. The others were treated with nivolumab for a shorter period, followed by a systemic treatment-free period or by a chemotherapy.

Relevance

The use of this machine learning method allows us to get a clear picture of treatment sequences observed in a large patient population with complex treatments.

insight into the much-needed optimal treatment sequences and durations.[11]

Using real-world health care claims data to analyze treatment sequences is an arduous task, and interpretation of the output is difficult. Mismatches in expected drug-dispensing dates, one-time treatment swaps to replace sold out drugs, or changes in clinical practice over time all add complexity to the task. In addition, the large number of patients and treatment combinations hinders an easy interpretation of the results. Thus, using artificial intelligence to tackle big data becomes unavoidable. In particular, machine learning is ideally positioned to overcome these challenges. In France, the availability of a comprehensive national hospital database (programme de médicalisation des systèmes d'information [PMSI]) offers a unique opportunity to analyze treatment patterns of a large number of patients in a real-world setting.[12,13]

Using the PMSI database, the objectives of this study were (1) to use a machine learning method to establish a typology of treatment sequences on patients with aNSCLC (stage IIIb-IV) who were alive 2 years after initiating a treatment with nivolumab in 2015-2016 and (2) to describe the patients' characteristics according to the typology of treatment sequences.

MATERIALS AND METHODS

Study Design and Study Population

The study design and patient identification process have been published elsewhere[4,7] and are summarized here. This was a retrospective observational study on the basis of the PMSI database, which includes records for patients receiving outpatient anticancer treatment infusions.[13] This database collects reason for hospitalization and health care resource utilization information, at an individual level, from all French public and private hospitals.

The study included all patients with lung cancer (International Classification of Diseases-10 code: C34*) who received at least one line of platinum-based chemotherapy and started nivolumab (the only ICI available) between January 01, 2015, and December 31, 2016 (ie, in the early access period). Data collection started on January 01, 2011, to capture history of lung cancer, comorbidities, and previous treatments. Patients were followed until December 31, 2018. Date of lung cancer diagnosis was defined as date of first hospital admission with a lung cancer diagnosis. The analysis was restricted to patients alive 2 years after nivolumab treatment initiation.

Study Outcomes

The input of the clustering analysis was all patient treatment sequences (Fig 1, 1. Preliminary data processing). For each patient, aNSCLC-related drug administrations were identified and dated. Sequences of the following systemic treatments were captured: nivolumab-, pembrolizumab-, pemetrexed-, and/or bevacizumab-based chemotherapy, or other chemotherapies, and treatment-free intervals (ie, systemic treatment-free, as we did not collect radiotherapies). Atezolizumab was not available at the time. For visualization purposes, in the cluster analysis, all chemotherapy-based protocols and single immunotherapies were combined into two categories (chemotherapy and immunotherapy, respectively). The duration of each systemic treatment, treatment-free period, and the median cumulative duration for each patient were determined. Deaths during hospital stays were also identified. Patients were censored either 6 weeks after the last day of the hospital stay for nivolumab and other systemic treatment administrations in the absence of subsequent hospital admission or the date of their last hospital admission.

All patients were modeled as discretized time vectors (with a 1-day step) representing treatments throughout the follow-up (Fig 1, 2. Time sequence vectors). This transformation from sparse drug dispensing to a fully characterized treatment sequence was made possible by two elements: first, drug information in the PMSI and second,

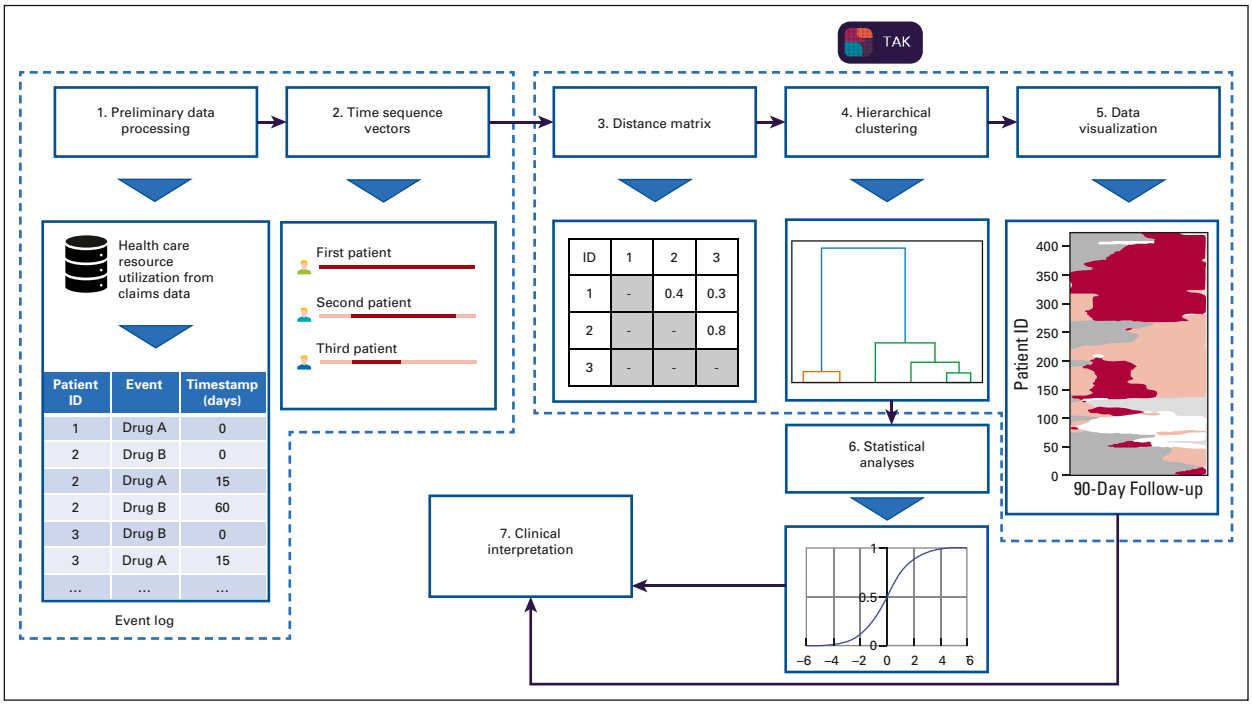

FIG 1. Examples of patient time vectors. In the boxes under time sequence vectors and data visualization, each color represents a different treatment. Each color represents a different treatment sequence, and the line's length represents the treatment duration. TAK, time sequence analysis through K-clustering.

medical experts' knowledge to determine thresholds and to define constraints on the basis of the European guidelines and therapeutic classes in use.[14] The vector size corresponded to the longest follow-up duration in the study.

