Study Protocol P2-C3-003

12/04/2024

Version 3.2

TABLE OF CONTENTS

DOCUMENT HISTORY

Version	Date	Description
V1.0	19/01/2024	**First draft for EMA review**
V2.0	21/02/2024	**Second draft including EMA's and industry comments**
V3.0	11/03/2024	**Third draft including EMA's comments**
V3.1	26/03/2024	**Inclusion of minor comments from EMA**
V3.2	12/04/2024	**Inclusion of Anton Barchuk to the study team**

Study Title	DARWIN EU® – Overall survival in patients with locally advanced or metastatic non-small cell lung cancer treated with selected immunotherapies as first line of treatment
Protocol version identifier	3.2
Date of last version of protocol	12/04/2024
EU PAS register number	EUPAS1000000112
Active substance	- Pembrolizumab - Nivolumab - Atezolizumab - Cemiplimab - Durvalumab - Ipilimumab - Chemotherapies (reference cohort for comparisons): cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine
Medicinal product	Keytruda, Opdivo, Tecentriq, Libtayo, Imfinzi, Yervoy
Research question and objectives	The **aim** of this study is to assess the overall survival of patients with locally advanced or metastatic NSCLC who initiate first-line treatment with selected immunotherapies (pembrolizumab, atezolizumab, cemiplimab, nivolumab, durvalumab and ipilimumab) and how it compares to the survival of locally advanced or metastatic NSCLC patients treated with chemotherapies as first line. The **specific objectives** of this study are: 1) To characterise patients at the time of initiating first line therapy as well as treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences. 2) To estimate the overall survival rates of patients with locally advanced or metastatic NSCLC **who initiated treatment** with immunotherapies (Pembrolizumab, Nivolumab, Atezolizumab, Cemiplimab, Durvalumab, Ipilimumab) and also with chemotherapies (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine) given **as monotherapy or in combination (as per the label) and as first line of**

	treatment, regardless of treatment discontinuation or treatment switch. 3) To compare the overall survival of patients with locally advanced or metastatic NSCLC under each immunotherapy to that of chemotherapy (reference cohort).
Country(-ies) of study	France, Spain, The Netherlands
Author	Talita Duarte-Salles

LIST OF ABBREVIATIONS

Abbreviation	Name
CDM	Common Data Model
CDWBordeaux	Clinical Data Warehouse of Bordeaux University Hospital
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DOI	Declaration Of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic Health Record
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
GDPR	General Data Protection Regulation
HR	Hazard Ratios
ICI	Immune checkpoint inhibitor
ICD-9-CM	International Classification of Diseases
IMASIS	Institut Municipal Assistència Sanitària Information System
IP	Inpatient
IRB	Institutional Review Board
LSPS	Large-scale propensity scores
MDRR	Minimum detectable rate ratio
NCR	Netherlands Cancer Registry
NOS	Not otherwise specified
NSCLC	Non-small cell lung cancer
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
PCT	Primary Care Teams
PD-L1	Programmed Death-Ligand 1
PSMar	Parc Salut Mar
RMST	Restricted Mean Survival Time
SD	Standard deviation
SNOMED	Systematized Nomenclature of Medicine
TNM	Tumour, nodes, metastasis
WHO	World Health Organisation

Abbreviation	Name
UICC	Union for International Cancer Control

1. TITLE

DARWIN EU® – Overall survival in patients with locally advanced or metastatic non-small cell lung cancer treated with selected immunotherapies as first line of treatment

2. RESPONSIBLE PARTIES – STUDY TEAM

Study team Role	Names	Organisation
Study Project Manager/Principal Investigator	Talita Duarte-Salles	Erasmus MC
Clinical Epidemiologists	Dina Vojinovic Julieta Politi Anton Barchuk*	IQVIA Erasmus MC Erasmus MC
Statistician	Maria de Ridder	Erasmus MC
Data Scientist	Maarten van Kessel Ross Williams	Erasmus MC Erasmus MC
Data Partner	**Names**	**Organisation – Database**
Local Study Coordinator/Data Analyst	Miguel-Angel Mayer Angela Leis Juan Manuel Ramirez Peter Prinsen Jelle Evers Ronald Damhuis Romain Griffier Guillaume Verdy	PSMAR - IMASIS PSMAR - IMASIS PSMAR - IMASIS Netherlands Cancer Registry Netherlands Cancer Registry Netherlands Cancer Registry University of Bordeaux - CDWBordeaux University of Bordeaux - CDWBordeaux

*Included to the study team on the 12th of April 2024.

3. ABSTRACT

Title

DARWIN EU® – Overall survival in patients with locally advanced or metastatic non-small cell lung cancer treated with selected immunotherapies as first line of treatment

Rationale and Background

Over the last decade, a better understanding of the molecular mechanism of lung cancer has led to the development of new therapies resulting in improvement in overall survival, mostly driven by advances in the treatment of non-small cell lung cancer (NSCLC). NSCLC has been recognized as a set of multiple diseases; therefore, numerous approved targeted therapies are now available in driver mutation positive NSCLC, and immunotherapies in the form of immune checkpoint inhibitor(s) (ICIs) are indicated in patients without a driver mutation.

