STATISTICAL ANALYSIS PLAN

Impact of Patient Characteristics on pneumo-oncologists NSCLC systemic treatment decision in Belgium: A cross-sectional study.

BEPACT Lung

NCT03959137

Version	Date	Author	Job Title	Status
1.0	04 November 2019	Sofie Van Waes	Biostatistician	Final

Approvals

Hermine Leroi, MSD Belgium, Brussels, Belgium

Signature: Leroi H. Date: 13 NOV 2019

Leonard Kaufman, Statistician, Data Investigation Company Europe, Brussels, Belgium

Signature: L Kaufman Date: 25 NOVEMBER 2019

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

bev	bevacizumab
BSC	Best supportive care
Chemo	Chemotherapy
CI	Confidence Interval
CT	Clinical trial
CRF	Case Report Form
EP	Eligible Population
IBD	Inflammatory bowel disease
ICF	Informed consent form
IO	Immunotherapy
NSCLC	Non-small cell lung cancer
SAP	Statistical Analysis Plan
SD	Standard deviation

1. METHODOLOGY

1.1 Introduction and Purpose

The purpose of this statistical analysis plan (SAP) is to describe the final analysis of the data of the study.

The documents used in preparation of this SAP are:

- Study Protocol Version 4.0, dated 10 September 2019.
- Study annotated CRF Version 2.0, dated 17 May 2019.

1.2 Objectives and Endpoints Mentioned in the Protocol

1.2.1 Study Objectives

The primary objective of this study is to investigate the relation between patient, tumor and site characteristics and systemic treatment choices of stage IV untreated non-small cell lung cancer (NSCLC).

It is hypothesized that there is a relation between the systemic treatment choice and several explanatory variables that can be categorized as follows:

- Patient demography characteristics
- Medical history
- Co-morbidities
- Autoimmune disease
- Current/recent medication
- Prior cancer treatment in earlier stage
- Site characteristics

A detailed list of the variables can be found in the Section 1.5. For each of the variables, it will be queried whether the variable had an (important) impact on the systemic treatment choice.

1.2.2 Endpoints

The outcome variable is systemic treatment choice or best supportive care, defined as:

1. Chemotherapy (chemo)
2. Immunotherapy (IO)
3. Immuno combined therapies (IO+IO)
4. IO+chemo
5. IO+bevacizumab+chemo (IO+bev+chemo)
6. Best supportive care (BSC)

In practice, only four categories of the outcome variable were observed: Chemo, IO, IO+chemo, and BSC. Therefore, in the analysis the outcome variable will be considered as having four levels.

1.3 Study Design

This is a multicenter non-interventional cross-sectional study on first-line systemic treatment choice in stage IV NSCLC patients coming to the consultation. The study was to involve approximately 21 Belgian hospitals where consecutive patients with metastatic stage IV NSCLC selected for systemic

treatment or BSC are included in the trial at the time the patient signs the informed consent form (ICF). Patients selected for systemic treatment, should have received at least their first dose and a maximum of one cycle of the same treatment.

The pneumo-oncologist who treats the patient fills in the CRF, indicating the systemic treatment or BSC option and the link with the variables that impact the choice of systemic treatment or BSC. The actual value of the considered variables is also recorded in the CRF.

A questionnaire must be filled in characterizing the site and allowing the site to be assigned to a group based on the number of newly diagnosed patients per year with lung cancer (diagnostic volume) and whether the site participates in interventional clinical trials (CT participation).

In order to minimize bias, it was emphasized to the participating centers that consecutive patients need to be included in the study. In addition, a maximum of 10% (20 patients) of the total study population can be enrolled per site.

1.4 Sample Size Justification

No formal sample size calculation was performed. The protocol states that a maximum of 210 eligible patients and a minimum of 200 patients are to be included to have a representative picture of the pneumo-oncologists' NSCLC systemic treatment decision in Belgium.

1.5 Recorded Data

This section describes the covariates, that are considered in the analysis. These are divided into Site characteristics, recorded at the level of the site, and Patient characteristics, recorded at the level of a patient.

1.5.1 Site Characteristics

The following main site characteristics are recorded and will be used to classify the sites into four groups:

- Approximate number of newly diagnosed NSCLC patients (all stages) in reference year 2018 (diagnostic volume)
- Participation in lung cancer interventional clinical trials (CT participation) (yes/no)

The following additional site characteristics are recorded:

- Genetic testing
  - In-house (yes/no)
  - Referral testing (yes/no)
- PD-L1 IHC
  - In-house (yes/no)
  - Referral testing (yes/no)
- NGS testing
  - In-house (yes/no), if yes: type (whole exome/tumor gene panel/foundationONE/unknown[1])
  - Referral testing (yes/no), if yes: type (whole exome/tumor gene panel/foundationONE/ unknown[1])

[1] Considered missing in the analysis

1.5.2 Patient Characteristics

After making the treatment choice, the investigator records the following variables and indicates whether the characteristic impacted the treatment choice (yes/no). Of the variables indicated as having an impact, the investigator choses at least one and a maximum of three most important characteristics.

