STATISTICAL ANALYSIS PLAN

Version v1.0

30/10/2019

1199-0296	OBSERVATIONAL ANALYSIS ON THE SOCIO-ECONOMIC IMPACT OF IPF IN SPAIN.

Boehringer Ingelheim España, S.A

Boehringer Ingelheim

Document:

SAP_OASIS_v1.0_20191030

PAGE OF SIGNATURES

	Responsible	Signature	Date (dd/mm/yyyy)
Prepared by:	Statistician		
Reviewed by:	Statistician 2		
	Project		
Approved by:	Boehringer Ingelheim		
	Boehringer Ingelheim España		

1. OVERVIEW OF THE STUDY

1.1. SPONSOR IDENTIFICATION

Boehringer Ingelheim España, S.A.

C/ Prat de la Riba, 50

08174 Sant Cugat del Vallès (Barcelona)

1.2. STUDY TITLE

Observational Analysis on the Socio-economic Impact of IPF in Spain.

1.3. PROTOCOL CODE

BI study number: 1199-0296

1.4. COORDINATOR INVESTIGATOR(S):

(Granada)

(Barcelona)

(Lugo)

(Madrid)

1.5. TYPES OF SITES WHERE THE STUDY WILL TAKE PLACE

It is planned that data of approximately 200 patients from approximately 25 sites (secondary care sites – Pulmonology services where IPF is diagnosed and managed) in Spain will be collected.

1.6. ETHICS COMMITTEE THAT EVALUATES THE STUDY

The study was presented to the specific ECs of the participating hospitals. The CEIC of acted as reference EC.

1.7. PRIMARY OBJECTIVE

To compare the economic impact of Idiopathic Pulmonary Fibrosis (IPF) according to FVC % predicted level (FVC<50%, FVC 50-80%, FVC>80%), in adult patients through estimation of annual direct costs (primary and secondary care visits, outpatient visits, emergency visits -

primary care and hospital-, hospitalizations,…), and indirect costs (days off work and informal caregiver) associated with the disease during one year.

1.8. STUDY DESIGN

Non-interventional multicenter study based on newly collected data of IPF patients followed-up for one year in secondary care settings (Pulmonology Services). IPF patients will be enrolled in a consecutive manner over a period of 6 month.

1.9. DISEASE(S) OR CONDITION(S) BEING STUDIED

Idiopathic Pulmonary Fibrosis (IPF)

1.10. DETAILS OF THE MEDICINAL PRODUCTS BEING STUDIED

Not applicable

1.11. STUDY POPULATION AND TOTAL NUMBER OF SUBJECTS

It is planned that data of approximately 200 IPF patients in Spain will be collected.

1.12. STUDY CALENDAR

Start of data collection: 29 November 2017

Planned end of data collection: August-2019

Planned final study report: December-2019

1.13. FUNDING SOURCE

Boehringer Ingelheim España, S.A.
C/ Prat de la Riba, 50
08174 Sant Cugat del Vallès (Barcelona)

1.14. DETAILS OF THE COORDINATOR SITE

Tel.:

Fax:

TABLE OF CONTENTS

2. DICTIONARY OF ABBREVIATIONS

A description of the meaning of the abbreviations used in this document can be found below:

ADR	Adverse Drug Reaction
AE	Adverse Event
AESI	Adverse Event of Special Interest
BI	Boehringer Ingelheim
BMI	Body Mass Index
CA	Competent Authority
CCDS	Company Core Data Sheet
CI	Confidence Interval
CML	Local Clinical Monitor
CRA	Clinical Research Associate
CRF	Case Report Form
CTCAE	Common Terminology Criteria for Adverse Events
CTP	Clinical Trial Protocol
DLCO	Diffusing capacity of the lung for carbon monoxide
eCRF	Electronic Case Report Form
EQ-5D-5L	EuroQoL five dimensions questionnaire 5L
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
FDA	Food and Drug Administration
FVC	Forced Vital Capacity
GCP	Good Clinical Practice
GEP	Good Epidemiological Practice
GPP	Good Pharmacoepidemiology Practice
GVP	Good Pharmacovigilance Practices
IB	Investigator's Brochure
IEC	Independent Ethics Committee
IPF	Idiopathic Pulmonary Fibrosis
IQR	Interval quartile range
IRB	Institutional Review Board
MAH	Marketing Authorisation Holder Activities
Max.	Maximum

