First-line anticancer drugs in patients with advanced, primary Non-Small Cell Lung Cancer: drug-utilization and effectiveness studies from Tuscany Region healthcare database

Protocol version 1.0

Title	First-line anticancer drugs in patients with advanced, primary Non-Small Cell Lung Cancer: drug-utilization and effectiveness studies from Tuscany Region healthcare database
Medicinal product(s) / Device(s)	Target therapies, immunotherapies and standard chemotherapies
Event(s) of interest	Not applicable
Research question and objectives	Research question: What is the drug-utilization and the effectiveness of first-line anticancer drugs used in advanced, primary NSCLC patients between 2009 and 2019 in Tuscany? Objectives: To describe drug utilization and effectiveness of first-line anticancer drugs in advanced primary NSCLC patients in Tuscany between 2009 and 2019
Country(ies) of study	Italy
Protocol author(s)	Andrea Spini, Rosa Gini, Giuseppe Roberto, Pietro Rosellini, Marina Ziche, Sandra Donnini, and Francesco Salvo

List of abbreviations

NSCLC	Non-small cell lung cancer
SCLC	Small cell lung cancer
PR	Pathology registry
AHD	Administrative healthcare database
HDS	Hospital data sources
HDR	Hospital discharge records
ODD	Outpatients drugs dispensing
OED	Outpatients encounter data
CT	Standard chemotherapy
uCT	Unspecified chemotherapy
TT	Target therapy
IT	Immunotherapy
uMAB	Monoclonal antibody-based therapy
NT	Not-treated patients
RT	Radiotherapy
HR	Hazard ratio
CI	Confidence Interval
PPV	Positive predictive value
SE	Sensitivity

Responsible parties

Name	Institution	Activity
Andrea Spini	Université de Bordeaux, Bordeaux, France; Università di Siena, Siena, Italy	Researcher
Francesco Salvo	Université de Bordeaux, Bordeaux, France;	Pharmaco-epidemiologist / Researcher
Giuseppe Roberto	Regional Health Agency of Tuscany, Florence, Italy	Pharmaco-epidemiologist / Researcher
Rosa Gini	Regional Health Agency of Tuscany, Florence, Italy	Data scientist / Researcher
Marina Ziche	University of Siena, Siena, Italy	Pharmacologist/ Researcher
Sandra Donnini	University of Siena, Siena, Italy	Pharmacologist/ Researcher
Rosellini Pietro	Université de Bordeaux, Bordeaux, France; Università di Siena, Siena, Italy	Oncologist / Researcher

Amendments and updates

Version	Description of changes	Study protocol section	Date of effectiveness
1.0			

1. Background

Lung cancer is the most commonly diagnosed cancer worldwide (2.09 million cases in 2018).(1) Based on biological characteristics, the World Health Organization considers lung cancer as small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). (2) The latter represents 85% of all cases of lung cancer and can be distinguished in two main histotypes: squamous, and non-squamous carcinoma. (3–5)

During the last fifteen years, the pharmacological treatment for advanced stage, non-resectable, NSCLC was revolutionized by the authorizations of innovative anticancer therapies, such as target- (TT) and immuno-therapies (IT). Before the approval of such innovative therapies, standard chemotherapy (CT) was the only possible pharmacological approach. (6)

Knowledge on efficacy and safety of authorized anticancer drugs mostly relies on evidence from clinical trials. Such studies are usually based on relatively small samples of strictly selected, well monitored, patient populations, which are generally followed for short time periods. (7) In the field of cancer, pharmacoepidemiology represent an essential tool to analyse drug utilization, survival and safety of anti-cancer medications in real word setting. (8,9)

Unfortunately, information needed to properly describe cancer outcomes are scattered in different data source. Administrative healthcare databases (AHD) are a valuable data source for observational studies and enables to describe real-life outcomes at population-based level. In the field of cancer, these databases, that mainly collect drugs reimbursements, professional services and hospitalizations, did not register fundamental cancer-related information as histology and stage that are needed to produce clear real-word evidence. Beyond this, hospital data source (HDS) as pathology registries (PR) or medical records often contain detailed clinical information, findings of physical examinations, and results of diagnostic tests. Unfortunately HDS are less accessible for research purposes than AHD, often due to time lags for updates or because they are not electronic and manual extraction of information is needed. Overall, whenever possible, the best approach to evaluate drug utilization and survival in cancer field seems to be record linkage between different data sources. (10)

Tuscany region has a region wide level PR linked with AHD collecting all healthcare information about Tuscan inhabitants reimbursed by the National Health Service.

The aim of the project is evaluate drug utilization and the effectiveness of first-line anticancer therapies approved for the treatment of advanced, non-resectable, NSCLC patients between 2009 and 2019 in the Tuscany population. Two studies will be thus performed on a cohort of patients with not-resectable, primary NSCLC: a drugs utilization study and a survival study.

2. Materials and Methods

2.1 Data sources

Two different, already pseudo-anonymized, data sources will be used: PR and AHD of the Tuscany region.

PR collects all the records of deidentified Tuscan inhabitants with a pathological evaluation coming from regional hospitals. This PR contains demographic characteristics, free-text fields (i.e. pathological diagnosis) and *Systematized NOmenclature of human and veterinary MEDicine* (SNOMED) morphology and topography codes, which identify the histology and the location of the disease respectively.

AHD collects longitudinal pseudonymized patient-level information on the utilization of healthcare services reimbursed by the National Healthcare Service, and dispensed to all subjects who are residents and registered with a general practitioner in the relevant catchment areas. For each subject registered in the database, through a pseudo-anonymized regional person identifier code, demographic data can be linked to different registries in which information on healthcare services delivered are recorded. For the purpose of this study, the following information will be used:

- *Inhabitant registry* (IR): this registry contains the following patient level information as age, gender and death date.

- *Outpatient Drug Dispensings* (ODD): this database contains national medicine approval code, brand name, formulation, with the associated ATC (Anatomical Therapeutic Chemical) code, as well as dispensing dates.
- *Hospital discharge records* (HDR): this database contains dates and codes of main diagnosis/procedure (ICD9-CM dictionary), secondary procedures, hospitalization dates. For the purpose of this study HDR will be used to select patients with a record of lung surgery and to investigate serious adverse events. Moreover, among the available procedures, ICD9-CM codes will be used to retrieve patients treated with unspecified chemotherapy or monoclonal base chemotherapy.
- *Outpatients Encounter Data* (OED): similarly to HDR, this database contains codes and dates for procedures, but only for ambulatory patients.

2.2 Cohort selection

All patients with a NSCLC record in PR registry of Tuscany Region between January 1, 2009 and June 30, 2019 will be identified. Patients with a diagnosis of NSCLC will be selected using a specific algorithm for the PR (the algorithm is reported in appendix: Table A1). This algorithm showed a PPV of 87.9% and a growing sensitivity (SE) up to 64.1% in 2017.