The clusters were described by their type of treatment sequence (treatments taken and treatment-free periods), the median cumulative duration of each treatment, and patient characteristics at inclusion.

Analytical Process

Machine learning analysis. The machine learning analysis enabled the identification of clusters of patients with similar treatment sequences. A preliminary analysis with a simple ordering strategy was applied to form a baseline. Patients were sorted according to the duration of their first nivolumab treatment without discontinuation (Fig 2). No further ordering of subsequent treatments was performed.

Then, we applied the Time sequence Analysis through K-clustering (TAK) method,[15,16] hereafter described.

Patients' time vectors were clustered using an unsupervised hierarchical Ward's clustering method.[17] It is a two-step algorithm. First, a matrix of pairwise distances between all patients is computed (Fig 1, 3. Distance matrix) to assess dissimilarities between patients. Second, Ward's linkage method was used to build patient clusters (Fig 1, 4. Hierarchical clustering). The linkage builds nested clusters from the pairwise distances by minimizing the distance between patients within the cluster. These clusters are structured as a tree: each patient is a singleton cluster (a leaf) and then patients are grouped in intermediate nodes, up to the root node—which contains the whole patient population. The top nodes below the root node divide the patient population into groups with similar treatment patterns. The appropriate number of clusters was determined with the medical experts' input.

Both the previous steps rely on the choice of a distance metric. We chose the Hamming distance[18] for two reasons: first, for its robustness to small discrepancies between two patients' sequences (eg, two 365-day-long vectors with one 1-day discrepancy are considered as 99.7% identical) and second, it provides a score on the basis of the vectors' length.

Following the optimal leaf ordering for the hierarchical clustering method developed by Bar-Joseph et al,[19] we sorted patients in the clustering tree by flipping each node to minimize the previously defined distance between the innermost leaves of two adjacent nodes. This computation simplifies the temporal patterns found in the treatment sequences. Thereby, we obtained a treatment matrix with the time dimension on the *x*-axis and all patients stacked on the *y*-axis.

Then, we applied a noise filtering technique to this matrix to only retain meaningful treatment patterns. This technique is a modal filter parametrized by a kernel size, which replaces each value by the most frequent value in the kernel. The kernel size was chosen on the basis of the population