MedDRA	*Medical Dictionary for Regulatory Activities*
Min.	Minimum
NHS	National Health Service
NIS	Non-Interventional Study
PASS	Post-Authorization Safety Study
QoL	Quality of Life
SAE	Serious Adverse Event
SAS	*Statistical Analysis System*
SD	Standard Deviation
SGRQ	St. George's Respiratory Questionnaire
UIP	Usual Interstitial Pneumonia
6MWD	6 Minutes Walk Distance test

3. STUDY OBJECTIVES

3.1. PRIMARY OBJECTIVE

To compare the economic impact of Idiopathic Pulmonary Fibrosis (IPF) according to forced vital capacity (FVC) % predicted level (FVC<50%, FVC 50-80%, FVC>80%), in adult patients through estimation of annual direct and indirect costs associated with the disease during one year.

3.2. SECONDARY OBJECTIVES

- To estimate the QoL of the patients with IPF according to FVC % predicted level, through SGRQ and EQ-5D-5L questionnaires and the Barthel Index.

- To explore the determinants of costs and QoL in patients with IPF according to FVC % predicted level.

- To characterize acute IPF exacerbations along one year (frequency and cost) according to FVC % predicted level at baseline.

- To describe the variation of costs and QoL with IPF progression (according to the FVC deterioration).

- To explore the impact of disease on the patient's caregiver through Zarit Burden Interview questionnaire at 6 and 12 month.

4. STUDY POPULATION

4.1. ELIGIBILITY CRITERIA

It is planned that data of approximately 200 patients from approximately 25 sites (secondary care sites – Pulmonology services where IPF is diagnosed and managed) in Spain will be collected. All Idiopathic Pulmonary Fibrosis patients who are diagnosed with IPF and attend to a routine visit during the inclusion period and fulfill inclusion/exclusion criteria and provide informed consent to participate will be included in the study.

Patients can be included if all of the following criteria are met:

4.1.1. Inclusion criteria

1. Female and male patients ≥ 40 years of age.

2. Patients diagnosed with Idiopathic Pulmonary Fibrosis (IPF) according to last ATS/ERS/JRS/ALAT IPF guideline for diagnosis and management consensus.

3. Written informed consent prior to participation.

4.1.2. Exclusion criteria

1. Inability for the patient to understand or complete the written Inform Consent or patients questionnaires or to understand Spanish.

2. Current participation in any clinical trial.

3. Patients for whom further follow-up is not possible at the enrolling site.

4.2. SAMPLE SIZE JUSTIFICATION

In view of the lack of Spanish studies that analyse the economic impact of IPF according to deterioration in the predicted FVC, and as this is a prospective observational study, it was decided to observe all the patients with IPF that meet the inclusion/exclusion criteria during the inclusion period. Based on the investigation into the viability of recruitment, we expect to include approximately 200 patients.

The primary objective of the study is to compare the economic impact of IPF according to FVC levels. This would mean opting for a balanced sample size for the three comparison groups.

However, knowing the approximate patient prevalence according to predicted FVC (5-11% predicted FVC <50%, 51-73% predicted FVC 50-80%, and 22-38% predicted FVC >80%) and the sponsor's internal data, we have estimated to find:

- 10-22 patients with predicted FVC <50%.
- 102-143 patients with predicted FVC 50-80%.
- 43-76 patients with predicted FVC >80%.

Depending on the final distribution obtained in terms of number of patients for each comparison group and given the exploratory nature of this study, it is expected that comparisons between some groups will have a descriptive approach, although the expected comparison analyses will be carried out as stated in this protocol (explorative statistical model if sample size allows). Although the expected number of patients with predicted FVC <50% is

small, it should at least provide first insights into the costs within this patient subgroup. In view of the lack of previous data on the cost of the disease according to the study's primary endpoint in Spain, this will be the first Spanish cost comparison study according to predicted FVC.

5. METHODS

5.1. DATA PROCESSING

A data management plan (DMP) was created to describe all functions, processes, and specifications for data collection, cleaning and validation (version v2.0_20180711).

5.2. DATA ANALYSIS AND STATISTICAL TESTS

5.2.1. Protocol deviations

The next table defined the different categories of protocol violations (PVs). The final columns describes which PVs were used to exclude subjects from the different patient analysis sets, in case any of them would be detected during the study.