The date of the first NSCLC diagnosis recorded in PR will be considered as the index date. Thereafter, the PR will be linked to regional AHD, in order to select patients with ≥18 years old and with at least two years of look back. In order to select only patients with a primary NSCLC, only patients without a cancer diagnosis other than lung cancer recorded in HDR in the 5 years before index date will be included in the study cohort (ICD-9CM codes: 14*,15*,160*,161*,163*,164*,17*,18*,190*, 191*,192*,193*,194*,195*)

Moreover, in order to avoid patients with a cancer recurrence, only patients with a lung cancer diagnosis (162* ICD-9CM) recorded between 3 months and 5 years before index date will be excluded.

In order to select only lung cancer patients with an advanced stage of the disease, patients without a lung surgery (ICD9-CM codes: 32*) in the six months before or after index date will be included. With this choice, we aim to exclude those patients receiving neoadjuvant therapy (generally less than six months), and those patients receiving adjuvant therapy (generally no more than six months after PR record).

Once this cohort will be constituted, SNOMED codes and specific morphology keywords listed in Table 1 will be used to identify non-squamous, squamous, mix and unknown histology of advanced NSCLC. (see Figure A1 in the appendix - flow chart)

Table 1: Morphology codes and keywords

	Morphology code /Keywords in diagnosis field	**Code description**
Non squamous	M-81403	Adenocarcinoma
	M-85503	Adenocarcinoma acinare
	M-83233	Adenocarcinoma misto
	M-82503	Adenocarcinoma bronchiolo alveolare
	M-80123	Carcinoma a grandi cellule
	M-80033	Tumore a grandi cellule maligno
	"*adenocarcin*"	-
	"*grandi cell*"	
Squamous	M-80703	Carcinoma squamocellulare
	M-84303	Carcinoma mucoepidermoide
	"*squamocel*"	-
Mix	M-85603	Adenocarcinoma e Squamo Misto

This cohort will be the source population to describe drug utilisation and the effectiveness of available pharmacological options in patients with a new diagnosis of a primary, not-resected, NSCLC.

2.3 Drug utilisation study

A descriptive, population-based historical prospective cohort study will be performed in order to describe the use of first-line anticancer drugs in patients with a new diagnosis of advanced NSCLC.

2.3.1 First-line identification

ODD, HDR and OED will be combined to identify the first-line treatment sequences. In ODD drugs are coded precisely, and the ATC fifth level is available: this will allow to distinguish between pharmacological treatments as active substances. In HDR/OED only the first date of a chemotherapy procedure is registered and drugs are recorded less specifically (no ATC code is available). Nevertheless, in HDR/OED there are ICD-9CM codes that can be useful to differentiate between undefined standard CT (uCT codes: V58.11, 99.25) and monoclonal antibody-based therapy (uMAB codes: V58.12, 99.28).

All treatment codes (ATC codes for ODD and ICD-9CM codes for HDR/OED) are listed in Table 2.

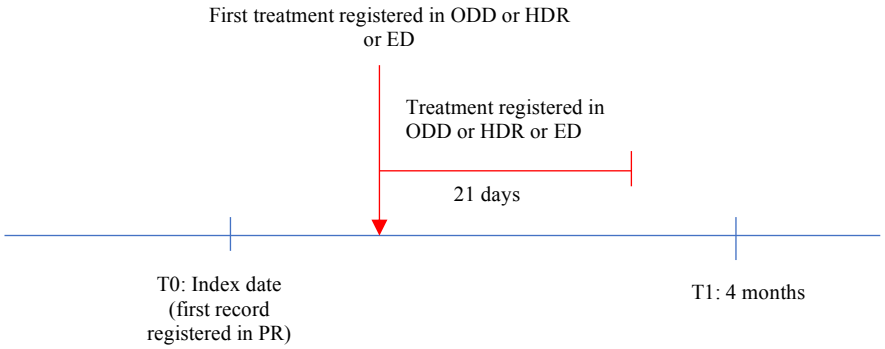
Table 2. **Anti-cancer therapy of interest for the study as available in Italy during the study period**

Treatment class	Treatment sub-class	Active substance	ATC/ICD-9CM codes	Data source
Standard Chemotherapy (CT)	Other chemotherapies	Gemcitabina	L01BC05	ODD
		Etoposide	L01CB01	ODD
		Vinorelbina	L01CA04	ODD
		Pemetrexed	L01BA04	ODD
	Taxanes	Docetaxel	L01CD02	ODD
		Paclitaxel	L01CD01 L01CD03	ODD
	Platin	Cisplatino	L01XA01	ODD
		Carboplatino	L01XA02	ODD
Undefined standard chemotherapy (uCT)	–	–	V58.11, 99.25	HDR, OED
Target therapies (TT)	Anti EGFR	Gefitinib	L01XE02	ODD
		Erlotinib	L01XE03	ODD
		Osimertinib	L01XE35	ODD
		Afatinib	L01XE13	ODD
		Necitumumab	L01XC22	ODD
	Anti-ALK	Crizotinib	L01XE16	ODD
		Ceritinib	L01XE28	ODD
		Alectinib	L01XE36	ODD
	Anti-BRAF	Trametinib	L01XE25	ODD
		Dabrafenib	L01XE23	ODD
	Anti-VEGF	Bevacizumab	L01XC07	ODD
		Nintedanib	L01XE31	ODD
		Ramucirumab	L01XC21	ODD
Immunotherapy (IT)	Anti-PD-1	Nivolumab	L01XC17	ODD
		Pembrolizumab	L01XC18	ODD
	Anti-PDL-1	Atezolizumab	L01XC32	ODD
		Durvalumab	L01XC28	ODD

Monoclonal antibody-based therapy (uMAB)	–	–	V58.12, 99.28	HDR, OED

Nevertheless, the pharmacological treatment of cancer (one treatment cycle) is rarely represented by the administration of a single anticancer drug but more often by the administration of a combination anticancer drugs in a time-lapse of maximum 21 days. Pharmacological treatment sequences, represented by all the records registered in ODD and HDR/OED in the 21 days following the first pharmacological treatment dispensing date, will be identified in the four months after the index date.