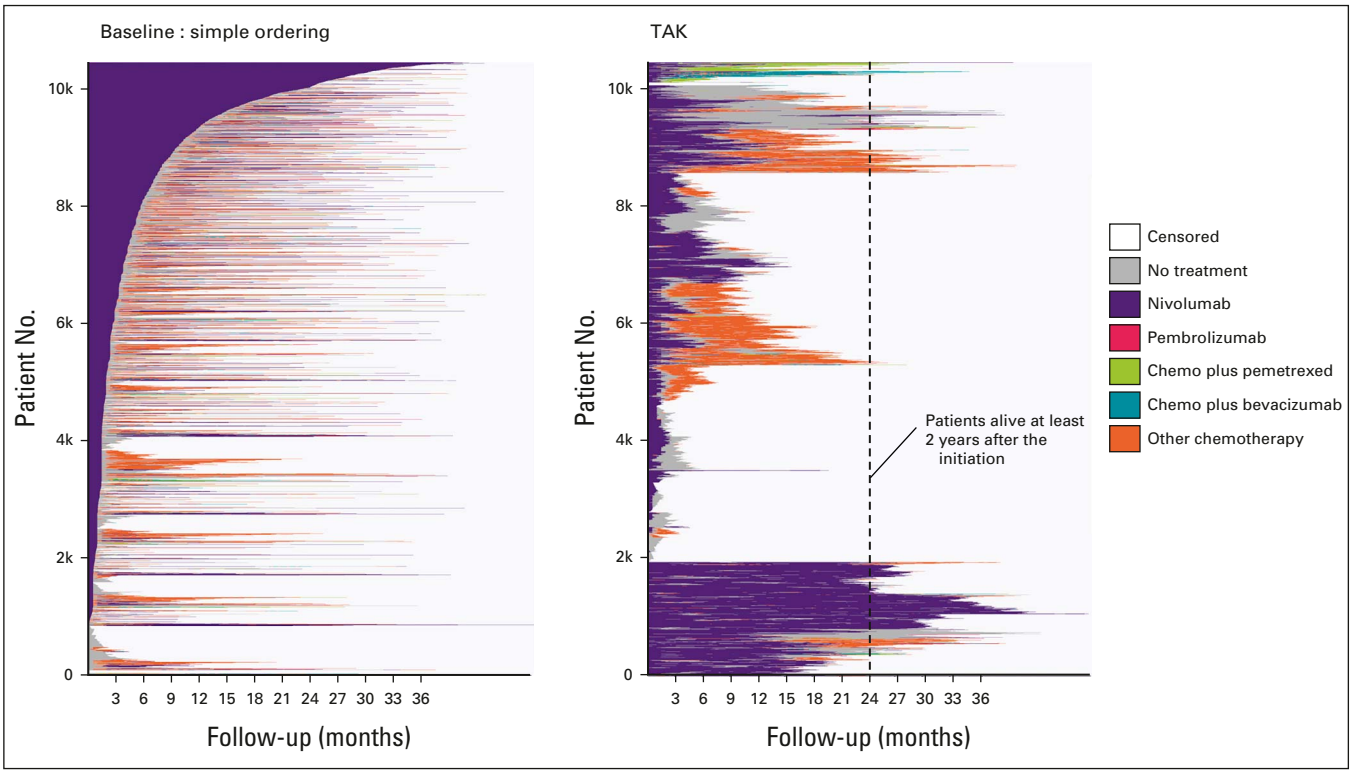

FIG 2. The superiority of machine learning over treatment duration ordering to find treatment sequence clusters in the population of all patients with advanced non–small-cell lung cancer initiating nivolumab treatment in 2015 and 2016, regardless of the 2-year survival outcome (N = 10,452). (A) Simple patient ordering by first nivolumab treatment duration. (B) TAK result: the patients are sorted by typology of treatment sequences. Each color represents a different treatment sequence, and the line's length represents the treatment duration. The dotted line marks down 2 years after the nivolumab treatment initiation. No treatment refers to no systemic treatment. Censored: dead or no longer followed. chemo, chemotherapy; OS, overall survival; TAK, time sequence analysis through K-clustering.

size and the expected image output dimensions. We used a 7 × 10 pixel kernel to exclude temporal patterns smaller than a week for 10 patients.

Finally, we deleted nonsensical artifacts (eg, patients receiving a treatment after being dead; Fig 1, 5. Data visualization).

Initially, we applied the TAK method to all patients initiating nivolumab during the inclusion period as a benchmark against the simple ordering of the baseline nivolumab treatment. Then, we only applied it to patients still alive 2 years after initiating nivolumab.

The TAK was implemented in Python 3.7, with the SciPy library.[20]

Statistical analyses on the patient clusters. We used a two-step analysis[21] to investigate the clusters identified using the TAK method (Fig 1, 6. Statistical analyses). First, to test for differences across clusters with respect to variables of interest, we used the chi-square test for categorical variables and one-way analysis of variance for continuous variables. Second, we investigated the significant differences across clusters at a $P \leq .10$ threshold using multinomial logistic regression analysis to identify associations of the clusters with the abovementioned variables of interest. Associations are presented as odds ratios with 95% CIs. Hence, our dependent variable had four possible categories (cluster 1, cluster 2, cluster 3, and cluster 4).

RESULTS

Treatment Sequences Analysis

During the inclusion period, 10,452 patients with aNSCLC initiated a nivolumab treatment. Figure 2 shows the advantage of the TAK method over simple ordering of patient treatment sequences by duration of the first nivolumab treatment, in all patients who initiated nivolumab during the inclusion period, regardless of their 2-year clinical outcome. As TAK allowed the combination of patients with similar treatment patterns, groups of patients with different OS became evident: short (about 3 months), medium (about 12 months), and long survival (at least 24 months).

The patients who were alive 2 years after treatment initiation (n = 2,212, 21.2%) were further analyzed. In this group, during the first 2 years after nivolumab treatment initiation, the median cumulative duration of nivolumab treatment was 14.3 months. Most patients had a treatment-free period (98.5%) and a sequence of chemotherapy

treatment (57.5%), for median cumulative durations of 5.7 and 4.1 months, respectively (Table 1). Percentages do not add up to 100% because patients have several types of treatments (nivolumab, chemotherapies) or absence of systemic treatment.

On the basis of the TAK analysis applied to patients alive 2 years after nivolumab initiation and input from medical experts, four basic types of treatment sequences administered to clusters of patients were identified (Fig 3). The largest cluster comprised 44.4% (982 of 2,212) of the patients. They were mainly treated with nivolumab (median cumulative nivolumab duration of 21.0 months; Table 1).

In the second cluster (327 of 2,212, 14.8%), patients received nivolumab for a long period (median 16.5 months), followed, in 99.4% of the cases, by a treatment-free period (median 5.3 months) and, in 63.6% of the cases, by a short sequence of chemotherapy (median 2.5 months).

Patients in the third cluster (385 of 2,212, 17.4%) received nivolumab for a limited time (median cumulative duration of 6.4 months), followed by a long treatment-free period of more than a year (median 14.4 months, 100% of patients). Between nivolumab and the treatment-free period, some patients received a short chemotherapy sequence (median of 4.5 months).

The last cluster (518 of 2,212, 23.4%) combined a short to medium nivolumab treatment period (median 5.5 months) followed by a long chemotherapy (median 9.5 months, 100% of patients).

Patient Characteristics

Upon initiation of nivolumab treatment, the 2,212 patients were, on average, 63.0 years old, 69.9% of them were men, 61.9% had a nonsquamous cell carcinoma, and 16.9% had brain metastases. The median time since lung cancer diagnosis was 22.9 months. Before inclusion in the study, 16.0% and 23.9% had been treated with curative surgery and curative radiotherapy, respectively. The most frequent comorbidities were hypertension (16.6%) and chronic obstructive pulmonary disease (12.1%; Table 2).

In univariate analysis, age, time since lung cancer diagnosis, presence of brain metastases, prior surgery, and prior radiotherapy were significantly different between clusters (Table 2). Pairwise comparisons showed that cluster 1 patients were younger than patients in the other three clusters, more of them had been treated with radiotherapy than cluster 3 and 4 patients, and a shorter time since diagnosis than cluster 3 patients. There were no significant differences between clusters 2 and 3 patients. Cluster 3 patients were older than cluster 4 patients. More cluster 2 patients had brain metastases than cluster 4 patients (Table 3, Appendix 1).