Description	Requirements	Patients excluded due to PV in Evaluable population	Patients excluded due to PV in Safety population
IC1. Patients under 40 years of age or with no specified age	not met as specified in the protocol	All	None
IC2. Patients without diagnosed with Idiopathic Pulmonary Fibrosis (IPF) according to last ATS/ERS/JRS/ALAT IPF guideline for diagnosis and management consensus	not met as specified in the protocol	All	None
IC3. No written informed consent prior to participation	not met as specified in the protocol	All	All
EC1. Inability for the patient to understand or complete the written Inform Consent or patients questionnaires or to understand Spanish	met as specified in the protocol	All	None
EC2. Current participation in any clinical trial	met as specified in the protocol	All	None
EC3. Patients for whom further follow-up is not possible at the enrolling site	met as specified in the protocol	All	None

5.2.2. Populations for analysis

Evaluable population: The analyses will be performed on a sample of assessable patients, which will include all patients who meet the selection criteria (no protocol deviations) and with predicted FVC classified.

Likewise, given the main objective of the study, patients for whom there is data on resource use due to IPF (primary care/secondary care/emergency visits, transport, hospitalization, outpatient tests, pharmacological and non-pharmacological treatment, caregivers [formal and informal], orthoprosthetic material, structural changes, economic aid, social services, lost work productivity, resource use due to acute exacerbations) will be considered valid for the main analysis.

Safety population: All patients screening with informed consent prior to participation.

5.2.3. Univariate descriptive analysis

Main sociodemographic and clinical variables will be collected in order to describe population included in the study.

Continuous variables will be described using means, standard deviations (SD) and 95% confidence intervals (IC), and medians and 25-75 percentiles. Categorical variables will be described using absolute and relative frequencies.

5.2.4. Analysis of the primary variable

To compare the economic impact of IPF according to predicted FVC% (FVC<50%, FVC 50-80%, FVC>80%), in adult patients through the estimation of annual direct and indirect costs associated with the disease)

Annual direct and indirect IPF-related costs will be quantified for each patient over the follow-up period of 12 months from the societal perspective and National Health Service (NHS).

Direct and indirect costs will be quantified according to resource use and days of sick leave.

5.2.5. Direct cost

Direct costs are costs related to the health care system and are divided into health and non-healthcare costs. By definition, **direct health costs** include hospital costs, outpatient costs, tests costs, treatment costs, etc., while **direct non-health costs** include transport costs, social services costs, essential adaptations of the patient's home costs, etc. Therefore, this study will be considered:

- **Direct health costs:**
    - Number of primary care visits due to IPF (T6, T12, acute exacerbations)
    - Number of secondary care visits due to IPF (pneumologist, nurse, nutritionist, psychiatrist, psychologist, respiratory rehabilitation, pneumologist home visit, nursing home visit, smoking consult cessation, other health professional) (T6, T12, acute exacerbations)
    - Number of emergency visits (primary care and hospital) due to IPF (T6, T12, acute exacerbations)
    - Number of hospitalizations due to IPF (time in emergency room, number of days in plant, number of days in ICU) (T6, T12, acute exacerbations)
    - Number of outpatient tests due to IPF (T6, T12, acute exacerbations):
        - Laboratory test (hemogram, biochemistry, coagulation profile, erythrocyte sedimentation rate, liver profile, angiotensin converting enzyme, rheumatoid factor, antinuclear antibodies, another tests).
        - Pulmonary function test (spirometry, pulmonary plethysmography, carbon monoxide diffusion capacity, 6-minute walk test, another test).
        - Other examinations (x-Ray, high resolution computed tomography, bronchoscopy, bronchoalveolar lavage, transbronchial biopsy, lung cryobiopsy, surgical lung biopsy, arterial blood gases, another test).
    - Non-pharmacological treatment due to IPF (liquid oxygen therapy, electric portable oxygen therapy, oxygen therapy concentrator, oxygen, oxygen therapy portable device, high flow nasal cannulas, non-invasive mechanical ventilation) (T6, T12, acute exacerbations).
    - Pharmacological treatments due to IPF (except treatments administered in hospitalization) (T6, T12, acute exacerbations).
- **Direct non-health costs:**
    - Transport (number of trips and km of displacement in taxi, number of trips and km of displacement in ambulance) (T6, T12, acute exacerbations).
    - Formal caregiver (time dedicated to patient care with IPF) (T6, T12, acute exacerbations).