Figure 1. First-line sequences identification



The single pharmacological treatments recorded within the time sequence will be associated with:

1) treatment registration date: the time difference between the date of the first drug treatment registered and the date on which the treatment is registered within the sequence. The time difference will be reported in weeks (7 days interval from day 1). Week 1 (days 1-7), week 2 (days 8-14), week 3 (days 15-21). Each treatment will then be associated with the number of the week to which it refers (i.e.1)
2) data source (i.e. ODD, HDR or OED)
3) treatment info (i.e. uCT)

Example:

Sequence i: 1 ODD cisplatin, 1 ODD bevacizumab, 2 HDR uMAB;

Sequence ii: 1 ODD erlotinib, 1 ODD erlotinib;

Sequence iii: 1 HDR uCT, 1 ODD pembrolizumab;

After pharmacological sequences identified, each sequence will be classified as follows:

1) CT: standard chemotherapy

2) TT: target therapy

3) IT: immunotherapy

4) uMAB: undefined monoclonal antibody-based therapy

If possible, a more detailed classification on the basis of pharmacological treatment sub-classes (see Table 2) or active substances will be done. For example, patients receiving TT will be grouped in those receiving anti-ALK, anti EGFR, anti VEGF or anti BRAF pharmacotherapies.

Patients not receiving any pharmacological treatment in the 4 months after index date (not treated patients, NT) will be considered as non-treated.

Two sensitivity analyses will be performed to check time windows.

1) The first-line cancer treatment will be identified within 6 months after the index date.

2) Pharmacological treatment sequences will be identified within 35 days following the first dispensing date.

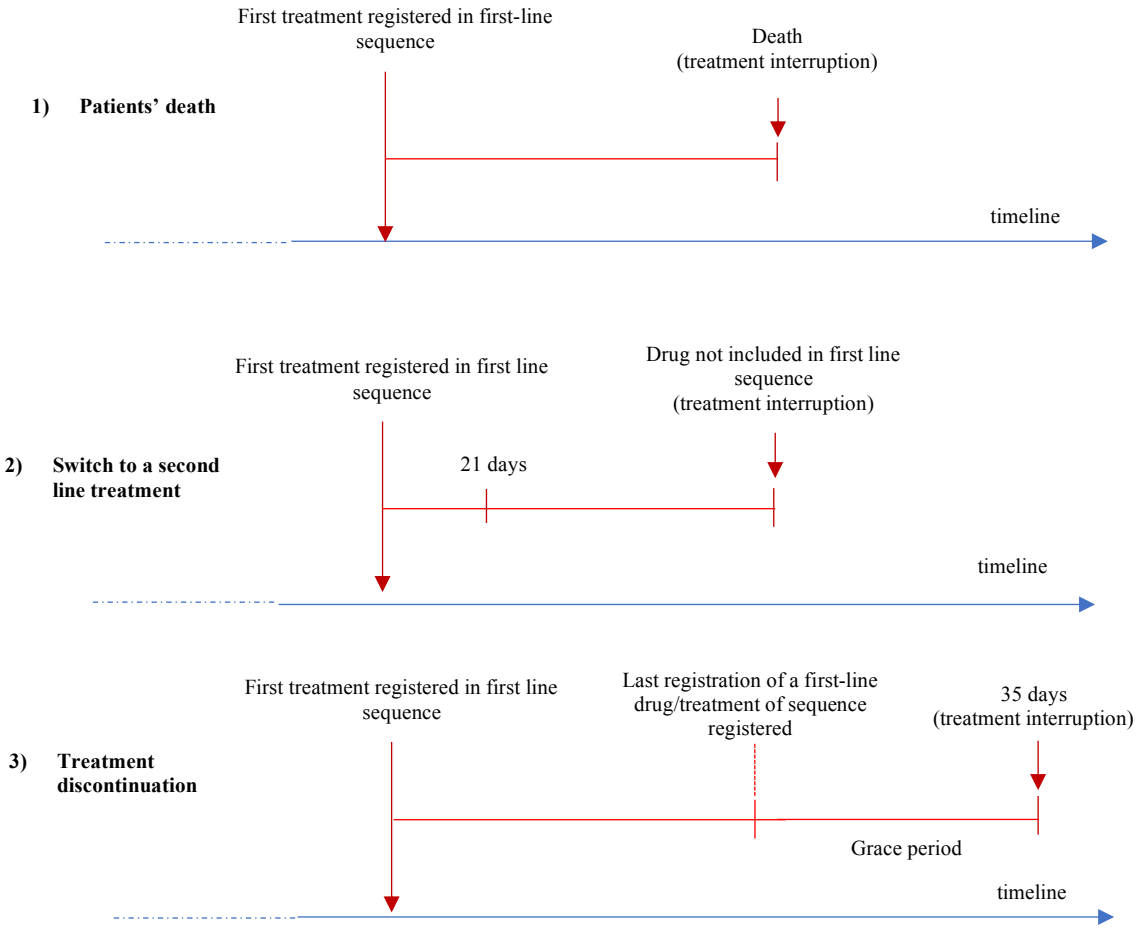
On the basis of the sensitivity analysis results, time windows could be maintained, reduced or extended.

2.3.2 Variables

- *First-line treatment duration*

First-line treatment duration will be calculated as the time difference between the date of first drug/treatment registered in first-line sequence and the putative date in which a treatment interruption will occur.

Figure 2. First-line treatment duration



Treatment interruption will be defined by the first occurring of one of the following conditions (see Figure 2):

1) death of the patient;

2) second-line switching: switch from a first-line to a second-line will occur when a pharmacological treatment, not included in first-line sequence, will be reimbursed > 21 days after the last reimbursement of the first anticancer drug/treatment identified in first-line (the 21 days time window could be extended on the basis of the results of sensitivity analyses in section 2.3.1);

3) treatment discontinuation: discontinuation will occur at 35 days after the last drug/treatment recorded in first-line. The introduction of a grace period was decided because of possible adverse events which may cause the discontinuation of treatment.

- *Concomitant radiotherapy*

HDR/OED will be screened to retrieve concomitant or subsequent radiotherapy (RT). For each selected patient with a first-line pharmacological treatment registered, the eventually concomitant RT will be searched between the date of the first-drug/treatment registered in first-line and treatment interruption (see Table 3 for codes and section 2.3.2 for first-line treatment duration). If the time window of first-line treatment identification will be extended to 6 months according to the sensitivity analysis (section 2.3.1), also RT time window will be modified.

Table 3. Radiotherapy codes

Description	Codes	Data source
Radiotherapy (RT)	92.24.1, 92.24.2, 92.24.3, 92.24.4, 92.29.H, 92.29.M, 92.24.5, 92.24.6, 92.24.7, 92.24.8, 92.47.9, 92.47.8, 92.25.1, V58.0	HDR, OED

- *Diagnostic imaging*

HDR/OED will be screened to retrieve diagnostic imaging. For each selected patient diagnostic imaging codes (see Table 4) will be searched in the six months before the index date.

Table 4. Diagnostic imaging codes

Description	Codes	Data source
Radiography	87.43.1	HDR, OED
	87.43.2	
	87.44.1	
Computed tomography	87.41	
	87.41.1	
	87.42.1	
	87.42.2	
Magnetic resonance imaging	88.92	
	88.92.1	

- *Other variables*

Gender and age will be retrieved from IR. Age will be calculated as the time difference (years) between the patients' birth date and the index date.