DISCUSSION

During the inclusion period, 10,452 patients with aNSCLC initiated nivolumab as a second-line or later and more than a fifth survived for at least 2 years. Using machine learning and expert knowledge, we were able to identify clusters of patients with four distinct treatment sequences over these 2 years. In the largest cluster (44% of patients), patients were almost continuously treated with nivolumab; in the second cluster (15%), patients received nivolumab most of the time, followed by a treatment-free interval or chemotherapy sessions. Altogether, the majority of long-term survivors were treated with nivolumab for well over 1 year. In the last two clusters, patients had a short or medium nivolumab treatment period (median duration around 6 months), followed by a long treatment-free interval (median duration of about

TABLE 1. Treatment Characteristics of Patients With Advanced Non–Small-Cell Lung Cancer, During the First 2 Years After Nivolumab Initiation, by Cluster

Treatment Characteristics	All Patients (N = 2,212)	Cluster 1 (n = 982)	Cluster 2 (n = 327)	Cluster 3 (n = 385)	Cluster 4 (n = 518)
Nivolumab					
Patients with at least one nivolumab treatment, No. (%)	2,212 (100.0)	982 (100.0)	327 (100.0)	385 (100.0)	518 (100.0)
Cumulative duration, median, months	14.3	21.0	16.5	6.4	5.5
Chemotherapies					
Patients with at least one chemotherapy, No. (%)	1,271 (57.5)	336 (34.2)	208 (63.6)	209 (54.3)	518 (100.0)
Cumulative duration, median, months	4.1	0.7	2.5	4.5	9.5
Treatment-free interval					
Patients with at least one treatment-free interval, No. (%)	2.179 (98.5)	956 (97.4)	325 (99.4)	385 (100.0)	513 (99.0)
Cumulative duration, median, months	5.7	2.7	5.3	14.4	7.5

NOTE. Nivolumab cumulative duration: sum of the times from administration until 14 days after administration, except in the case of death or censoring. Chemotherapy cumulative duration: sum of the times from administration until 21 days after administration, except in the case of death or censoring. Chemotherapy: any chemotherapy. Treatment-free interval refers to systemic treatment-free interval. Percentages do not add up to 100% because patients have several types of treatments (nivolumab, chemotherapies) or the absence of systemic treatment.

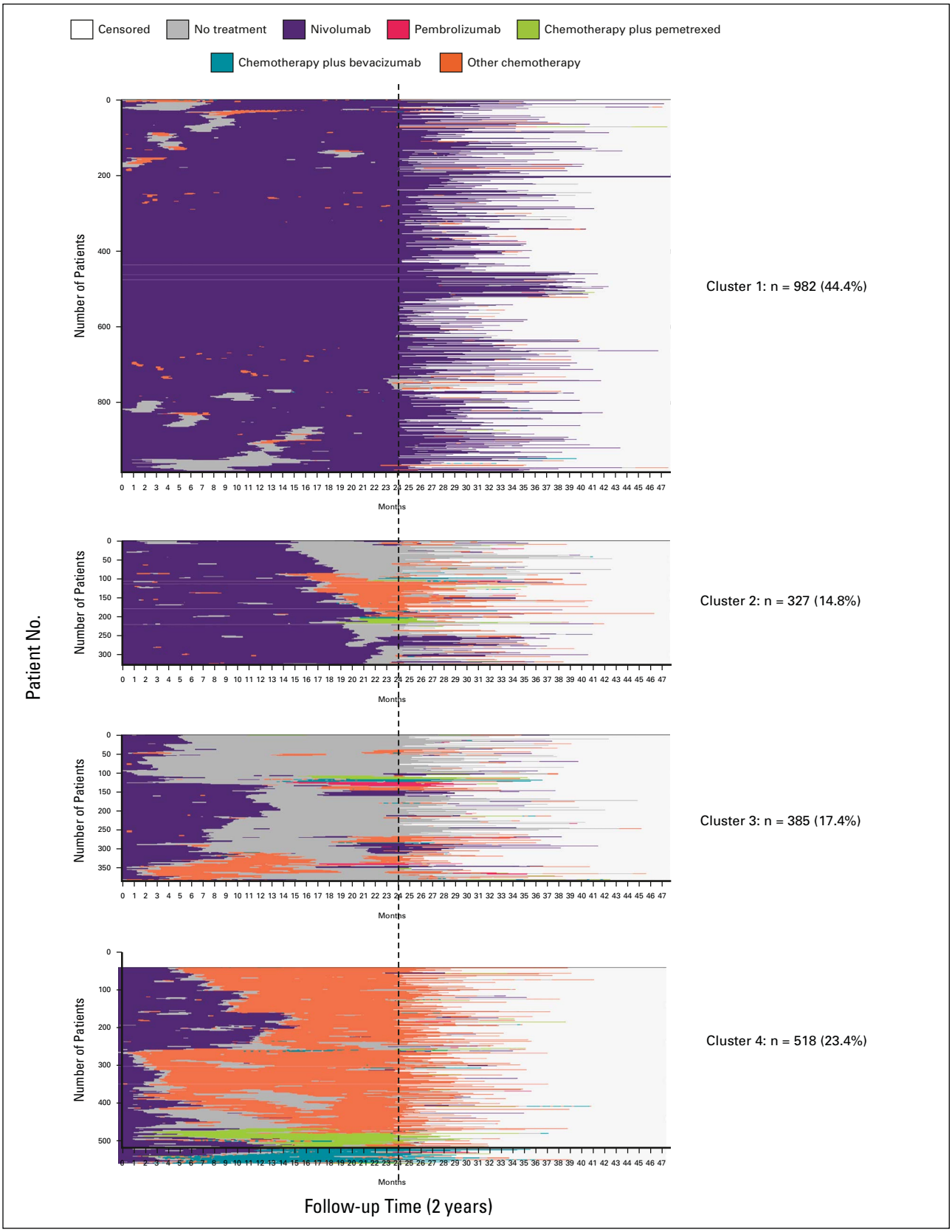

FIG 3. Four treatment sequence clusters in patients alive 2 years after initiating nivolumab treatment (n = 2,212). The dotted line marks down 2 years after the nivolumab treatment initiation. Clusters were searched on treatment sequences between nivolumab treatment initiation and 2 years after. No treatment refers to no systemic treatment. Censored: dead or no longer followed.

