Key Performance Indicators for the assessment of diagnostic and therapeutic pathway of NSCLC patients: a multicenter study (KIND NSCLC)

Protocol Code: IRST 162.13

Date and Version: 19/09/2019 – Version 1.0

Protocol approval and Investigator agreement

Key Performance Indicators for the assessment of diagnostic and therapeutic pathway of NSCLC patients: a multicenter study (KIND NSCLC)

The undersigned agree and confirm that:

the following protocol has been agreed and accepted and the Chief Investigator agrees to conduct the trial in compliance with the approved protocol and will adhere to the principles outlined in ICH GCP guidelines, Sponsor/Promoter SOP's and other regulatory requirements as amended.

The confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the clinical investigation without the prior written consent of the Sponsor/Promoter.

The findings of the study will be made publically available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and any discrepancies from the study as planned in this protocol will be explained.

Chief Investigator	Signature	Date
Trial Statistician	Signature	Date

By signing this document I am confirming that I have read the protocol for the above study and I agree to conduct the study in compliance with the protocol and ICH GCP.

Mattia Altini

Principal Investigator _____ Signature _____ Date

SUMMARY

Title	**Key Performance Indicators for the assessment of diagnostic and therapeutic pathway of NSCLC patients: a multicenter study (KIND NSCLC)**
Protocol Code	**IRST 162.13**
Acronym	**KIND NSCLC**
Rationale	Lung cancer is the most common cancer diagnosed globally and the leading cause of cancer death worldwide. Non-Small Cell Lung Cancer (NSCLC) accounts for approximately 85% to 90% of all lung cancer with 5-year survival rate of 22.1%. Approximately 79% of NSCLC patients are diagnosed at advanced stage (IIIB or IV) NSCLC. Furthermore, recurrence rates after complete surgical resection range from 30% to 75% among patients initially diagnosed with stage I, II, or III, with the majority of patients who experiencing metastatic recurrence. However, significant advanced in testing and treatment have been made in recent years with the potential to positively impact lung cancer outcomes. These include rapid positive progress in the use of

	immunotherapy, which has become an important therapeutic alternative influencing the clinical approach and becoming the first choice of treatment in many cases of advanced NSCLC. Moreover treatment of NSCLC is highly complex, resulting in wide variations in the patterns of care. An appropriate care of patients with NSCLC requires managerial skills merged with clinical knowledge and experience, and coordination among multidisciplinary specialists. Nevertheless, to date the management of patients with NSCLC is characterized by a sequence of medical investigations and visits of various specialists, whose timeframes are often not short, whose assessment is fragmentary, poorly coordinated and often followed by inappropriate decisions. The organizational framework described above makes the analysis of NSCLC particularly interesting. Currently, there is an increasing interest in the use of administrative healthcare databases in clinical and health services research as they provide timely and easy access to a large source of information regarding subjects in a defined geographical area. Administrative data have been widely used in different types of epidemiologic research, health services, and outcomes cancer research. The investigated population can be identified through disease-specific codes, prescriptions and disease-specific procedure codes among administrative databases. In light of the rapid change of NSCLC therapeutic options, we are proposing a non-interventional study to generate real-life practice data on NSCLC patient from administrative databases. The focus of this study is to allow a deeper knowledge about the quality of NSCLC care pathway in three hospital of Emilia-Romagna region, monitoring the routinely clinical practice patterns through the measurement of key performance indicators (KPIs). Data will be entirely collected from administrative databases.

	This study can help to define best clinical practice and to understand the management of resources for NSCLC patients, through a fast and cheap approach, much reflective of the real world.
Study Design	This is a multicenter, observational study involving a retrospective collection of data. A total of potential 16 key performance indicators (KPIs) had been developed from a panel of experts (clinicians, IT experts, etc..) to investigate the appropriateness of care in NSCLC patients, with a special focus on the use of immunotherapy. The eligible population and data will be gathered retrospectively using an algorithm. Administrative databases will be used as unique resource: • to identify target population; • to collect patient's data with which measure KPIs.
Study duration	Enrollment period: from January 2017 to December2017 Total duration of the study: 1 year from approval date of study by Ethics Committee.
Study Promoter	IRST
Study Centers	This study is an Italian Multi-Center study. The participating sites are: • Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST-IRCCS); • Department of Oncology and Hematology, Policlinico of Modena; • Department of Oncology, Arcispedale S. Maria Nuova Reggio Emilia.
Objectives	The project aims to measure a set of potential indicators (KPIs) derived from administrative databases in order to investigate the appropriateness and quality of NSCLC care among participating sites. The secondary aim is to develop and validate an algorithm to identify the eligible population of the study from SDO (hospital discharge