Follow-up duration will be also retrieved from IR and will be considered as the time between index date and the date they were last known to be alive or the date patient was lost, whichever comes first.

Mention of molecular tests (EGFR, ALK, BRAF) will be searched in diagnosis field in PR in the three months before or after the index date. A free-text search using the following specific keywords "EGFR", "ALK", "BRAF", and "PD-1" will be performed.

2.3.3 First-line description

Characteristics of patients treated with a first-line will be described using Table A2 in the appendix. According to histology, selected patients (see section 2.2: cohort selection) will be described in terms of sex (n, %), mean age (years, SD), age band distribution (n,%; 18-54; 55-69; 70-84; 85+), median follow-up (months and range), diagnostic imaging (n, %; see section 2.3.2 - Table 4 for codes), median duration of treatment (months, range), mention of molecular testing (n, %) and concomitant radiotherapy (n, %). This results will be presented referring to the overall cohort and to first-line treatment classes: CT, TT, IT and uMAB. Similarly, characteristics of NT patients will be described in table A3.

Distribution of selected NSCLC patients per first-line treatment classes (CT, TT, IT, uMAB) will be described per each year of observation and per histology in the Figure A2 of the appendix.

Finally, reasons of treatment interruption per first-line treatment and per histology will be described in table A5 (see section 2.3.2. – Follow up duration). As for those patients discontinuing treatment, they will be also grouped in those switching or dying or restarting first-line treatment 35 days after the last drug recorded in first-line.

2.4 Survival study

A descriptive, population-based historical prospective cohort study will be performed in order to describe the Overall Survival (OS) of the selected cohort of patients with NSCLC (see section 2.2). Due to the very different biological, molecular and clinical characteristics of patients treated with innovative anticancer drugs approved for NSCLC, it is not reasonable to perform a comparative effectiveness study among pharmacological classes. Nevertheless, if the cohort of patients will be large enough and first-line identification will allow a more detailed classification, a comparison of survival among drugs used for the same indications will be performed (e.g. survival of patients receiving gefitinib vs afatinib, or vs erlotinib; crizotinib vs ceritinib).

Three main analyses will be conducted on the basis of patients' histology characteristics (squamous and non-squamous). For each analyses surviving patients at the time of data collection will be censored at the date they were last known to be alive. The main results of the first two analyses will be three-years survival and five-years survival (whenever follow-up time is available). Moreover, median OS (CI: 95%) will be calculated and the number of events reported.

1) Survival per each year of observation: Selected NSCLC patients will be grouped in ten groups on the basis of the year of their index date. Survival time will be defined as the time between patients index date and death. Results of this analysis will be reported in table A6 and a Kaplan-Meier will be used to represent survival.

2) Survival per first-line treatment received: OS will be described on the basis of patients' first-line treatment classes (see 2.3.1 section; CT, TT, IT, uMAB). If possible, a more detailed analysis on the basis of pharmacological treatment sub-classes will be done. Survival time will be defined as the time between the date of first drug reimbursement in first-line sequence and death. Results

of this analysis will be reported in table A7 and a Kaplan-Meier will be used to represent survival of patients.

3) Survival per each year of observation on the basis of first-line received: Selected NSCLC patients will be grouped in three groups on the basis of first-line treatment classes (CT, NT and other treatment) and the survival will be described on the basis of the year of their index date. Survival time will be defined as the time between the date of first drug reimbursement in first-line sequence and death. Results of this analysis will be represented using the Kaplan-Meier method (see figure A3).

As for these survival analyses, whenever will be possible, results will be also stratified by concomitant radiotherapy and by gender.

2.5 Statistical analysis

Descriptive analyses will be conducted to assess demographic and clinical characteristics of selected NSCLC patients in relation to the histology of NSCLC (squamous, non- squamous, unknown). Continuous variables will be described by means and standard deviation or by median and range while categorical variables will be described by patient counts and percentages. Characteristics of patients according to first-line therapy registered will be compared using chi-square tests or Fisher exact test for categorical variables and ANOVA tests for continuous variables. In order to evaluate if squamous and non-squamous patients have a different profile on the basis of first-line treatment (CT+NT vs other treatments), a logistic regression model will be used and odds ratios and their 95% confidence interval (CI95%) will be calculated (see table A4 of the Appendix). Trend, seasonality and cyclical irregularity about the percentage of NT patients and CT patients will be evaluated using a time series model where tthe percentage of patients will be reported by year quarters. Significance of trend will be assessed from the January 2009 until June 2019 and will be analyzed using Ordinary

Least Square method. As for survival analysis, a COX proportional hazards model will be used to analyze predictors of survival. The Cox model will be controlled for the following factors: age, sex, histology, year of observation, molecular test mention and concomitant use of radiotherapy. Results of the Cox model will be reported as HR with 95%CI. Finally, Kaplan Maier method will be used to describe OS.

3. Data management and processing

Data will be analysed using the statistical software STATA and R version 1.2.5001

4. Limitation of study methods

HDR and OED data source used for this study cannot track the use of active substances precisely, nor the dates of drug administrations after the first procedure of uMAB or uCT. This may affect the identification of treatment lines and treatment duration.

5. Ethical considerations

The study was approoved by the ethical committee of the Azienda Ospedaliera Universitaria Senese (Comitato Etico Regionale per la Sperimentazione Clinica della Regione Toscana; Sezione: AREA VASTA SUD EST).

6. Dissemination and communication strategy

A study report summarizing all main results will be produced and shared with data partners before December 2020. The findings from this study will be submitted to a peer-review international journal in the first semester of 2021.

7. Bibliography

1. World Health Organization. Cancer [Internet]. [cited 2019 Oct 4]. Available from: https://www.who.int/news-room/fact-sheets/detail/cancer
2. World Health Organization. PRGlobocanFinal.pdf [Internet]. [cited 2019 May 22]. Available from: https://www.who.int/cancer/PRGlobocanFinal.pdf
3. AIOM. Linee guida Associazione Italiana di Oncologia Medica - 2018: Tumore al polmone [Internet]. Available from: https://www.aiom.it/wp-

content/uploads/2018/11/2018_LG_AIOM_Polmone.pdf
4. NCCN. nscl.pdf [Internet]. [cited 2019 Mar 25]. Available from: https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf
5. Ho C, Tong KM, Ramsden K, Ionescu DN, Laskin J. Histologic classification of non-small-cell lung cancer over time: reducing the rates of not-otherwise-specified. Curr Oncol. 2015 Jun;22(3):e164–70.
6. Toschi L, Rossi S, Finocchiaro G, Santoro A. Non-small cell lung cancer treatment (r)evolution: ten years of advances and more to come. Ecancermedicalscience. 2017;11:787.
7. Fahey T. Applying the results of clinical trials to patients to general practice: perceived problems, strengths, assumptions, and challenges for the future. Br J Gen Pract. 1998 Apr;48(429):1173–8.
8. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. THE ROLE OF INFORMATICS IN PROMOTING PATIENT-CENTERED CARE. Cancer J. 2011;17(4):211–8.
9. Houser SH, Colquitt S, Clements K, Hart-Hester S. The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama. Perspect Health Inf Manag [Internet]. 2012 Apr 1 [cited 2018 Nov 29];9(Spring). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3329210/
10. Hiatt RA, Tai CG, Blayney DW, Deapen D, Hogarth M, Kizer KW, et al. Leveraging state cancer registries to measure and improve the quality of cancer care: a potential strategy for California and beyond. J Natl Cancer Inst. 2015 Mar 11;107(5).