14 months; cluster 3) or a long chemotherapy treatment period (median duration of about 10 months; cluster 4).

Consistent with the results of long-term clinical trials, we showed that more than 25% of patients who received an ICI in the second or subsequent line of therapy survived for at least 2 years.[22,23]

The optimal ICI treatment duration is still being investigated,[11] and whether patients with a complete response should withdraw from ICI treatment is still being debated. Yet, few real-world studies have explored the characteristics of long-term survivors after second-line immunotherapy[24] and even fewer have focused on the description of the therapeutic sequences administered to these long-term survivors. We observed that most 2-year survivors were continuously (or almost continuously) treated with nivolumab. Interestingly, we found that cluster 1 patients (the largest cluster, those continuously treated with nivolumab) were the most different from patients of the other clusters. They were younger and more often had been treated with radiotherapy. This possible synergy of ICI and radiotherapy has previously been noted, and its potential mechanisms are being explored.[25,26] Another salient cluster is the one with patients briefly treated with nivolumab (cluster 3) and receiving no further systemic treatment or only a short chemotherapy. The prolonged treatment-free interval may lead one to think that the disease of some of these patients shows a stable complete response to treatment. Our analyses revealed that patients in this cluster appeared somewhat different from patients in other clusters (older, fewer radiotherapies than in cluster 1, a longer time since diagnosis) without pinpointing what really differentiates them from the other patients. From a clinical and an economic point of view, future research to determine these patients' profiles would be particularly advantageous as they are those whose nivolumab treatment has been the most cost-effective.

Classical descriptive tools are insufficient to describe the complexity of long-term survivors' management.[24,27] The simple ordering of treatment sequences by one treatment duration only gives a hint at the heterogeneity of practices and fails to provide clear, interpretable results. To improve understanding of treatment sequences, previous studies mostly used analytic tools on the basis of probabilistic state

TABLE 2. Patient Characteristics at Nivolumab Initiation

Patient Characteristics	All Patients (N = 2,212)	Cluster 1 (n = 982)	Cluster 2 (n = 327)	Cluster 3 (n = 385)	Cluster 4 (n = 518)	Univariate *P* Values
Age, mean ± SD, years	63.0 ± 9.5	62.2 ± 9.7	63.7 ± 9.7	64.6 ± 9.0	63.2 ± 8.9	**.0045**
Sex (men), No. (%)	1,546 (69.9)	692 (70.5)	235 (71.9)	267 (69.4)	352 (68.0)	.628
Time since lung cancer diagnosis, mean ± SD, months	22.9 ± 22.2	21.6 ± 21.9	24.6 ± 24.4	25.1 ± 23.0	22.8 ± 20.4	**.060**
Histologic subtype (nonsquamous), No. (%)	1,369 (61.9)	614 (62.5)	191 (58.4)	4,783 (56.3)	328 (63.3)	.502
Malnutrition, No. (%)	267 (12.1)	133 (13.5)	35 (10.7)	44 (11.4)	55 (10.6)	.288
Brain metastases, No. (%)	373 (16.9)	194 (19.8)	59 (18.0)	52 (13.5)	68 (13.1)	**.002**
Comorbidities,[a] No. (%)						
Hypertension	367 (16.6)	165 (16.8)	62 (19.0)	69 (17.9)	71 (13.7)	.174
Diabetes	168 (7.6)	82 (8.4)	25 (7.6)	30 (7.8)	31 (6.0)	.434
Renal impairment	87 (3.9)	40 (4.1)	16 (4.9)	9 (2.3)	22 (4.2)	.311
Chronic obstructive pulmonary disease	267 (12.1)	110 (11.2)	51 (15.6)	48 (12.5)	58 (11.2)	.175
Pulmonary insufficiency	34 (1.5)	13 (1.3)	9 (2.8)	6 (1.6)	6 (1.2)	.264
Other chronic pulmonary diseases	193 (8.7)	86 (8.8)	35 (10.7)	33 (8.6)	39 (7.5)	.466
Treatment history, No. (%)						
Prior curative surgery	353 (16.0)	138 (14.1)	56 (17.1)	77 (20.0)	82 (15.8)	**.053**
Prior radiotherapy	528 (23.9)	272 (27.7)	75 (22.9)	74 (19.2)	107 (20.7)	**.001**
Type of hospital, No. (%)						.453
Local hospital	778 (35.2)	350 (35.6)	123 (37.6)	138 (35.8)	167 (32.2)	
University hospital	652 (29.5)	286 (29.1)	101 (30.9)	97 (25.2)	168 (32.4)	
Center for cancer care and research	225 (10.2)	96 (9.8)	22 (6.7)	46 (11.9)	61 (11.8)	
Others	557 (25.2)	250 (25.5)	81 (24.8)	104 (27.0)	122 (23.6)	

NOTE. The values in bold indicates a global significant difference ($P < .10$) across clusters.
Abbreviation: SD, standard deviation.
[a]Comorbidities are those identified within 1 year before patient inclusion.

TABLE 3. Synthesis of the Results of Pairwise Multinomial Logistic Regression Analyses