	card) against to a clinical database (only at IRST site).
Number of Subjects	About 400 patients are expected globally.
Diagnosis and Main Inclusion Criteria	1. Patients identified through the algorithm as described at paragraph 4.1; 2. Male or female; 3. At least 18 years at date of first index hospitalization
Procedures and data collection	Patient information will be obtained only from administrative databases both for out-patient and in-patient settings.
Statistical Methodology	For primary objective, for each indicator, the proportion of patients who receive the procedure, in the defined time window, among those eligible, will be calculated. Descriptive statistics as arithmetic mean and standard deviation (SD) as well as median and range value (maximum-minimum) will be presented for KPI related to average time. Two-sided 95% confidence interval (95% CI) will be reported as appropriate and specified for KPI expressed as proportion. For secondary objective, the validity of algorithm will evaluated by individual matching among cases retrieved from administrative database and medical chart obtained from IRST. Sensitivity, positive predictive value as well as specificity and negative predictive value will be calculated, and exact (Clopper-Pearson) 95% confidence limits for binomial distribution will be applied.

ABBREVIATIONS

ADI	Assistenza domiciliare integrata
AFT	Assistenza farmaceutica territoriale
ASA	Assistenza ambulatoriale
CC	Coordinating Center
CI	Chief Investigator
CRF	Case Report Form
FED	Farmaci ad erogazione diretta
GCP	Good Clinical Practice
ICF	Informed Consent Form
ICH	International Conference of Harmonization
ICP	Integrated Care Pathway
IEC	Independent Ethics Committee
KPI	Key Performance Indicator
mNSCLC	Metastatic Non Small Cell Lung Cancer
NSCLC	Non Small Cell Lung Cancer
PI	Principal Investigator
REM	Registro mortalità regione Emilia Romagna
ASA	Assistenza ambulatoriale
SDHS-Hospice	Scheda di dimissione Hospice
SDO	Scheda di dimissione ospedaliera

TABLE OF CONTENTS

1. BACKGROUND AND RATIONALE

Lung cancer is the most common cancer diagnosed globally and the leading cause of cancer death worldwide [1]. Non-Small Cell Lung Cancer (NSCLC) accounts for approximately 85% to 90% of all lung cancer [2,3] with 5-year survival rate of 22.1% [4]. Approximately seventy-nine percent (79%) of NSCLC patients are diagnosed with advanced (stage IIIB or IV) NSCLC [4]. Furthermore, recurrence rates after complete surgical resection range from 30% to 75% among patients initially diagnosed with stage I, II, or III, with the majority experiencing metastatic recurrence [5–7].

However, significant advanced in testing and treatment, which have the potential to positively impact lung cancer outcomes have been made in recent years. These include rapid positive progress in the use of immunotherapy, which has become an important therapeutic alternative influencing the clinical approach and becoming one of the main choices of treatment in many cases of advanced NSCLC. Moreover, treatment of NSCLC is highly complex, resulting in wide variations in the patterns of care. An appropriate care of patients with NSCLC requires managerial skills merged with clinical knowledge and experience, and coordination among multidisciplinary specialists [8]. Therefore, to date the management of patients with NSCLC is characterized by a sequence of medical investigations and visits of various specialists, whose timeframes are often not short, and the risk of fragmentary assessments, poorly coordinated and often followed by inappropriate decisions, is high. The organizational framework described above makes the analysis of NSCLC care pathway particularly interesting. This objective could be pursued measuring some key performance indicators (KPIs) of care pathway in NSCLC. The measure of KPIs in NSCLC is not an easy task. Available literature on KPIs in NSCLC is relatively limited and there is no universal agreement on quality of care indicators. However, the employment of quality indicators is an important tool that can help to improve patient's care and outcomes, promote patients safety, increase patient satisfaction, and to optimize the use of resources.

Currently, there is an increasing interest in the use of administrative healthcare databases in clinical and health services research as they provide timely and easy access to a large source of information regarding subjects in a defined geographical area. Administrative data have been widely used in different types of epidemiologic research, health services, and outcomes cancer research[9–11] .