8. Appendix

Table A1: Algorithm codes and keywords

Algorithm: (Morfo_#1.0 AND NOT Morfo_K_neuroend) AND (Topog_#1.0 AND NOT **Topog_Upper_airways**)

Algorithm 1.0 component Acronym	Description	Tumoral characteristic identified	Terminology	Codes/free text words	Use
Morpho_#1.0	NSCLC codes and free text words	Morophologic	Snomed e testo libero	M-81403 OR M-80703 OR M-82463 OR M-85603 OR M-83233 OR-85503 OR M-84303 OR M-82503 OR M-80033 OR M-80123 OR M-80463 OR (M-8* AND (Stringhe_in_diagnosi:"*non microcit*" OR "*adenocarcin*" "*squamocell*" OR "*grandi cell*"))	Inclusion
Morpho_K_neuroend	Neuroendocrin carcinoma	Morphologic	Snomed	M-82463	Exclusion
Topog_#1.0	Chest topography	Topographic	Snomed e testo libero	T-2* OR Stringhe in diagnosi: "POLM*" OR "BRONCH*"	Inclusion

Topog_Upper_airways	Upper airways	Topographic	Snomed	T-21* OR T-22* OR T-23* OR T-24* OR T-25*	Exclusion

Figure A1. Flow chart exemple

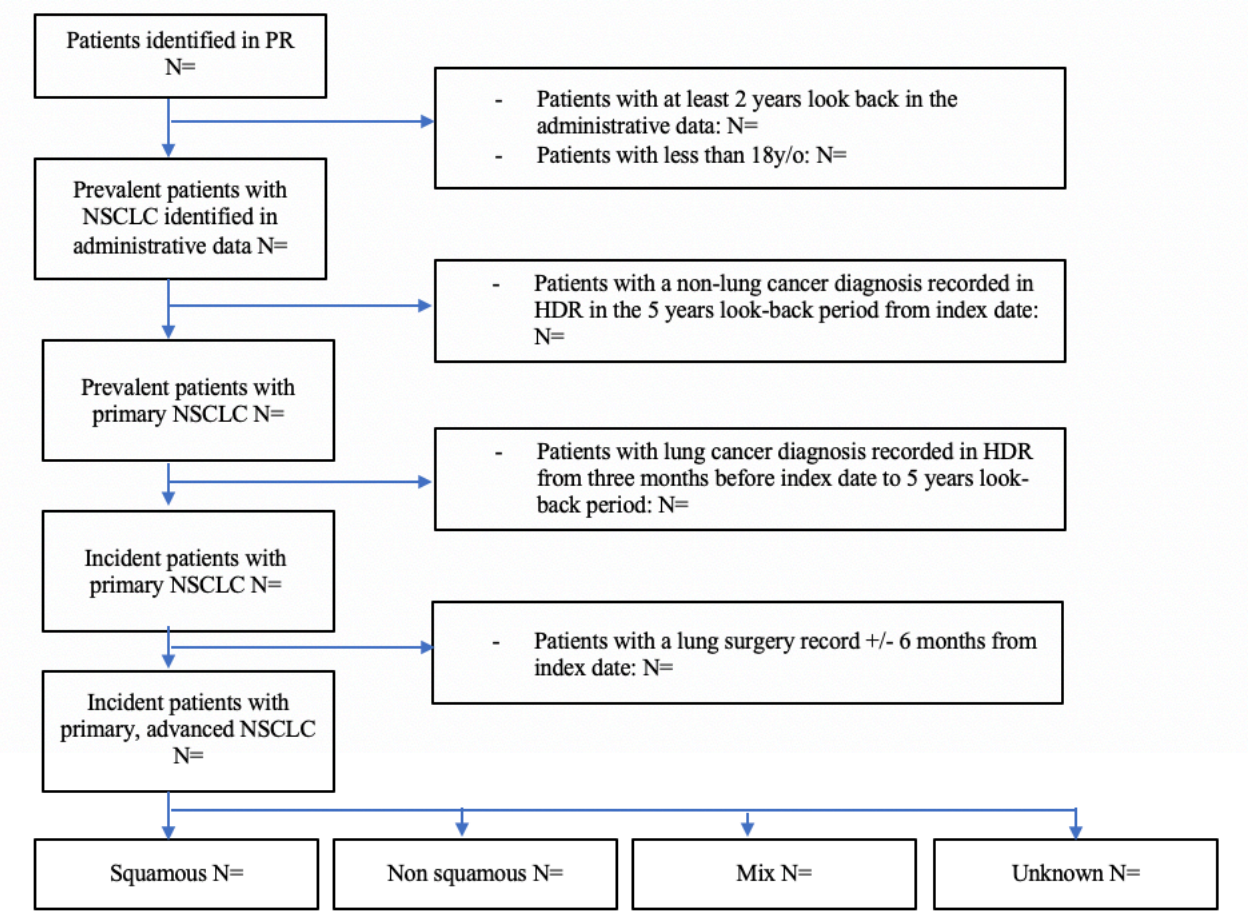


Table A2. Characteristics of patients on the basis of first-line received and histology

	Squamous						Non squamous						Unknown					
	CT N, (%)	TT N, (%)	IT N, (%)	uMAB N, (%)	P value	Overall N	CT N, (%)	TT N, (%)	IT N, (%)	uMAB N, (%)	P value	Overall N	CT N, (%)	TT N, (%)	IT N, (%)	uMAB N, (%)	P value	Overall N
Female																		
Mean age (y, SD)																		
Age (n, %)																		
18-54																		
55-69																		
70-84																		
85+																		
Months of follow-up from ID																		
Median (Range)																		
First-line duration, days																		
Median (range)																		
Diagnostic imaging																		
Contrast/not contrast TC																		
Radiography																		