Clusters Pairwise Analyses	Cluster 1: Full-Time Nivolumab (n = 982)	Cluster 2: Nivolumab Most of the Time Followed by Treatment-Free Interval or Chemotherapy (n = 327)	Cluster 3: Short or Medium Nivolumab Treatment, Followed by Long Treatment-Free Interval (n = 385)
Cluster 2: nivolumab most of the time followed by treatment-free interval or chemotherapy (n = 327)	Cluster 1 *v* cluster 2: younger < 60 years, OR = 0.26 (0.97 to 1.74) > 70 years, OR = 1.56 (1.11 to 2.19)	NA	NA
Cluster 3: short or medium nivolumab treatment, followed by a long treatment-free interval (n = 385)	Cluster 1 *v* cluster 3: younger. < 60 years, OR = 0.29 (0.22 to 0.37) > 70 years, OR = 1.42 (1.02 to 1.96) Cluster 1 *v* cluster 3: more radiotherapy OR = 0.61 (0.45 to 0.82) Cluster 1 *v* cluster 3: shorter time since diagnosis, more than 24 months OR = 1.52 (1.12 to 2.06)	No significant difference	NA
Cluster 4: short or medium nivolumab treatment followed by a long chemotherapy treatment (n = 518)	Cluster 1 *v* cluster 4: younger < 60 years: OR = 0.47 (0.37 to 0.59) > 70 years, OR = 1.16 (0.87 to 1.56) Cluster 1 *v* cluster 4: more radiotherapy OR = 0.69 (0.53 to 0.90)	Cluster 2 *v* cluster 4: older < 60 years: OR = 1.82 (1.34 to 2.47) > 70 years OR = 0.75 (0.51 to 1.09) Cluster 2 *v* cluster 4: more brain metastases OR = 0.66 (0.45 to 0.99)	Cluster 3 *v* cluster 4: older < 60 years OR = 1.63 (1.22 to 2.19) > 70 years OR = 0.82 (0.57 to 1.18)

NOTE. The results of the pairwise multinomial logistic regressions are detailed in Appendix Table A1. The treatment-free interval refers to the systemic treatment-free interval. OR, followed by the 95% CI in brackets.

Abbreviations: NA, not applicable; OR, odds ratio.

models (decision trees, Markov models, discrete event simulation, and survival analysis). However, the limits to these models are threefold. First, they lack scalability beyond a few hundred patients.[8,10,28] Second, treatment sequences need to be oversimplified in the modeling phase, thereby disregarding valuable information. Such simplification results in the omission of the time dimension of the treatment sequence and of some inherent characteristics of real-world treatment sequences (such as short treatment phases and treatment-free intervals).[8,10,28,29] Third, these previous studies only focused on the analytic performances of the ensuing clusters, not on the clinical interpretation of the outputs. Le Meur et al,[30] Vogt et al,[31] and Roux et al[32] used methods similar to ours in various therapeutic areas, allowing for the scale-up of the modeling beyond a 1,000 patients while keeping the time dimension. In our study, we further developed algorithms to enable the analysis of > 10,000 patients while keeping all the information of their treatment sequences. In addition, the straightforward visual display of the results made it possible for the medical experts to discover compelling treatment patterns.

The limitations of the study mainly pertain to the database. As a claims database, the PMSI does not provide results of imaging procedures or laboratory tests such as programmed death-ligand 1 expression nor clinical records. For instance, performance status upon nivolumab initiation, corticosteroid treatment at baseline, and metastasis' localization might have an impact on the duration of nivolumab and its efficacy. This information would ease the interpretation of treatment switches and treatment-free intervals, which could be due to response to treatment or treatment toxicity. Access to clinical information would also expand our knowledge of the patient profiles in the four clusters. Another limitation is that only in-hospital deaths were recorded in our study. In the absence of hospital care for at least 6 weeks, the patient was censored. Although this decision was reasonable for several reasons (the 2-year metastatic lung cancer survival rate is only 27%,[33] 77% of lung cancer deaths occur at the hospital,[34] and lung cancer treatment requires frequent hospital admissions), it is possible, although unlikely, that we failed to include patients who stopped all hospital interactions but were alive 2 years after initiating nivolumab treatment. The treatment sequences of 2-year survivors cannot be compared with patients living for < 2 years. In the first place, their treatment duration is shorter (median treatment duration of the whole nivolumab-treated cohort: 2.8 months). Finally, clinical practices have evolved since the study was carried out: some—but not all—patients with aNSCLC are now administered nivolumab as a first-line treatment.[1]

This study also has multiple strengths. To our knowledge, this is the first study to evaluate treatment sequences of patients who initiated nivolumab for aNSCLC in France, in a real-world setting. On the basis of a database covering 100% of hospital admissions and nivolumab being administered at the hospital, all patients benefiting from the treatment were captured in the study. The large study population allowed for meaningful subgroup (clusters) analyses. Finally, identification of study population and treatments in the database is accurate because recording of treatments in the PMSI is essential to hospitals' funding.

In conclusion, using a large population of patients with aNSCLC living 2 years after the initiation of nivolumab in the second-line setting or beyond, this machine learning approach enabled the identification of a typology of four clusters of patients with four quintessential treatment sequences. Nevertheless, an in-depth study of patients' clinical profiles is required to better understand the best treatment sequences for the longest OS according to patients' characteristics.

AFFILIATIONS

[1]Service de pneumologie, Centre Hospitalier Intercommunal de Créteil, Créteil, France
[2]Bristol Myers Squibb France, Rueil-Malmaison, France
[3]HEVA, Lyon, France
[4]Centre Hospitalier Intercommunal de Cornouaille, Quimper, France
[5]Centre Hospitalier Universitaire Grenoble Alpes (CHUGA), Grenoble, France
[6]Centre de Recherche des Cordeliers, Inserm, Université de Paris, Sorbonne Université, Functional Genomics of Solid Tumors Laboratory, Paris, France

CORRESPONDING AUTHOR

Francois-Emery Cotté, PhD, Laboratoire Bristol Myers Squibb, Health Economics and Outcomes Research, 3 rue Joseph Monier 92500, Rueil-Malmaison, France; e-mail: Francois-Emery.Cotte@bms.com.

PRIOR PRESENTATION

Presented as an abstract and a poster at the 2020 virtual ISPOR congress (the professional society for health economics and outcomes research), November 19, 2020.

SUPPORT

Supported by Bristol Myers Squibb (Princeton, NJ). J.-B.A. was supported by grants from Fondation pour la Recherche Médicale (FRM).