Healthcare administrative databases were also used to identify study population through disease-specific codes, prescriptions or specific procedure codes.

In light of the recent introduction of immunotherapies, we are proposing a non-interventional study to generate real-life practice data on NSCLC care pathway. The focus of this study is to allow a deeper knowledge about the quality of NSCLC care in three hospitals of Emilia-Romagna region, monitoring the routinely clinical practice through the measurement of key performance indicators (KPIs). Data will be entirely collected from administrative databases.

This study can help to define best clinical practice and to understand the management of resources for NSCLC patients, through a fast and cheap approach, much reflective of the real world.

2. OBJECTIVES

2.1 Primary Objective

To measure a set of potential indicators (KPIs) derived from administrative database in order to investigate the appropriateness and quality of NSCLC care among participating sites.

2.2 Secondary Objectives

The secondary aim is to develop and validate an algorithm to identify the eligible population of the study from administrative databases against to a clinical database (only at IRST site. Algorithm will be used to identify eligible patients and measure KPIs in the other two participating sites only through administrative database.

3. STUDY DESIGN

This is a multi-center, observational study involving retrospective collection of NSCLC patients information. All consecutive patients who had a newly diagnosis of NSCLC in 2017 from healthcare administrative database and received at least one anti-cancer drug treatment between January 2017 and December 2017, identified through the proposed algorithm, will be considered. All data needed for KPI calculation, even if they fall outside January-December 2017 period, will be collected. The end of data collection will be defined as patient death or 30 June 2018, whichever came first.

A set of potential KPIs had been developed from a panel of experts (clinicians, IT experts etc..) to investigate the appropriateness of activities within NSCLC care pathway, with a special focus on the use of immunotherapy.

The eligible population and data will be gathered retrospectively through an algorithm from administrative databases (SDO, REM, ASA, FED, AFT, ADI, SDHS-hospice).

Administrative data will be used as unique resource:

- to identify patients cohort;
- to measure KPIs along care pathway ;

4. STUDY POPULATION

4.1 Definition of study groups

An algorithm will be developed to reliably identify eligible subjects from administrative database SDO. The algorithm will apply specific criteria to patients collected in administrative databases from January 2017 to December 2017. It will be developed as the combination of ICD-9-CM codes with other criteria. Selection of codes and criteria was based on a discussion between senior Oncologists with expertise in lung cancer and the outcome research working group at IRST.

A first data extraction from the administrative database will be performed to identified potential study subjects. These data are related to patients with at least one hospitalization due to diagnosis of malignant tumors of trachea, bronchi and lungs between January and December 2017 using the International Classification of Diseases, 9th EDITION, (ICD-9-CM codes: 162* [12,13] in all position of diagnosis codes).

The following criteria will be applied to obtained patients, to exclude prevalence NSCLC cases (incident cases in 2014-2016 period), patients with mesothelioma (cod 163*) and SCLC patients in order to obtain the correct population.

- Only NSCLC incident cases in 2017 will be considered, excluding prevalence cases that had the same diagnosis in the 3 years before the period of interest (2014-2016). Patients with at least one hospitalization due to diagnosis of malignant tumors of trachea, bronchi and lung according to ICD-9CM 162* in the 3 years prior to the first index hospitalization on the year of evaluation (2017) will be excluded

- Patients with at least one hospitalization due to malignant tumor different than tumors of trachea, bronchi and lungs (ICD-9CM 140*-161* or 163*- 195* or 200* - 208* or V.10* excluding V10.11 and V10.12), in the 3 years previous 2017 should be excluded.
- All patients with at least one ATC Code L01CB01
- Etoposide (VP-16) therapy administration in the 180 days following the first hospitalization index will be excluded.

After criteria application, the final population shall include only NSCLC patients.

4.2 Validation of algorithm (only at IRST site)

Algorithm will be validated against the electronic medical record review. Validation will be performed applying the criteria described in 4.1 paragraph on electronic medical record of patients treated in IRST. The population obtained from administrative database and identified by the algorithm as NSCLC or non-NSCLC will be then compared with the corresponding cases identified in the medical chart, and performance of the algorithm will be evaluated. The corresponding medical chart of the selected sample cases will be consulted only for validation purposes. The definitions of algorithm will be subsequently revised to maximize the agreement and delete discrepancies with the patient medical chart to maximize the accuracy of the algorithm.