Demographic data

- Age (years)
- Gender (male/female)
- Weight loss (%)
- Smoking status (current/former (≥1 year)/never/unknown[1])
- ECOG status (0/1/2/3-4/unknown[1])
- Patients' preference for treatment choice (yes/no/unknown[1])

Medical history

- Metastatic disease status (M1a/M1b/M1c/unknown[1])
- Tumor burden:
  - Tumor size diameter (mm)
  - Number of metastatic sites
- Brainmets (yes/no), if yes:
  - Symptomatic/asymptomatic
  - Treated with SRT/treated with WBRT/not treated
- Livermets (present/not present/unknown[1])
- Concomitant malignancies (present/not present/unknown[1])
- Histology (squamous/non-squamous/not otherwise specified[1])
- IHC PD-L1 (PD-L1 TPS <1% / 1-49% / ≥50% /unknown[1])
- High TMB (≥ 10 mutations/Mb) (present/not present/unknown[1])
- Genetic testing complications (turnaround time/biopsy availability/not tested[1]/not applicable[1])

Co-morbidities

Presence (yes/no) of the following diseases:

- Myocardial infarction in the last year
- Congestive heart failure
- Peripheral vascular disease
- Hypertension
- Cerebrovascular event in the last year
- Transient ischemic attack in the last year
- Diabetes with complications
- Renal disease
- Liver disease
- Hematological disease
- Chronic obstructive pulmonary disease
- Chronic infections
- Mental disorders

[1] Considered missing in the analysis

Autoimmune disease

Presence, and in case present, active (yes/no), of the following diseases:

- Inflammatory interstitial lung disease
- Inflammatory bowel disease (IBD)
- Inflammatory joint disease
- Psoriasis
- Connective tissue disease

Current/recent medication

- Corticosteroids (yes/no), and if yes, low dose/high dose (≥10mg/day Prednisolone or equivalent)
- Other immunosuppressants
- Antibiotics

Prior cancer treatment in earlier stage NSCLC

Did the patient undergo the following treatment (yes/no) and if yes, the year of the treatment:

- Single-modality treatment
- Multi-modality treatment
- Surgery/radiotherapy
- Previous IO treatment

1.6 Definitions and Derived Variables

Site type

Sites will be divided into four groups based on the two main site characteristics, diagnostic volume and CT participation:

- High number of patients per year (more than the median number of NSCLC patients per year) and participating in interventional clinical trials
- High number of patients and not participating in interventional clinical trials
- Low number of patients (less than or equal to the median number of NSCLC patients per year) and participating in interventional clinical trials
- Low number of patients and not participating in interventional clinical trials

1.7 Analysis Set

The following analysis set is considered:

Eligible Population (EP): All subjects enrolled in the study, who have given informed consent and who satisfied all inclusion and none of the exclusion criteria.

The patients that should be excluded from the EP will be discussed during a data review meeting.

All analyses will be performed on the EP.

1.8 General Methodology

The statistical analysis will be performed using the SAS software for WINDOWS, Version 9.2. It will consist of a descriptive analysis and statistical modelling. Furthermore, lists of individual data will be provided for all patient characteristics.

1.8.1 Descriptive Analysis

Throughout this document the following terminology is used for indicating the type of descriptive analysis:

- 'Descriptive statistics' is the description to be provided for quantitative variables, and consists of the mean, standard deviation (SD), 95% confidence interval (CI) on the mean, minimum, 1st quartile, median, 3rd quartile, maximum, number of available observations, and number of missing observations.
- 'Frequency distribution' is the description to be provided for ordinal and nominal variables and consists of numbers and percentages for each of the scores or categories. Exact 95% CIs will be provided for key variables.

The descriptive analysis consists of:

1. **Site Characteristics**

Descriptive statistics will be provided for the diagnostic volume. Other site characteristics, including the site type, will be described by frequency distributions.

2. **Patient Characteristics by Outcome Variable and by Site Type**

Descriptive statistics and frequency distributions will be used to describe numerical and ordinal/nominal patient characteristics, respectively. The analysis will be performed for:

- All patients in the EP
- By outcome variable (Chemo+IO, Chemo, IO, BSC)
- By site type (4 subgroups).

Also, frequency distributions of the outcome variable will be given for all patients in the EP and by site type.

3. **Impact of Covariates on the Outcome Variable**

The impact of variables, as indicated by the investigator, on the outcome variable, is described as one of the following:

- Important impact (variable selected as one of the three most important characteristics)
- Impact (variable indicated as having an impact, but not selected among the three most important characteristics)
- No impact (variable not indicated as having impacted the treatment choice)

For each variable the frequency distribution of the impact is provided per level of the outcome variable (chemo, IO, IO+chemo, and BSC).

1.8.2 Statistical Modelling

An exploratory analysis will be performed, consisting of a multinomial logistic regression model that assesses the effect of patient characteristics and the site type on the outcome variable.