While there is evidence on the clinical efficacy of these immunotherapies, there is still uncertainty on the benefits for a more diverse patient population treated outside clinical trials. A better understanding of the effectiveness of these medicines in real-life, which is the aim of this study, can help inform health technology assessment. This is particularly important considering the high costs of these immunotherapies, their increasing widespread use and population aging coupled with high incidence in older age groups.

Research question and Objectives

The **overall aim** of this study is to assess the overall survival of patients with locally advanced or metastatic NSCLC who initiate first-line treatment with selected immunotherapies (pembrolizumab, atezolizumab, cemiplimab, nivolumab, durvalumab, and ipilimumab) and how it compares to the survival of locally advanced or metastatic NSCLC patients treated with chemotherapies as first line.

The **specific objectives** of this study are:

1) To characterise patients at the time of initiating the first line therapy as well as to characterise treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences.

2) To estimate the overall survival rates of patients with locally advanced or metastatic NSCLC **who initiated treatment** with immunotherapies (Pembrolizumab, Nivolumab, Atezolizumab, Cemiplimab, Durvalumab, Ipilimumab) and also with chemotherapies (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine) given **as monotherapy or in combination (as per the label) and as first line of treatment**, regardless of treatment discontinuation or treatment switch.

3) To compare the overall survival of patients with locally advanced or metastatic NSCLC under each immunotherapy to that of chemotherapy (reference cohort).

Research Methods

An interim report will be developed upon completion of the first objective to characterise treatments in the target patient population. Based on the results, the specific cohorts for objectives 2 and 3 will be established.

To address objectives 2 and 3 a target trial emulation approach will be used. The estimand of the target trial is defined as per the following attributes:

- Population: Patients with locally advanced or metastatic NSCLC.
- Treatments:
    - Pembrolizumab
    - Nivolumab
    - Atezolizumab
    - Cemiplimab
    - Durvalumab
    - Ipilimumab
    - Chemotherapies (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine)

  given as monotherapy or in combination (as per the label) and as first line of treatment.
- Variable/outcome: overall survival, i.e. time from initiation of treatment to death from any cause.
- Summary measure: The hazard ratio will be used for comparison between immunotherapy and chemotherapy treatment groups.
- Intercurrent events: treatment discontinuation and treatment switch. Both will be dealt with a 'treatment policy strategy', i.e. the interest lies on effectiveness of the above treatment regardless of treatment discontinuation and treatment switch.

Based on the above attributes, the estimand (precise research question of interest) targeted with objective 3 can be described as: what is hazard ratio of time to death from any cause in selected immunotherapies given as first line of treatment compared to chemotherapies given as first line of treatment regardless of treatment discontinuation or switch?

The statistical analysis will be based on a Cox regression model. All time at risk of patients in the cohorts will be used regardless of treatment discontinuation and switch.

In a supplemental estimand, the summary measure will be changed to restricted mean survival time (RMST) to 1, 2 and 3 years. Comparison with the comparator cohort will be based on the difference in RMST.

Study design

New user matched cohort study.

Population

Patients aged 18 or above with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies listed above between 01/01/2016 and 31/12/2022. One cohort will be created for each immunotherapy treatment (target cohort) and a cohort of new users of chemotherapies (comparator cohort).

Data sources

1. Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France
2. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain
3. Netherlands Cancer Registry (NCR), The Netherlands

Exposures

Initiation of the following first-line treatments after diagnosis of locally advanced or metastatic NSCLC: pembrolizumab, nivolumab, atezolizumab, cemiplimab, durvalumab, ipilimumab and chemotherapy (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine).

Primary outcome of interest

Overall survival since the start of therapy for locally advanced or metastatic NSCLC.

Follow-up

Participants will be followed in each cohort from therapy initiation date until date of death, loss to follow-up or end of data availability in each database.

Data analyses

All analyses will be conducted separately for each database, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data.

First, we will run cohort diagnostics to evaluate data availability and data quality in terms of identification of locally advanced or metastatic NSCLC as well as recording of cancer treatments of interest.

A minimum cell count of 5 will be used when reporting results, with any smaller counts reported as "<5" to comply with the database's privacy protection regulations. All analyses will be reported by database, overall and stratified by age and sex when possible (minimum cell count reached).

Objective 1: To characterise patients at the time of initiating the first-line therapy as well as to characterise treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences (including combinations), and sequences (first line, second line, etc.)

The number and percentage of patients receiving each of the pre-specified NSCLC treatment/s as monotherapy/combinations will be described at index date and including all treatments up to 42 days following index date, which will represent the first-line treatment. Additionally, sunburst plots and Sankey diagrams will be used to describe treatment combinations and sequences over time when available. This will also be used to inform and evaluate sample size for the conduction of objectives 2 and 3. The results of this analysis will be reported in an interim report.