4.3 Inclusion Criteria

4. Patients identified through the algorithm as described at paragraph 4.1;
5. Male or female;
6. At least 18 years at date of first index hospitalization;

4.4 Exclusion criteria

The subjects who lives outside the regional territory of competence will be excluded from analysis due to the difficulty in obtaining the medical data.

5 STUDY PROCEDURES AND DATA COLLECTION

To evaluate primary objective of the study, patient information will be obtained only from regional administrative databases both for out-patient and in-patient settings. These data will be related only to the year 2017.

5.1 Clinical variables of interest for the study

The only source for clinical information will be administrative databases in all its parts.

We will focus primarily on the following set of clinical data:

- Biopsy at diagnosis;
- Presence of metastatic NSCLC at diagnosis;
- Neoadjuvant treatment before lung cancer surgical resection;
- Lung cancer main surgical resection;
- Length of stay in hospital after lung major surgery;
- Main information regarding anti-cancer drug treatments at the first diagnosis and for subsequent lines
- Multidisciplinary team (MDT) meeting discussion;
- Radiotherapy (RT) in the end of life;
- Death;
- Integrated home care utilization;
- Hospice care utilization.

5.2 Key performance indicators for monitoring care pathway in NSCLC patients

The following key performance indicators (KPIs) will be measured to investigate the appropriateness of NSCLC patient care pathways, with a special focus on the use of immunotherapy.

BOZZA - KPI -	
1	**Percentage of patients who performed biopsy within 7 days from the date of examination prescription** NUM: Number of patients who performed biopsy within 7 days from prescription; DEN: Total number of patients identified by the algorithm who performed biopsy;
2	**Percentage of mNSCLC patients at diagnosis** NUM: Number of patients with metastases (presence of one of the following codes (197*,198*,199*) in all position of diagnosis codes for index hospitalization); DEN: Total number of patients identified by the algorithm;
3	**Percentage of major surgical resections within the year of analysis.** NUM: Number of patients who underwent to lung cancer major surgical resection; DEN: Total number of patients identified by the algorithm;

<table>
<tr><td></td><td>To identify major surgical procedures for lung, the following codes of ICD-9-CM classification will be identified:
<table>
<tr><th>ICD-9-CM coding related to major surgery on lung</th></tr>
<tr><td>32.01 ENDOSCOPIC EXCISION OR DESTRUCTION OF LESION OR TISSUE OF BRONCHUS</td></tr>
<tr><td>32.29 OTHER LOCAL EXCISION OR DESTRUCTION OF LESION OR TISSUE OF LUNG</td></tr>
<tr><td>32.3 SEGMENTAL RESECTION OF LUNG</td></tr>
<tr><td>32.4 LOBECTOMY OF LUNG</td></tr>
<tr><td>32.5 COMPLETE PNEUMONECTOMY</td></tr>
<tr><td>34.6 SCARIFICATION OF PLEURA</td></tr>
</table></td></tr>
<tr><td>4</td><td>Percentage of surgical patients on the total number of patients with exploratory thoracotomy in inpatient setting.
NUM: Total of patients with only exploratory thoracotomy procedure (Cod 34.02);
DEN: Number of surgical patients;
For major surgery, ICD-9-CM codes were selected as reported for KPI 3.
Surgical patients defined as patients who receive any surgical procedure (major and/or explorative)</td></tr>
<tr><td>5</td><td>Percentage of patients undergoing neoadjuvant therapy (CT or CT+RT within 6 months prior to the first major resection date)
NUM: Number of patients who underwent induction treatment as CT or CT+RT in the 6 months prior to surgery;
DEN: Total number of surgical patients;
The algorithm will not be able to identify patients who received neoadjuvant therapy in clinical trials.
For major surgery, ICD-9-CM codes were selected as reported for KPI 3.</td></tr>
<tr><td>6</td><td>30-day mortality rate from major surgery
NUM : Number of patients who died within 30 days from major surgery;
DEN: Total number of surgical patients;
For major surgery, ICD-9-CM codes were selected as reported for KPI 3.</td></tr>
<tr><td>7</td><td>Percentage of patients discussed in the multidisciplinary team (MDT) meeting at least once in the year of incidence
NUM: Number of patients discussed at least one time in MDT meetings;
DEN: Total number of patients identified by the algorithm;</td></tr>
<tr><td>8</td><td>Percentage of patients who performed imaging exams such as MRI, TC and PET within 7 days from MDT meeting
NUM: Number of patients who performed imaging exams within 7 days from MDT meeting;
DEN : Total number of patients discussed in MDT meeting;</td></tr>
<tr><td>9</td><td>Average and range of length of stay calculated as the mean difference among date of discharge and date of admission.</td></tr>
<tr><td>10</td><td>Average time between biopsy and the first administration of systematic anti-cancer drug (excluded neoadjuvant patients of KPI 5) calculated as the average difference among biopsy and the first date of administration of chemotherapy.
For patients enrolled in experimental clinical trial will not be possible to retrieve information on therapies from administrative database.</td></tr>
</table>