AUTHOR CONTRIBUTIONS

Conception and design: Valentine Grumberg, Alexandre Batisse, Alexandre Batisse, Matteo Giaj Levra, Anne-Françoise Gaudin, Martin Prodel, Francois-Emery Cotté
Collection and assembly of data: Alexandre Batisse, Martin Prodel
Data analysis and interpretation: Christos Chouaïd, Valentine Grumberg, Alexandre Batisse, Alexandre Batisse, Matteo Giaj Levra, Anne-Françoise Gaudin, Martin Prodel, Joannie Lortet-Tieulent, Jean-Baptiste Assié, Francois-Emery Cotté
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Christos Chouaïd
Honoraria: Amgen, Bristol Myers Squibb, MSD, AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, Takeda, Roche
Consulting or Advisory Role: Roche, MSD, Bristol Myers Squibb, AstraZeneca
Research Funding: AstraZeneca (Inst), Bristol Myers Squibb (Inst), MSD (Inst), Takeda (Inst), Pfizer (Inst), Novartis (Inst), Roche (Inst), Pierre Fabre (Inst)
Travel, Accommodations, Expenses: AstraZeneca, Roche, Pfizer, MSD, Bristol Myers Squibb, Boehringer Ingelheim

Valentine Grumberg
Employment: Bristol Myers Squibb

Alexandre Batisse
Employment: Stragen Services (I), HEVA
Travel, Accommodations, Expenses: HEVA

Romain Corre
Honoraria: Bristol Myers Squibb, MSD Oncology, AstraZeneca/MedImmune
Consulting or Advisory Role: AstraZeneca, Sanofi, Takeda

Matteo Giaj Levra
Consulting or Advisory Role: AstraZeneca/MedImmune, Roche, BMS
Research Funding: BMSi
Travel, Accommodations, Expenses: MSD

Anne-Françoise Gaudin
Employment: Bristol Myers Squibb

Joannie Lortet-Tieulent
Employment: HEVA

Jean-Baptiste Assié
Uncompensated Relationships: Bristol Myers Squib France

Francois-Emery Cotté
Employment: Bristol Myers Squibb

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank Baptiste Jouaneton for his supervision of the overall operational conduct of the study; Florent Daydé for the data extraction, data management, and part of the statistical analysis; and Hannah Lennon for the remaining part of the statistical analyses. All three of them work for HEVA, the company contracted by Bristol Myers Squibb France to carry out the study.

REFERENCES

1. Planchard D, Popat S, Kerr K, et al: Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 29:iv192-237, 2018 (suppl 4)
2. Alexander M, Kim SY, Cheng H: Update 2020: Management of non-small cell lung cancer. Lung 198:897-907, 2020

3. Nadal E, Massuti B, Dómine M, et al: Immunotherapy with checkpoint inhibitors in non-small cell lung cancer: Insights from long-term survivors. Cancer Immunol Immunother 68:341-352, 2019
4. Assié J-B, Corre R, Levra MG, et al: Nivolumab treatment in advanced non-small cell lung cancer: Real-world long-term outcomes within overall and special populations (the UNIVOC study). Ther Adv Med Oncol 12:175883592096723, 2020
5. Dixmier, Asselain B, Barlesi F, et al: IO-synthesise NSCLC: A pooled analysis of real-world survival outcomes for non-small cell lung cancer patients treated with nivolumab in France and Germany. Ann Oncol 30:v614, 2019
6. Costantini A, Corny J, Fallet V, et al: Efficacy of next treatment received after nivolumab progression in patients with advanced nonsmall cell lung cancer. ERJ Open Res 4:00120-2017, 2018
7. Giaj Levra M, Cotté F-E, Corre R, et al: Immunotherapy rechallenge after nivolumab treatment in advanced non-small cell lung cancer in the real-world setting: A national data base analysis. Lung Cancer 140:99-106, 2020
8. Hess LM, Kern DM, Carter GC, et al: Real-World Treatment Sequences and Outcomes among Patients with non-small cell lung cancer (RESOUNDS) in the United States: Study protocol. JMIR Res Protoc 6:e195, 2017
9. Davis KL, Goyal RK, Able SL, et al: Real-world treatment patterns and costs in a US Medicare population with metastatic squamous non-small cell lung cancer. Lung Cancer 87:176-185, 2015
10. Hochmair MJ, Morabito A, Hao D, et al: Sequential treatment with afatinib and osimertinib in patients with *EGFR* mutation-positive non-small-cell lung cancer: An observational study. Future Oncol 14:2861-2874, 2018
11. Friedlaender A, Kim C, Addeo A: Rethinking the optimal duration of immune checkpoint inhibitors in non-small cell lung cancer throughout the COVID-19 pandemic. Front Oncol 10:862, 2020
12. Bezin J, Duong M, Lassalle R, et al: The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 26:954-962, 2017
13. Tuppin P, Rudant J, Constantinou P, et al: Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique 65:S149-S167, 2017
14. Reck M, Popat S, Reinmuth N, et al: Metastatic non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 25:iii27-39, 2014 (suppl 3)
15. De Oliveira H, Pierres M, Raguideau F, et al: Étude des régimes d'affiliations dans le Système national des données de santé : Exploration à l'aide d'une méthode innovante de data visualisation. Revue d'Épidémiologie et de Santé Publique 68:S45-S46, 2020
16. Laurent M, Prodel M, Léotoing Lde, et al: Analysis of treatment sequences from the French national SNIIRAM database: Case study of incident people living with HIV in 2013. Value Health 22:S663, 2019
17. Ward JHJ: Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236-244, 1963
18. Hamming RW: Error detecting and error correcting codes. Bell Syst Tech J 29:147-160, 1950
19. Bar-Joseph Z, Gifford DK, Jaakkola TS: Fast optimal leaf ordering for hierarchical clustering. Bioinformatics 17:S22-S29, 2001 (suppl 1)
20. Virtanen P, Gommers R, Oliphant TE, et al: SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nat Methods 17:261-272, 2020
21. Larsson B, Gerdle B, Bernfort L, et al: Distinctive subgroups derived by cluster analysis based on pain and psychological symptoms in Swedish older adults with chronic pain—A population study (PainS65+). BMC Geriatr 17:200, 2017
22. Gettinger S, Horn L, Jackman D, et al: Five-year follow-up of nivolumab in previously treated advanced non–small-cell lung cancer: Results from the CA209-003 study. J Clin Oncol 36:1675-1684, 2018
23. Garon EB, Hellmann MD, Rizvi NA, et al: Five-year overall survival for patients with advanced non–small-cell lung cancer treated with pembrolizumab: Results from the phase I KEYNOTE-001 study. J Clin Oncol 37:2518-2527, 2019
24. Barlesi F, Dixmier A, Debieuvre D, et al: Effectiveness and safety of nivolumab in the treatment of lung cancer patients in France: Preliminary results from the real-world EVIDENS study. Oncoimmunology 9:1744898, 2020
25. Cheng M, Durm G, Hanna N, et al: Can radiotherapy potentiate the effectiveness of immune checkpoint inhibitors in lung cancer? Future Oncol 13:2503-2505, 2017
26. Spaas M, Lievens Y: Is the combination of immunotherapy and radiotherapy in non-small cell lung cancer a feasible and effective approach? Front Med 6:244, 2019
27. Horn L, Spigel DR, Vokes EE, et al: Nivolumab versus docetaxel in previously treated patients with advanced non–small-cell lung cancer: Two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). J Clin Oncol 35:3924-3933, 2017
28. Nishino K, Imamura F, Morita S, et al: A retrospective analysis of 335 Japanese lung cancer patients who responded to initial gefitinib treatment. Lung Cancer 82:299-304, 2013
29. Pereira-Salgado A, Kwan EM, Tran B, et al: Systematic review of efficacy and health economic implications of real-world treatment sequencing in prostate cancer: Where do the newer agents enzalutamide and abiraterone fit in? Eur Urol Focus 7:752-763, 2021
30. Le Meur N, Gao F, Bayat S: Mining care trajectories using health administrative information systems: The use of state sequence analysis to assess disparities in prenatal care consumption. BMC Health Serv Res 15:200, 2015
31. Vogt V, Scholz SM, Sundmacher L: Applying sequence clustering techniques to explore practice-based ambulatory care pathways in insurance claims data. Eur J Public Health 28:214-219, 2018
32. Roux J, Grimaud O, Leray E: Use of state sequence analysis for care pathway analysis: The example of multiple sclerosis. Stat Methods Med Res 28:1651-1663, 2019
33. Chouaïd C, Debieuvre D, Durand-Zaleski I, et al: Survival inequalities in patients with lung cancer in France: A nationwide cohort study (the TERRITOIRE study). PLoS One 12:e0182798, 2017
34. Gisquet E, Aouba A, Aubry R, et al: Où meurt-on en France? Analyse des certificats de décès (1993-2008). Bulletin Epidémiologique Hebdomadaire 48:547-551, 2012