Pet therapy																		
Molecular test mention																		
EGFR (yes)																		
ALK (yes),																		
BRAF (yes)																		
Concomitant Radiotherapy (n, %)																		

Table A3. Characteristics of NT patients on the basis of histology

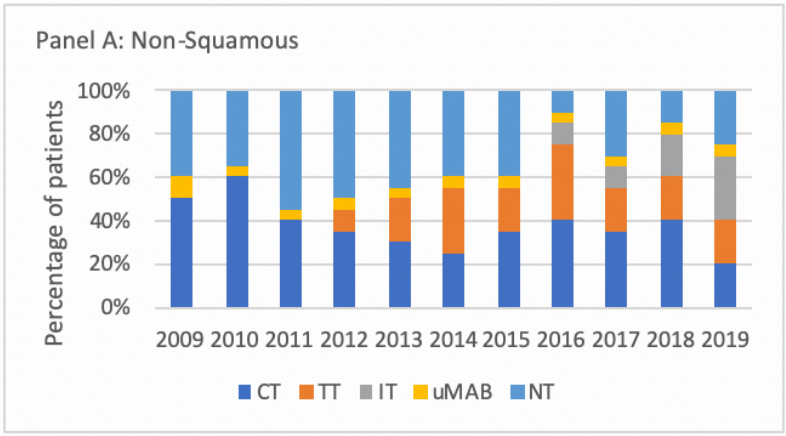
	Not treated patients		
	Squamous	Not squamous	Unknown
Female			
Mean age (y, SD)			
Age (n, %)			
18-54			
55-69			
70-84			
85+			

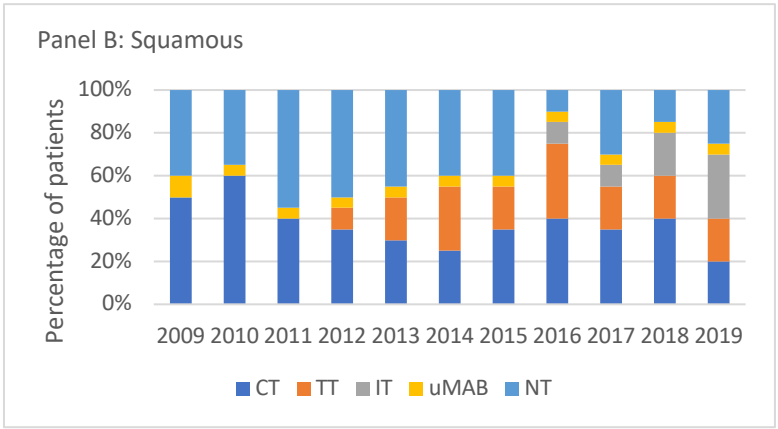
Months of follow-up from ID			
Median (Range)			
Diagnostic imaging			
Contrast/not contrast TC			
Radiography			
Pet therapy			
Death within 6 months after index date			

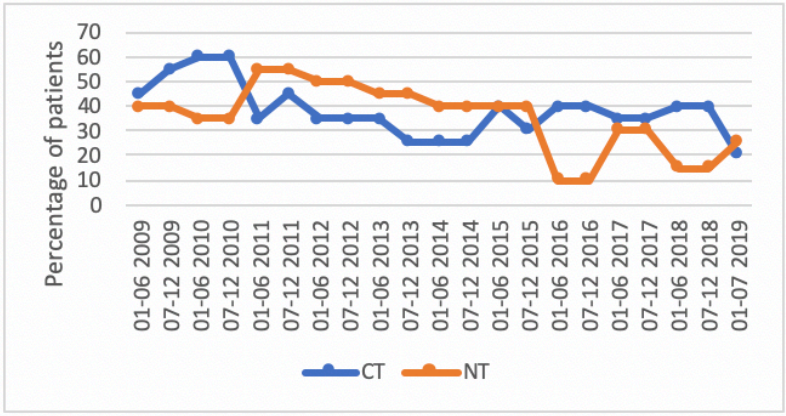
Table A4. Different profile of patients on the basis of first-line treatment and histology

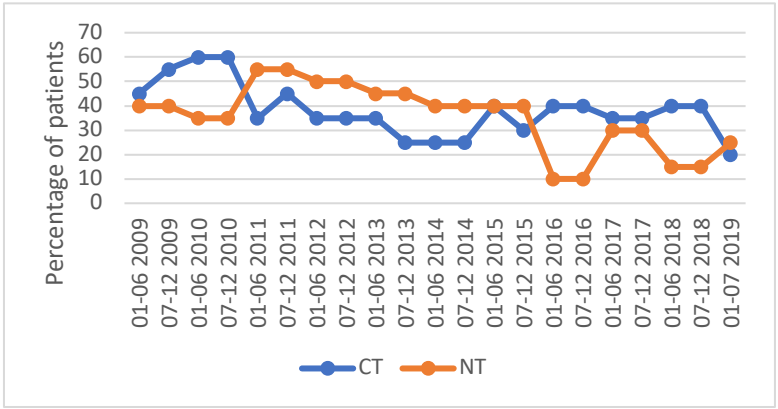
	Squamous	Non squamous
	CT + NT vs other treatments	CT + NT vs other treatments
Female (0,male; 1, female)	Odds ratio (CI 95%)	
Age (0,<70; 1 ≥70)		
Molecular test mention (0, no; 1, yes)		
Concomitant Radiotherapy (0, no; 1, yes)		

Figure A2. Distribution of first-line pharmacological classes per year (2009-2019) and time series analysis for CT and NT patients. Panel A: Non-Squamous patients; Panel B: squamous patients;