11	**KPIs on THERAPIES** For patients enrolled in experimental clinical trial will not be possible to retrieve information on therapies from administrative database. **a) Percentage of patients treated with immunotherapies** NUM: Patients with at least one administration of immunotherapy According to Anatomical Therapeutic Chemical Classification System ATC, the following codes will be investigated: L01XC17 nivolumab L01XC18 pembrolizumab L01XC32 atezolizumab DEN: Total number of patients identified by the algorithm who underwent anti-cancer drug treatment for NSCLC, as single or double agent chemotherapy, target therapy etc..(excluded neoadjuvant patients of KPI 5) **b) Percentage of patients who underwent immunotherapy for more than 6 months after first dose.** NUM: Number of patients who underwent immunotherapy more than 6 months after the first administration; DEN: Total number of patients identified by the algorithm who underwent immunotherapy;
12	**Percentage of patients who underwent anti-cancer drug treatment in the last 30 days of life** NUM: Number of patients who underwent anti-cancer drug treatment in the last 30 days of life; DEN: Total number of patients identified by the algorithm who deceased
13	**Percentage of patients who underwent the first administration of systematic anti-cancer drug within 30 days from discussed in MDT meeting** NUM: Number of patients who underwent first systematic anti-cancer drug administration within 30 days from MDT; DEN: Total number of patients patients identified by the algorithm discussed in MDT meeting;
14	**Percentage of patients who underwent RT in the last 30 days of life (excluded palliative treatments single fraction)** NUM: Number of patients who performed RT in the last 30 days of life (excluded palliative treatments)**;** DEN: Total number of patients identified by the algorithm who deceased;
15	**Percentage of patients who underwent at least one Integrated Home Care (IHC)** NUM: Number of patients who underwent IHC service; DEN: Total number of patients identified by the algorithm;
16	**Percentage of patients who underwent at least one hospice care** NUM: Number of patients who underwent hospice care; DEN: Total number of patients identified by the algorithm;

6. INFORMED CONSENT AND DATA PROTECTION

6.1 Informed Consent

This study is an observational and retrospective analysis, therefore it does not provide direct involvement of patients (see the AIFA Guidelines for Observational Studies, Determination 20/03/08 - CURI No. 76 of 31/03/2008). Since collected data will be anonymous, they are not subject to Privacy Code regarding the protection of personal data. Data will be collected exclusively from regional administrative databases. To respect privacy, regional institution anonymised all data that is sent monthly by regional hospitals. Anonymization is carried out by constructing a unique code to replace the patient's fiscal code. This means that data retrieved from the regional databases are anonymous and cannot be traced back to identity individual patient in any way. For this reason, not even the informed consent of patients of Modena and Reggio Emilia hospitals will be collected.

Clinical data will be referred only for IRST patients to validate the algorithm. This is a non-interventional study based on secondary use of data collected for other purposes. So according to subparagraph 4 of Article 110-bis of the Privacy Code, which reports that " The processing of personal data collected for clinical purposes, for research purposes, by Scientific Institute for Research and Healthcare, public and private, does not constitute further processing by third parties because of the instrumental nature of health care activities carried out by these institutions compared to research, in compliance with the provisions of Article 89 of the Regulation" the IRCCS as IRST institute are exempt from collection of individual informed consent. Moreover, institutional informed consent form submitted to all patients at the first access in IRST included consent to use the patient's data, materials and/or test results for research purposes. In compliance with the "Code regarding the protection of personal data" (Legislative Decree 196/2003) appropriate safeguard will be implement to protect privacy.