The previously defined outcome variable has four levels (chemo, IO, IO+chemo and BSC). The level with the highest frequency will be used as the reference in the multinomial logistic model.

Variables considered as predictors in the model are the site type and patient characteristics with less than 20% missing or unknown values. In case of yes/no variables with further conditional questioning (e.g. Inflammatory interstitial lung disease (yes/no) and if yes, Active/non-active) only the yes/no variable is included in the model. Selection of the final model, containing only predictors with a significant effect on the outcome variable, will be done using automatic selection procedures.

1.8.3 Safety Analysis

This is an epidemiological study. No safety data were to be collected.

1.9 Deviations from the Analysis Foreseen in the Protocol

The protocol foresaw an exploratory analysis of the outcome variable using a penalized hierarchical logistic regression model for discrete choices.

Since a penalized logistic regression is not available in SAS, it was decided to use the classical multinomial logistic regression approach, with a stepwise selection procedure.

2. DETAILS OF THE STATISTICAL ANALYSIS

2.1 Sample Description

- Number of subjects enrolled, and number in the EP, overall and broken down by site
- First and last inclusion date, overall and by site

2.2 Site Characteristics

- Descriptive statistics for diagnostic volume
- Frequency distribution and 95% CI for CT participation
- Frequency distribution for site type
- Frequency distribution for Genetic testing, in-house
- Frequency distribution for Genetic testing, Referral testing
- Frequency distribution for PD-L1 IHC, in-house
- Frequency distribution for PD-L1 IHC, Referral testing
- Frequency distribution for NGS testing, in-house (yes/no) and frequency distribution for Type
- Frequency distribution for NGS testing, Referral testing (yes/no) and frequency distribution for Type

2.3 Patient Characteristics by Outcome Variable and by Site Type

- Frequency distribution and 95% CI for the outcome variable (treatment choice).
- Frequency distributions and 95% CI for the outcome variable (treatment choice) by site type.

The following descriptive analysis is performed broken down by the outcome variable (all patients in the EP and 4 subgroups), and by site type (4 subgroups):

Demographic data
- Descriptive statistics for Age and Weight loss
- Frequency distributions for Gender, Smoking status, and ECOG
- Frequency distribution for Patients' preference for treatment choice

Medical history
- Frequency distribution for Metastatic disease status
- Descriptive statistics for Tumor size diameter and Number of metastatic sites
- Frequency distributions for Brainmets, Symptomatic/asymptomatic, and Treatment
- Frequency distributions for Livermets, Concomitant malignancies, Histology, IHC PD-L1, High TMB, and Genetic testing complications

Co-morbidities
- Frequency distribution for each Co-morbidity

Auto-immune disease
- Frequency distributions of presence and activity of each Auto-immune disease

Current/recent medication
- Frequency distributions for use of Corticosteroids and for Dose
- Frequency distributions for use of Other immunosuppressants and Antibiotics

Prior cancer treatment in earlier stage NSCLC
- Frequency distribution for each Prior cancer treatment in earlier stage NSCLC

2.4 Impact of Covariates on the Outcome Variable

For each variable for which the impact has been evaluated, frequency distributions will be given of the impact, for all patients and per level of the outcome variable (treatment choice).

Demographic data
- Age and Weight loss
- Gender, Smoking status, and ECOG
- Patients' preference for treatment choice

Medical history
- Metastatic disease status
- Tumor size diameter and Number of metastatic sites
- Brainmets, Symptomatic/asymptomatic, and Treatment
- Livermets, Concomitant malignancies, Histology, IHC PD-L1, High TMB, and Genetic testing complications

Co-morbidities
- Each Co-morbidity

Auto-immune disease
- Presence and activity of each Auto-immune disease

Current/recent medication
- Use of Corticosteroids and Dose
- Use of Other immunosuppressants and Antibiotics

Prior cancer treatment in earlier stage NSCLC
- Each Prior cancer treatment in earlier stage NSCLC

2.5 Exploratory Analysis

The following variables will be considered in a multinomial stepwise logistic regression analysis:
- Site type (4 subgroups)
- Age
- Gender
- Weight loss
- Smoking status (current/former (≥1 year)/never)
- ECOG status (0/1/2/3-4)
- Patients' preference for treatment choice (yes/no)
- Metastatic disease status (M1a/M1b/M1c)
- Tumor size diameter
- Number of metastatic sites
- Brainmets (yes/no)
- Livermets (present/not present)
- Concomitant malignancies (present/not present)
- Histology (squamous/non-squamous)
- IHC PD-L1 (PD-L1 TPS <1% / 1-49% / ≥50%)
- High TMB (present/not present)
- Genetic testing complications (turn-around time/biopsy availability)
- Each co-morbidity (yes/no)
- Each auto-immune disease (yes/no)
- Use of corticosteroids, other immunosuppressants, and antibiotics (yes/no)
- Use of each prior cancer treatment in earlier stage NSCLC (yes/no)