- Orthopaedic material (manual wheelchair, electric wheelchair, anti-bedsore cushion, orthopaedic bed, orthopaedic chair, other) (T12, acute exacerbations).
- Economic aid and formal social services (T12, acute exacerbations)
- Structural adaptations (change in the structure of housing, elevators on the stairs in common areas, complementary technical aids, vehicle adaptation, other) (T12, acute exacerbations).

Calculation of direct health costs

Visits: The cost of visits will be calculated by multiplying the number of visits (emergency, medical) by the corresponding unit cost, which will depend on the health care area (primary or secondary) in which the visit occurred.

Hospitalizations: To calculate the cost of hospitalizations, days of stay in each unit (ICU, ward) will be multiplied by the unit cost corresponding to each unit. The sum of the costs of each service will then be calculated. In addition, emergency room admission will be added: if length of emergency stay was "less than 24 hours", emergency room visit cost will be considered; if length of emergency stay was "more than 24 hours", one day of emergency stay cost will be considered. Data on intubation during hospitalization (yes/no) will be used as qualitative information and not calculated as a cost.

Outpatient tests and non-pharmacological treatment: The cost of outpatient tests and non-pharmacological treatment will be obtained by multiplying the natural units of resource use (number of tests) by the unit cost of each outpatient test.

Pharmacological treatment: The costs of pharmacological treatment will be calculated taking into account:

- **Pharmacological doses**

  The total dose prescribed of each active ingredient during the last 6 months will be calculated by multiplying the dose prescribed by the time on treatment. The time on treatment will be determined according to the start and end date of treatment. The scheduled dose may be daily, twice a day, three times a day, alternate days, weekly, every 2 weeks, monthly or on demand. If the schedule dose is *'on demand'* an average dose according to the technical data sheet will be calculated.

The calculations described above will provide the total dose that each patient has received for the observational period.

- **Costs**

  For treatments dispensed in the pharmacy office, the unit costs for each treatment were calculated taking into account the retail price (RRP) plus value added tax (VAT). From the National Health System (NHS) perspective, the co-payment of the Spanish Royal Decree-Law 16/2012 will be applying:

  - Reduced contribution medication: 10% of the retail price for medication belonging to ATC groups qualifying for reduced contribution, with a maximum contribution expressed in euros. For reduced contribution medicines with a maximum of €4.26 per package.
  - Normal contribution medicines: copayment in this medicines group should be applied according to each individual situation and on declared annual income:

Medicines Co-payment	
Annual resident tax declaration	**Patient's copayment**
Income ≥100.000 €	60%
Income 18.000-100.000 €	50%
Income <18.000 €	40%
Pensioners	10%
Civil servant	30%

The cost of each pharmacological treatment will be obtained by multiplying the total dose (taking into account the prescribed dose and the administration schedule) each patient has received for the observational period by the unit cost of each treatment. If the commercial brand and presentation are available, the unit costs will be those that correspond to the said commercial brand and presentation. If this information is not available, an average cost will be calculated taking into account all commercialized presentations.

The annual direct health cost will be calculated as the sum of the costs of medical visits, emergency room visits, hospital admissions, outpatient tests, non-pharmacological treatments, and pharmacological treatments.

Calculation of direct non-health costs

Transport: The cost of ambulances will be obtained by multiplying the number of trips made per unit cost. For journeys by taxi, the kilometres recorded for each journey will be multiplied by the total cost of journeys made during the observation period, and the result will be multiplied by the corresponding unit cost.

Paid caregivers: The cost of paid caregivers will be calculated by multiplying the number of hours per day of care given by the number of days. The result will be the number of hours of paid care during the observational period, and will be multiplied by the corresponding unit cost.

Orthoprosthetic material: The cost of technical aids will be calculated by multiplying the number of days the patient has required the aid by the daily cost, calculated according to the catalogues of orthoprosthetic benefits of the autonomous communities.

Financial aid and social services: The cost registered in the eCRF will be considered. The monthly cost will be multiplied by the time aid was received. If the monthly amount is 'unknown', the mean cost of patients with full data will be used.

Structural changes: The cost registered in the eCRF will be considered.

Direct non-health costs will be calculated as the sum of transport costs, paid caregivers costs, orthoprosthetic material costs, financial aid and social services costs, and structural changes costs.