APPENDIX 1. SUPPLEMENTARY MATERIAL

A machine learning–based analysis of treatment sequences typology in advanced non–small-cell lung cancer long-term survivors treated with nivolumab

TABLE A1. Pairwise Multinomial Logistic Regression Analysis of Variables Associated With Derived Clusters

	Cluster 1			Cluster 2		Cluster 3
	Cluster 2	Cluster 3	Cluster 4	Cluster 3	Cluster 4	Cluster 4
Clusters Pairwise Analyses	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Age, years						
< 60	0.26 (0.97 to 1.74)	**0.29 (0.22 to 0.37)**	**0.47 (0.37 to 0.59)**	1.11 (0.80 to 1.55)	**1.82 (1.34 to 2.47)**	**1.63 (1.22 to 2.19)**
60-70	1.30 (0.97 to 1.74)	**1.37 (1.04 to 1.80)**	1.24 (0.97 to 1.59)	1.05 (0.74 to 1.49)	0.96 (0.69 to 1.32)	0.91 (0.67 to 1.24)
> 70	**1.56 (1.11 to 2.19)**	**1.42 (1.02 to 1.96)**	1.16 (0.87 to 1.56)	0.91 (0.61 to 1.35)	0.75 (0.51 to 1.09)	0.82 (0.57 to 1.18)
Disease status						
Brain metastases	1.07 (0.76 to 1.50)	0.80 (0.56 to 1.13)	0.71 (0.52 to 0.97)	0.75 (0.49 to 1.14)	**0.66 (0.45 to 0.99)**	0.89 (0.59 to 1.33)
Time since lung cancer diagnosis, months						
< 12	1.00	1.00	1.00	1.00	1.00	1.00
12-24	1.08 (0.79 to 1.48)	1.30 (0.96 to 1.75)	1.20 (0.92 to 1.57)	1.20 (0.83 to 1.75)	1.12 (0.79 to 1.58)	0.93 (0.66 to 1.30)
> 24	1.21 (0.88 to 1.67)	**1.52 (1.12 to 2.06)**	1.30 (0.99 to 1.71)	1.26 (0.86 to 1.83)	1.07 (0.75 to 1.53)	0.86 (0.61 to 1.20)
Treatment history						
Prior curative surgery	1.19 (0.83 to 1.70)	1.34 (0.96 to 1.86)	1.05 (0.77 to 1.44)	1.13 (0.75 to 1.69)	0.89 (0.60 to 1.32)	0.79 (0.55 to 1.14)
Prior radiotherapy	0.77 (0.56 to 1.04)	**0.61 (0.45 to 0.82)**	**0.69 (0.53 to 0.90)**	0.79 (0.55 to 1.16)	0.91 (0.64 to 1.28)	1.14 (0.81 to 1.60)

NOTE. The values in bold indicate a significant difference ($P < .05$) between clusters. Only variables of interest statistically significantly different, at the $P \leq .10$ threshold (Table 2), were included in the pairwise multinomial logistic regression analysis.

Abbreviation: OR, odds ratio.