6.2 Patient Data Protection

IRST patient data retrieved from electronic health record will be used to validate algorithm Data will be anonymised and aggregated, stored in a computer database, maintaining confidentiality in accordance with national data legislation. For data verification purposes, authorized representatives of Sponsor/Promoter, a regulatory authority, an Ethics Committee may require direct access to parts of the hospital or practice records relevant to the analysis. The

responsible investigator will ensure that this project is conducted in agreement with either the Declaration of Helsinki (Tokyo, Venice, Hong Kong, Somerset West and Edinburgh amendments) or the laws and regulations of the country, whichever provides the greatest protection to the patient. The protocol has been written, and the project will be conducted according to the ICH Harmonized Tripartite Guideline for Good Clinical Practice (ref:http://www.wma.net/en/30publications/10policies/b3/index.html). The protocol will be approved by the Local, Regional or National Ethics Committees.

6.3 Source data

For Reggio Emilia and Modena hospitals, patient information will be only retrieved from regional administrative health databases. Only for IRST patients data will be retrieved from electronic health records in order to validate algorithm.

The investigator is responsible for storage of essential documents before, during and after study completion, according to GCP (Good Clinical Practice). The Investigator must make study data accessible to the monitor, other authorized representatives of the Sponsor/Promoter (or designee), IEC, and Regulatory Agency inspectors upon request.

All study documents must be kept secured for a period of 15 years after completion or discontinuation of the study or for the length of time required by relevant national or local health authorities, whichever is longer. There may be other circumstances for which the Sponsor/Promoter is required to maintain study records and, therefore, the Sponsor/Promoter should be contacted prior to removing study records for any reason.

6.4 eCRF

Not applicable.

6.4.1 Paper CRF

Not applicable.

6.4.2 Study Database (applicable only for mono centric studies)

Not applicable.

6.5 Laboratory examinations

Not applicable.

6.6 Study Monitoring

Principal Investigator agrees to perform the study in accordance with ICH Good Clinical Practice and to provide all information requested in the Study Database, in an accurate manner according to the instructions provided. The Investigator has responsibilities to the Health Authorities to take all reasonable steps to ensure the proper conduct of the study as regards ethics, protocol adherence, integrity and validity of the data recorded on the case report forms/study database.

If a monitoring visit is required the trial Monitor will contact the site to arrange a date for the proposed visit and will provide the site with written confirmation. Investigators will allow the trial staff access to source documents as requested.

The main duty of the Coordinating Center/Trial Monitor is to help the Investigator and the Study Coordinators to maintain a high level of ethical, scientific, technical and regulatory quality in all aspects of the study. According to the guidelines on ICH Good Clinical Practice, the trial monitoring aim is to check subject informed consent and the study database entries against the source documents. Trials staff will check data for compliance with the protocol, data consistency, missing data and timing. Investigator will be sent requests missing data or clarification of inconsistencies or discrepancies.

7. STATISTICAL CONSIDERATIONS

7.1 Study design and objectives

This is a multi-center, observational study involving retrospective collection of NSCLC patients information from regional administrative databases. All Patients identified through the algorithm as described at paragraph 4.1 will be considered.

Primary objective will be to measure a set of potential indicators (KPIs) derived from regional administrative database in order to investigate the appropriateness and quality of NSCLC care among participating sites.

The secondary aim will be to develop and validate an algorithm to identify the eligible population of the study from administrative databases against to a clinical database (only at IRST site).

7.2 Population size

Due to the descriptive nature of the study without a hypothesis testing, a formal calculation of sample size and statistical power was not performed. Considering a preliminary analysis of clinical

medical charts from previous years in participating centers, it will be feasible to include about 400 patients with defined characteristics as defined at 4.1 paragraph.

7.3 Data analysis

For primary objective, for each indicator, the proportion of patients who receive the procedure, in the defined time window, among those eligible will be calculated. Descriptive statistics as arithmetic mean and standard deviation (SD) as well as median value and minimum–maximum statistics will be presented for KPI related to average time. Two-sided 95% confidence interval (95% CI) will be reported as appropriate and specified for KPI expressed as proportion. For secondary objective, the validity of algorithm will be evaluated by individual matching among cases retrieved from administrative database and medical chart obtained from IRST. Following definition will be used to explain this objective: NSCLC cases appearing as incident cases in administrative database, but not in IRST clinical chart will be considered as false positive and those appearing in IRST clinical chart but not in administrative database will be considered as false negatives.