5.2.6. Indirect cost

Indirect costs are defined as the costs associated with the impact of the disease borne by the patient and will include:

- Patients' days off work along 12 months (T6, T12, acute exacerbations).
- Informal caregiver (time dedicated to patient care with IPF) (T6, T12, acute exacerbations).

The calculation will be based on the human capital method, assuming that the wage/salary reflects the productivity of the worker. The opportunity cost method will be used to calculate informal care costs. The indirect costs will be estimated by applying salary costs based on the latest data published by the Spanish Instituto Nacional de Estadística (INE) from the salary structure survey, adjusted to age. With this method, the indirect cost of patients/caregiver not actively employed during the study period (e.g., unemployed and retired workers) will be 0.

5.2.6.1 Total cost

In the analysis from the social perspective, the total costs will be obtained as the sum of the direct health costs (without co-payment), direct non-health costs and indirect costs.

In the analysis from the perspective of the Spanish NHS, health costs will be direct health costs and direct non-health cost considering only the part financed by the NHS.

5.2.6.2 Unit cost

The following unit costs and cost data sources will be used:

- Hospitalizations, emergency room visits, medical visits, outpatient tests, and non-pharmacological treatments: the unit costs will be established from the representative national rates collected in the ESALUD Spanish health costs database.
- Pharmacological treatments: prices published in the Drug Database of the General Council of Official Associations of Pharmacists.
- Lost productivity: wage/salary costs will be calculated according to the latest data published by the Spanish National Statistics Institute on the survey of the salary structure.

In order to obtain the most recent unit costs, they will be consulted at the time of the analysis. All costs will be expressed in euros for the year in which the analysis is performed, updating as necessary the costs from previous periods according to the cumulative consumer price index (CPI).

For the primary objective of the study, the characteristics of the subpopulations defined by the categorisation of the deterioration in predicted FVC (FVC <50%, FVC 50-80%, FVC >80%) at the time of inclusion (T=0) will be described by means of: mean, SD, 95% CI of the mean, median, IQR (P25-P75), max. and min. The CI of the costs will be calculated using bootstrapping techniques using replacement samples of the same size as the original sample. In total 10 000

simulations will be made and the 2.5 and 97.5 percentiles of the distribution will be used to determine the 95% CI.

Analysis of variance (ANOVA) or Kruskal-Wallis tests will be performed to compare the direct and indirect costs by subpopulation (FVC <50%, FVC 50-80%, FVC >80%). Results of all statistical tests must be interpreted purely exploratorily.

5.2.7. Analysis of secondary variables

The following analyses will be performed according to secondary objectives.

1) ***To estimate the QoL of patients with IPF according to predicted FVC% through the SGRQ and EQ-5D-5L questionnaires and the Barthel index***

Health-related quality of life (QoL) will be investigated with the St George's Respiratory Questionnaire (SGRQ) and the EQ-5D-5L questionnaire. Quality of life will be assessed at visits T0, T6 and T12.

The **St. George's Respiratory Questionnaire (SGRQ)** is a 50-item questionnaire developed to quantify the impact of the disease on the health and QoL perceived by patients with respiratory diseases. It consists of 50 items divided into 3 scales: symptoms (frequency and severity of respiratory symptoms), activity (activity limitations due to dyspnoea) and impact (psychological and social functioning disorders caused by the disease). The final scores range from 0 (fewest limitations) to 100 (most limitations). Spanish validated version will be used.

For the **SGRQ:** mean, SD, 95% CI of the mean, median, IQR (P25-P75), max. and min. will be calculated from the scores obtained on the three scales (symptoms, activity, impact) and from the total score in the total population and in each subpopulation (FVC <50%, FVC 50-80%, FVC >80%).

The **EQ-5D-5L** consists of a questionnaire and a visual analogue scale (EQ-VAS). The EQVAS is a self-rated health status using a VAS (0-100). The EQ-VAS records the subject's perceptions of their own current overall health. The self-assessment questionnaire is self-reported description of the subject's current health in 5 dimensions i.e., mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The subject is asked to grade their own

current level of function in each dimension into one of five degrees of disability (no problems, slight problems, moderate problems, severe problems, and extreme problems).

This decision results in a 1-digit number expressing the level selected for that dimension. The

digits for 5 dimensions can be combined in a 5-digit number describing the respondent's health state. Spanish validated version will be used.