	Trend	Trend P value	Seasonality	Cyclical irregularity
CT				
NT				

	Trend	Trend P value	Seasonality	Cyclical irregularity
CT				
NT				

Table A5. Reasons for first-line treatment interruption

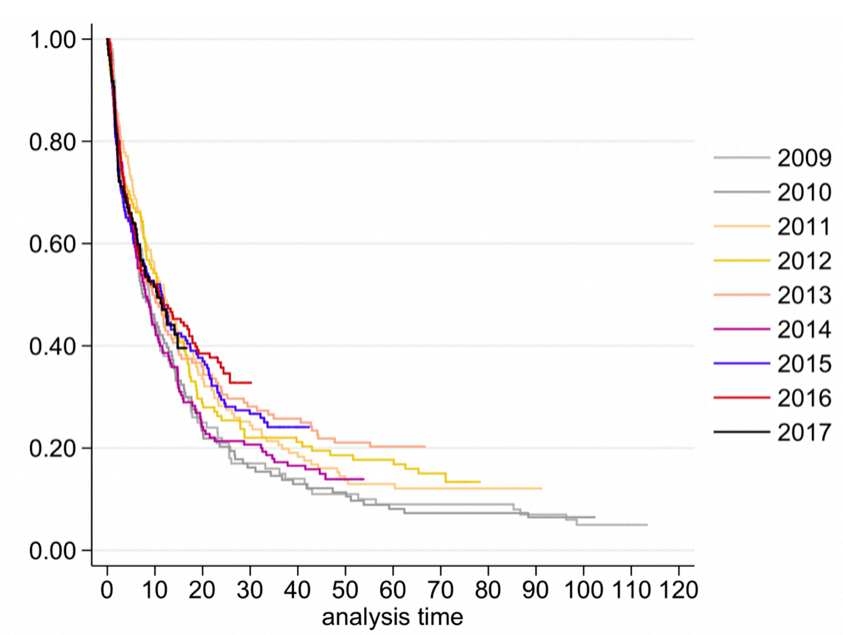
	Squamous					Non squamous				
	Death n (%)	Switch to second line n (%)	Treatment discontinuation			Death n (%)	Switch to second line n (%)	Treatment interruption		
			Switch to second line	First-line restart	Death			Switch to second line	First-line restart	Death
CT n, (%)										
Taxanes										
Platin										
Other chemotherapies										
TT n, (%)										
Anti ALK										
Anti BRAF										
Anti EGFR										
Anti VEGF										
IT n, (%)										
Anti PD-1										
Anti PD-L1										
uMAB n, (%)										

Table A6. Survival of selected NSCLC patients per date of entry in the cohort

	N	Three year survival (%)	Five year survival (%)	Median OS	Range	Events n (%)
Squamous						
2009						
2010						
2011						
2012						
2013						
2014						
2015						
2016						
2017						
2018						
2019						
Overall						
Non Squamous						
2009						
2010						
2011						
2012						
2013						

2014						
2015						
2016						
2017						
2018						
2019						
Overall						

Panel A: Kaplan-Meier per year of the study: squamous patients



Panel B: Kaplan-Meier per year of the study: non-squamous patients

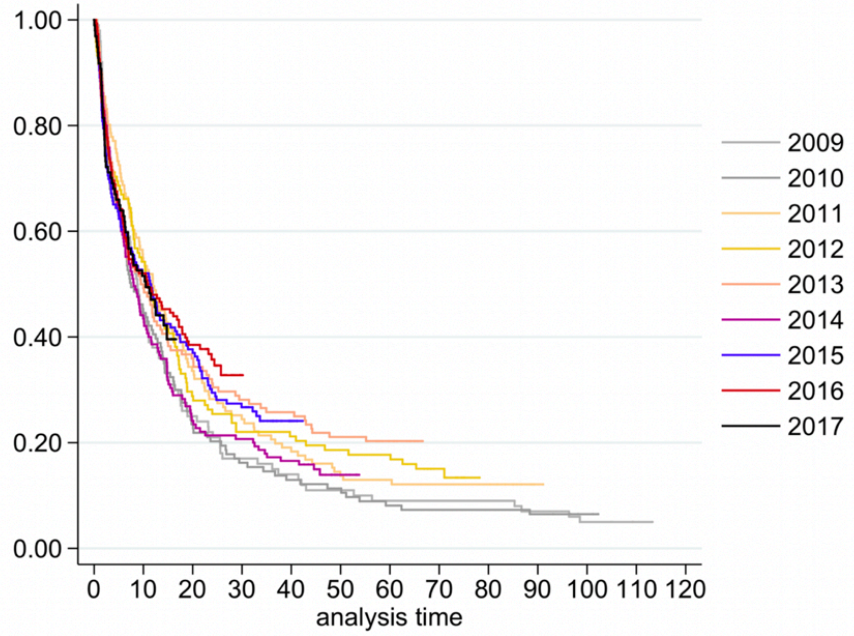
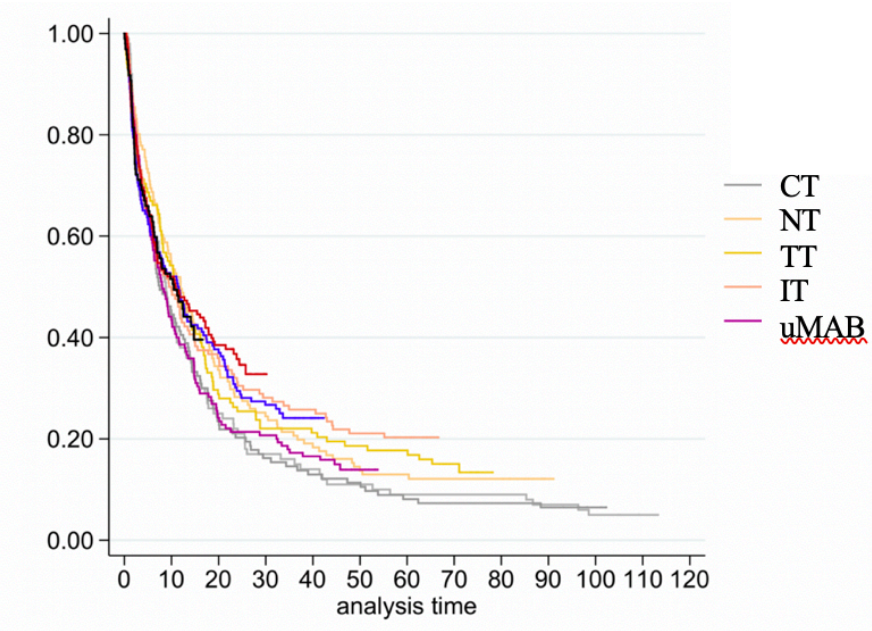


Table A7. Survival of selected NSCLC patients per first-line received

	N	Three year survival	Five year survival	Median OS	Range	Events n (%)
Squamous						
CT						
TT						
IT						
uMAB						
NT						
Non-Squamous						
CT						
TT						
IT						
uMAB						
NT						