Sensitivity will be defined as the percentage of NSCLC cases correctly identified by the algorithm and calculated as the proportion among NSCLC cases identified both in administrative database and IRST clinical chart (‘true positive’ cases) and all cases identified as NSCLC cases from IRST clinical chart (‘true positive’ + ’false negative’ cases). Positive predictive value will be defined as the percentage of NSCLC patients identified from IRST clinical chart were really NSCLC. This proportion will be calculated as the proportion among NSCLC cases identified both in administrative database and IRST clinical chart (‘true positive’ cases) and all cases identified as NSCLC cases from administrative database (‘true positive’ + ’false positive’ cases).

Likewise, specificity and negative predictive value will be calculated, and exact (Clopper-Pearson) 95% confidence limits for binomial distribution will be applied.

Statistical analysis will be performed using Stata software 15.1/SE for Windows (Stata Corp LP, USA).

7.4 Study duration

For this retrospective study, duration will be strictly related to time used for algorithm development and application both for administrative database and IRST clinical chart, for revision of discrepancies and time for statistical analysis. This time will be estimated in about six months.

8. ETHICAL ASPECTS

8.1 Local regulations/Declaration of Helsinki

The responsible Investigator will ensure that this study is conducted in compliance with the protocol, following the instructions and procedures described, adhering to the principles of Good Clinical Practice ICH Tripartite Guideline (December 2000) and in accordance with the principles laid down by the 18th World Medical Assembly (Helsinki, 1964 and further amendments) or with the laws and regulations of the country in which the research is conducted, whichever affords the greater protection to the individual.

8.2 Independent Ethical Committee

The protocol and any accompanying material provided to the patient will be submitted by the investigator to an Independent Ethical Committee for review. Approval from the committee must be obtained before starting the study. Any modifications made to the protocol, informed consent or material provided to the patient after receipt of the Ethics Committee approval must also be submitted by the investigator to the Committee in accordance with local procedures and regulatory requirements. The Independent Ethical Committee approval report must contain details of the trial (title, protocol number and version), documents evaluated (protocol, informed consent if required, accompanying material) and the date of the approval.

9. ADMINISTRATIVE REGULATIONS

The group of Outcome Research is responsible for drawing up the final version of the protocol and implementing the algorithm, defining general organizational procedures and organizing periodic meetings and newsletters. The Outcome Research will also undertake the following: support for the preparation of all documents needed for EC submission of the study protocol for each participating center, training of staff assigned to data collection, definition of monitoring procedures.

The Outcome Research is responsible for sending (by mail) the specific study document to the Register of the Observational Studies (RSO) established by AIFA, with the aim of collecting, in a single national archive, the data related to non-interventional clinical research focused on drugs in a prospectively manner.

10.2 Curriculum vitae

An updated copy of the curriculum vitae of each Principal Investigator, duly signed and dated, will be provided to the CC prior to the beginning of the study.

10.3 Secrecy agreement

All goods, materials, information (oral or written) and unpublished documentation provided to the Investigators, including this protocol, shall be considered confidential and may not given or disclosed to third parties.

10.4 Insurance

As this is an observational study, a specific insurance policy is not necessary in addition to the hospital insurance policy for daily clinical practice, according to "Determination AIFA 20 March 2008: Guidelines for the classification and conduction of observational study".

11 OWNERSHIP OF THE DATA AND USE OF THE STUDY RESULTS

The full ownership of the results (measured KPIs) generated in this study is retained by IRST IRCCS.

12 PUBLICATION POLICY

Publications regarding the main study objectives will be prepared by the Chief Investigator. Authorship will be proportional to the accrual of each center. All the recruiting investigators will be included in the authors list.

Once the Chief Investigator has presented the main study publication, any participating center may, eventually, use its own data (data generated in its own center) for educational purposes, publications and presentations. These should be sent to the Promoter for approval with a 15 days notice for abstracts, presentations or educational material and a 30 days notice for publications.

The manuscripts will be sent to major peer-reviewed scientific journals. The Chief Coordinator must approve all publications, abstracts and presentations based on patients included in this study. This is applicable to any individual patient registered in the study, or any subgroup of the study patients. Such a publication cannot include any analysis of any of the study end-points unless the final results of the study have already been published by the Study Coordinator

10 PROTOCOL AMENDMENTS

It is specified that the appendices, attached to this protocol and referred to in the main text of this protocol, form an integral part of the protocol.