For the **EQ-5D-5L questionnaire:** as the data are not continuous but ordinal, the information will be presented in terms of absolute and relative frequencies for each subpopulation (FVC <50%, FVC 50-80%, FVC >80%) by levels in each dimension. The EQ-5D-5L comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 five degrees of disability (no problems, slight problems, moderate problems, severe problems, and extreme problems).

Barthel Index will be used to score the ability of a patient to care for himself. It consists of 10 items, the values assigned to each item are based on time and amount of actual physical assistance required if a patient is unable to perform the activity. The final score range from 0 and 100. Patient scoring 100 is continent, feeds himself, dresses himself, gets up out of bed and chairs, bathes himself, walks at least a block, and can ascend and descend stairs. Spanish validated version will be used.

For the **Barthel index**: the items will be presented in terms of absolute and relative frequencies for each subpopulation (FVC <50%, FVC 50-80%, FVC >80%). The final score (0-100) will be presented by mean, SD, 95% CI of the mean, median, IQR (P25-P75), max. and min. will be calculated in each subpopulation (FVC <50%, FVC 50-80%, FVC >80%).

b) To explore the determinants of costs and QoL in patients with IPF according to predicted FVC%

In order to explore the determinants of costs and QoL in patients with IPF according to predicted FVC%, the following variables will be collected in the CRF at visit T0, T6 and/or T12:

- Sociodemographic variables: age, gender, employment, net incomes (T0).

- Anthropometric variables: weight, height and BMI (T0).

- Characteristics of IPF: time of IPF diagnosis (T0), FVC (L) (T0, T6, T12), FVC % predicted (T0, T6 and T12), FVC rate change by year (T0 and T12), DLCO (T0, T6 and T12), the

Barthel Index (T0, T6 and T12), 6MWD (T0 and T12), concomitant diseases related to IPF (T0, T6 and T12), pharmacological and non-pharmacological treatments related to IPF (T0, T6 and T12).

- Smoking status (T0, T6 and T12).

- Work with animals (now or in the past) (T0)

- Use of formal and/or informal caregiver (T0, T6 and T12).

If sample size allows, bivariate exploratory methods e.g. analysis of variance (ANOVA), Kruskal-Wallis test, Pearson/Spearman correlation will also be applied to explore predictors of costs and quality of life for each study subpopulation (FVC <50%, FVC 50-80%, FVC >80%) such as sociodemographic and clinical variables collected.

Taking into account that the expected final number of patients of the FVC <50% stratum could be small, with less of 20 patients, and the limitations that a multivariate analysis might have, an exploratory bivariate analysis could be performed with those variables considered as relevant.

Some variables may be categorised to provide practical information in the clinical setting.

If sample size allows, multivariate exploratory analyses (e.g. generalized linear models) will be applied which include costs and quality of life as dependent variables, and the explanatory variables will be discussed based on previous bivariate analyses and clinical relevance (with a $p<0.1$). Sample size and constraints imposed by the analysis models may prevent to include all variables of interest in this setting.

c) To characterize acute IPF exacerbations (frequency and cost) according to FVC% predicted level

In order to characterize acute IPF exacerbations during one year according to FVC% predicted level at one year, the following variables will be collected in the eCRF along the study (12 months).

- Acute exacerbation related resource use for direct cost estimation: primary and secondary care visits, emergency visits (primary care and hospital), hospitalizations, ICU with and without intubation (qualitative analysis), outpatient tests and other examinations, use of transport, use of formal caregiver, pharmacological and non-

pharmacological treatments (except treatments administered in hospitalization), 25rthopaedic material, formal social services, economic aid and structural adaptations.

- Acute exacerbation related resource use for indirect cost estimation: patients' days off work and informal caregiver.

- Costs will be quantified as explained in section 5.2.3.

The total number of acute IPF-related exacerbations, as well as the use of resources and costs associated with the events during one year according to the predicted FVC at the time of inclusion (FVC <50%, FVC 50-80%, FVC >80%) will be described.

For the quantitative (continuous) variables, the mean, the SD, the 95% CI of the mean, the median, IQR, min and max will be calculated. Qualitative (categorical) variables will be presented as absolute and relative frequencies (percentages).

d) To describe the cost and QoL variation associated with the lung capacity deterioration (change among the FVC % predicted level along 12 month)

In order to estimate the direct and indirect costs according to FVC deterioration the following variable will be collected in the eCRF along the study (T0, T6 and T12):

- FVC % predicted along the study (T0, T6 and T12)

  - Men: FVC % predicted (%)= 100 FVC/(0,0678 T – 0,0147 E – 6.0548)
  - Women: FVC % predicted (%)= 100 FVC /(0,0454 T – 0,0211 E – 2.8253)
  - (FVC is FVC in liters, T is height in cm and E is age in years)

The calculated variable will be stratified into the following subgroups between T0 and T12:

- ≤-10%
- from -10% to -5%
- >-5%

Descriptive analyses will be carried out of direct and indirect costs and patient quality of life according to the change in FVC (<-10%, from -10% to -5%, and >-5%) between T0 and T12.