Panel A: Kaplan-Meier per first-line received: squamous patients



Panel B: Kaplan-Meier per first-line received: non-squamous patients

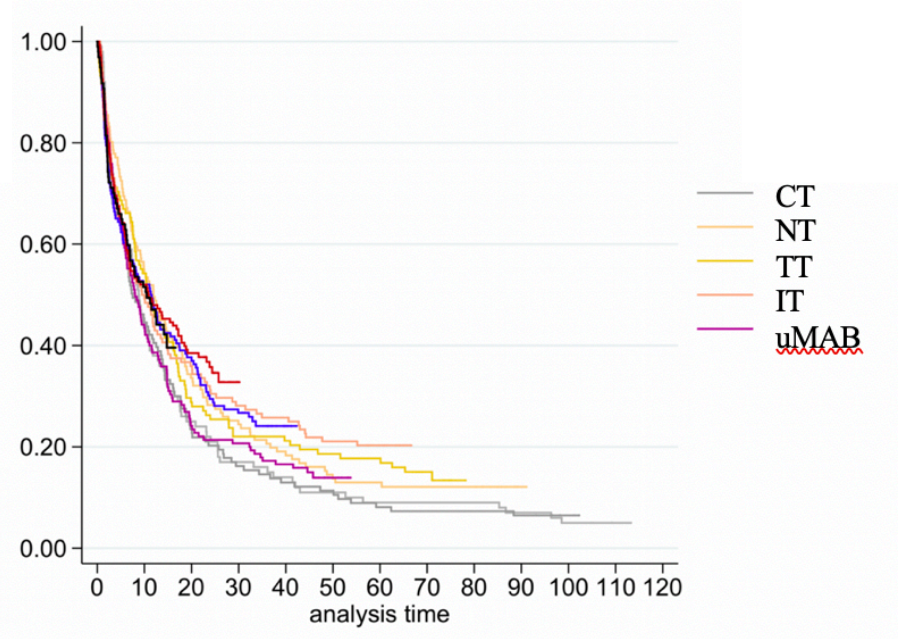
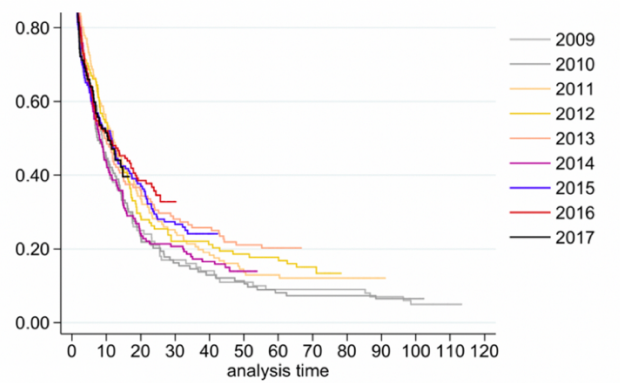
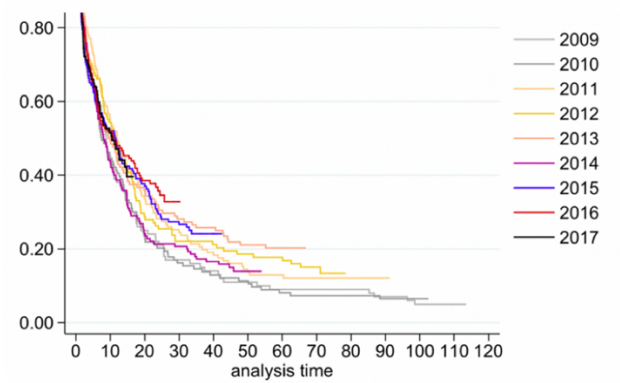


Figure A3. Survival of patients on the basis of first line received per year of the study and histology

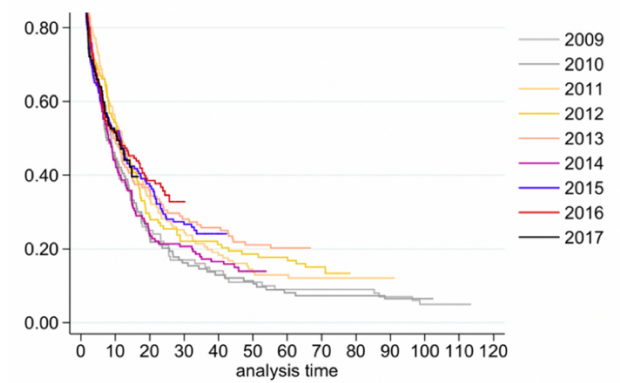
Panel A: Squamous patients treated with chemotherapy (CT)



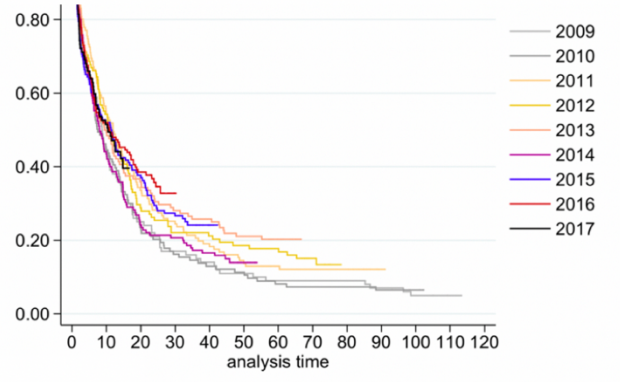
Panel B: Squamous patients treated with other treatments (IT, TT, uMAB)



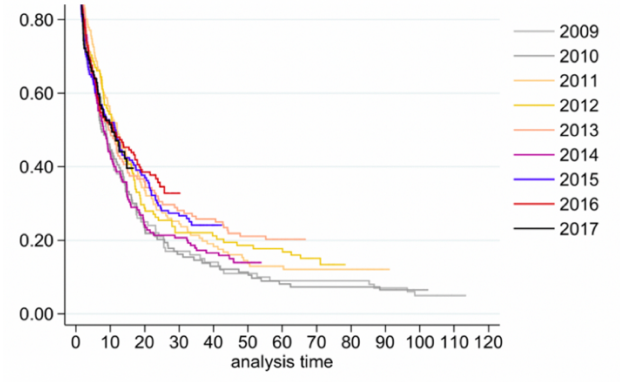
Panel C: Squamous patients not treated (NT)



Panel D: Non-squamous patients treated with chemotherapy (CT)



Panel E: Non-squamous patients treated with with other treatments (IT, TT, uMAB)



Panel F: Non-squamous patients not treated (NT)

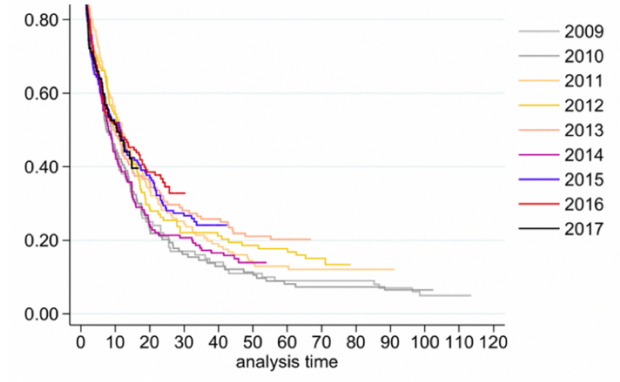


Table A8. Predictors of survival

Covariates	Coefficient	Standard error	P value	HR	Ci Lower	CI higher
Gender (1, F; 0, M)						
Age (1,<70; 0, ≥70)						
Year of observation						
Histology (1 non squamous; 0, squamous)						
Molecular test mention (1 yes; 0, no)						
Concomitant Radiotherapy (1 yes; 0, no)						