No changes or amendments to this protocol may be made by the Investigators after the protocol has been agreed to and signed by both parties. Any change agreed upon will be recorded in writing, the written amendment will be signed by the Chief Investigator and by the Principal Investigator and the signed amendment will be appended to this protocol.

Approval / advice of amendments by Ethical Committees or similar body is required prior to their implementation, unless there are overriding safety reasons.

If the change or deviation increases risk to the study population, or adversely affects the validity of the clinical investigation or the subject's rights, full approval / advice must be obtained prior to implementation. For changes that do not involve increased risk or affect the validity of the investigation or the subject's rights, approval / advice may be obtained by expedited review, where applicable.

In some instances, an amendment may require a change to a consent form. The Investigator must receive approval / advice of the revised consent form prior to implementation of the change.

11 REFERENCES

1. (WHO) WHO. Cancer Fact sheet. [cited May, 2018]. Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.
2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. 2008;83(5):584-594. doi:10.4065/83.5.584
3. Soerjomataram I, Lortet-Tieulent J, Parkin DM, et al. Global burden of cancer in 2008: a systematic analysis of disability-adjusted life-years in 12 world regions. *Lancet (London, England)*. 2012;380(9856):1840-1850. doi:10.1016/S0140-6736(12)60919-2
4. Institure NC. Surveillance, Epidemiology, and End Results program (SEER). Available on line: http://seer.cancer.gov/csr/1975_2012/browse_csr.php.
5. Martini N, Bains MS, Burt ME, et al. Incidence of local recurrence and second primary tumors in resected stage I lung cancer. *J Thorac Cardiovasc Surg*. 1995;109(1):120-129. doi:10.1016/S0022-5223(95)70427-2
6. al-Kattan K, Sepsas E, Fountain SW, Townsend ER. Disease recurrence after resection for stage I lung cancer. *Eur J Cardiothorac Surg*. 1997;12(3):380-384. doi:10.1016/s1010-7940(97)00198-x
7. Martin J, Ginsberg RJ, Venkatraman ES, et al. Long-term results of combined-modality therapy in resectable non-small-cell lung cancer. *J Clin Oncol*. 2002;20(8):1989-1995. doi:10.1200/JCO.2002.08.092
8. Tabchi S, Kassouf E, Rassy E El, et al. Management of stage III non–small cell lung cancer. *Semin Oncol*. 2017;44(3):163-177. doi:10.1053/j.seminoncol.2017.10.009
9. Cutrona SL, Toh S, Iyer A, et al. Design for validation of acute myocardial infarction cases in Mini-Sentinel. *Pharmacoepidemiol Drug Saf*. 2012;21 Suppl 1:274-281. doi:10.1002/pds.2314
10. Lo Re V 3rd, Haynes K, Goldberg D, et al. Validity of diagnostic codes to identify cases of severe acute liver injury in the US Food and Drug Administration's Mini-Sentinel Distributed Database. *Pharmacoepidemiol Drug Saf*. 2013;22(8):861-872. doi:10.1002/pds.3470
11. Whyte JL, Engel-nitz NM, Teitelbaum A, Rey GG, Kallich JD. An Evaluation of Algorithms for

Identifying Metastatic Claims Data. 2015;53(7):49-57.

12. Baldi I, Vicari P, Di D, et al. A high positive predictive value algorithm using hospital administrative data identified incident cancer cases. 2008;61. doi:10.1016/j.jclinepi.2007.05.017
13. Abraha I, Serraino D, Giovannini G, et al. Validity of ICD-9-CM codes for breast , lung and colorectal cancers in three Italian administrative healthcare databases : a diagnostic accuracy study protocol. 2016:6-10. doi:10.1136/bmjopen-2015-010547
14. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity , speci Þ city and predictive values Positive Predictive Value (PPV). 2008;(February):45-50.

APPENDIX A: Autorizzazione n. 9/2016

"Autorizzazione generale al trattamento dei dati personali effettuato per scopi di ricerca scientifica (15/12/2016)"

https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/5805552

APPENDIX B: Determination AIFA 20 March 2008

"Guideline for the classification and conduction of observational study"

https://www.agenziafarmaco.gov.it/ricclin/sites/default/files/files_wysiwyg/files/Normativa/DETERMINAZ_AIFA_20_Marzo_2008_ST_OSS.pdf