For the quantitative (continuous) variables, the mean, the 95% CI, the median, IQR, min and max will be calculated. Qualitative (categorical) variables will be presented as absolute and relative frequencies (percentages).

When sample size is minimum, subgroup analyses will be explored for the following subgroup variables:

- % FVC predicted in T0 (FVC <50%, FVC 50-80%, FVC >80%)

- antifibrotic treatment vs. non antifibrotic treatment during study period

- one acute IPF exacerbation vs. more than one acute IPF exacerbation during study period

e) To assess the QoL of the caregiver through the Zarit Burden Interview questionnaire

Caregiver quality of life will be explored using the Zarit questionnaire. Caregiver quality of life will be assessed at visits T0, T6 and T12.

Caregivers who have signed the specific written informed consent form will be asked to complete the Zarit Burden Interview. It is a self-report measure. The revised version contains 22 items. Each item on the interview is a statement which the caregiver is asked to endorse using a 5-point scale. Response options, in the panish version, range from 0 (Never) to 4 (Nearly Always). The final scores range from 0 and 88: Little or no burden (≤ 421), mild to moderate burden (22-40), moderate to severe burden (41-60) and severe burden (≥ 61).

The responses to the 22 items that make up the instrument will be reported, as well as the total score for the questionnaire. The distribution of caregiver burden will be described by means of absolute and relative frequencies (percentages), according to the following categories:

- Little or no burden: Zarit score ≤21
- Mild to moderate burden: Zarit score >21 and ≤40
- Moderate to severe burden: Zarit score >40 and ≤60
- Severe burden: Zarit score ≥61

5.2.8. Statistical methodology

A descriptive analysis will be performed of all the variables recorded for the study population. For the continuous variables, the mean, the SD, the 95% CI of the mean, the median, IQR (P25-P75), min. and max, and valid n will be calculated. Categorical variables will be presented as absolute and relative frequencies (percentages).

For bivariate analysis, continuous endpoints may be compared across subgroups of population using two-sample t-tests or ANOVA (if normality criteria are met), or the Mann–Whitney U test or Kruskal-Wallis test (if normality criteria are not met). The categorical endpoints will be analysed using the Chi-square or Fisher test, as appropriate. If apply, lineal correlations between continuous variables will be analysed (Pearson or Spearman correlations).

Multivariate exploratory analyses (e.g. generalized linear models) will be applied which include costs (Total annual IPF-related costs) and quality of life (SGRQ and EQ-5D-5L in baseline visit-T0) as dependent variables, and the explanatory variables will be discussed based on previous bivariate analyses and clinical relevance with $p<0.1$. Sample size and constraints imposed by the analysis models may prevent to include all variables of interest in this setting.

For all exploratory statistical tests, a significance level alpha of 0.05 will be assumed.

The data analysis will be performed by using the SAS statistical package, version 9.4 or later.

All costs will be estimated per patient per year, and will be specified with respect to the different FVC % predicted level (FVC<50%, FVC 50-80%, FVC>80%) and total sample.

5.2.9. Handling of missing data

Missing values will not be imputed in the main analysis. However, if considering only observed values for an analysis is felt to bias the results post-hoc sensitivity analyses including imputation techniques might be applied.

For each questionnaire, its indications will be followed (for example, calculating the score considering the number of available items or missing all data). A missing response in either of the questionnaires will exclude the patient from the specific analyses.

In context of concomitant medication date in T0, missing data in the month will be imputed as June.

In context of pharmacological and non-pharmacological treatment related to FPI, if the initial treatment date is missing and the treatment is ongoing, the cost corresponding to a full year will be imputed (in case the diagnosis of the PFI is earlier than 2018).

No other missing data will be imputed. If patients have missing values for any other endpoints, these patients will be excluded from the analysis of that endpoint.

5.3. CHANGES IN PLANNED ANALYSES

No changes.