Objective 2: To estimate the overall survival rates of patients who initiated treatments of interest given as monotherapy or in combination and as first line of treatment, regardless of treatment discontinuation or treatment switch.

Overall survival will be calculated for each study cohort using data on time at risk of death from any cause since start of therapy and the Kaplan-Meier method. Results will be reported as plots of the estimated survival curves as well as the estimated probability of survival at years 1, 2, and 3.

Objective 3: To compare the overall survival under each immunotherapy to that of chemotherapy (reference cohort).

We will use a propensity score-matched cohort design, where target and comparator cohort participants will be matched to 1:1 based on propensity scores, and exact-matched on year of birth and calendar year of index date. Large-scale propensity scores (LSPS) will be estimated as the probability of exposure (target cohort) conditional on all covariates available in the data with a prevalence >1%, which will have been previously described in the results of cohort diagnostics as described above. LSPS will be estimated using Lasso regression. Hazard Ratios (HR) and 95% confidence intervals for overall survival will be estimated using Cox proportional hazards models comparing the target vs comparator (reference) cohorts after LSPS matching. Kaplan-Meier plots and/or cumulative incidence plots will be used to illustrate survival. The RMST to 1 and 2 years will also be calculated and differences in RMST between target and comparator cohorts will be provided.

4. AMENDMENTS AND UPDATES

Number	Date	Section of study protocol	Amendment or update	Reason

5. MILESTONES

STUDY SPECIFIC DELIVERABLE	TIMELINE
Draft Study Protocol	18/01/2024
Final Study Protocol	February 2024
Creation of Analytical code	February-March 2024
Execution of Analytical Code on the data	April/May 2024
Interim Study Report	May/June 2024
Draft Study Report	To be confirmed
Final Study Report	To be confirmed

6. RATIONALE AND BACKGROUND

Over the last decade, a better understanding of the molecular mechanism of lung cancer has led to the development of new therapies resulting in improvement in overall survival, mostly driven by advances in the treatment of non-small cell lung cancer (NSCLC) (1, 2). NSCLC has been recognized as a set of multiple diseases; therefore, numerous approved targeted therapies are now available in driver mutation positive NSCLC (2), and immunotherapies in the form of immune checkpoint inhibitor(s) (ICIs) are indicated in patients without a driver mutation (2).

The first ICIs used in treatment of lung cancer was in the form of programmed cell death ligand-1 (PD-1) inhibitor nivolumab as second-line therapy for patients with advanced NSCLC. Randomized phase III trials showed higher objective response rate and overall survival with nivolumab compared to docetaxel in patients with advanced squamous and non-squamous NSCLC following progression on platinum-based chemotherapy (2-4). Thereafter, another PD-1 inhibitor pembrolizumab and PD-L1 inhibitor atezolizumab were approved

for the same indication, based on higher efficacy of these agents compared to docetaxel in second-line setting (2, 5, 6).

The success of ICIs in second-line settings led to their use in first-line treatment of advanced NSCLC. Several phase III clinical trials showing durable responses and improvement in overall survival with ICI or ICI plus platinum-based chemotherapy compared to chemotherapy alone have rapidly expanded first-line treatment options for patients with advanced NSCLC not harbouring sensitizing EGFR mutations or ALK translocations (2). These options include pembrolizumab (7), atezolizumab (8), cemiplimab (9), nivolumab (10, 11), and durvalumab (12), which are indicated as monotherapy or in combination with other treatments and/or platinum-based chemotherapy (See Annex I, Table 3).

While there is evidence on the clinical efficacy of these immunotherapies, there is still uncertainty on the benefits for a more diverse patient population treated outside clinical trials (13, 14), as well as the potential differences in effectiveness by immunotherapies related to important effect modifiers (15). A better understanding of the effectiveness of these medicines in real-life, which is the aim of this study, can help inform health technology assessment. This is particularly important considering the high costs of these immunotherapies, their increasing widespread use and population aging coupled with high incidence in older age groups.

7. RESEARCH QUESTION AND OBJECTIVES

The **overall aim** of this study is to assess the overall survival of patients with locally advanced or metastatic NSCLC who initiate first-line treatment with selected immunotherapies (Pembrolizumab, Atezolizumab, Cemiplimab, Nivolumab, Durvalumab, Ipilimumab) and how it compares to the survival of locally advanced or metastatic NSCLC patients treated with chemotherapies as first line.

The **specific objectives** of this study are:

1) To characterise patients at the time of initiating the first-line therapy as well as to characterise treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences.

2) To estimate the overall survival rates of patients with locally advanced or metastatic NSCLC who initiated treatment with immunotherapies (Pembrolizumab, Nivolumab, Atezolizumab, Cemiplimab, Durvalumab, Ipilimumab) and also with chemotherapies (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine) given **as monotherapy or in combination (as per the label) and as the first line of treatment**, regardless of treatment discontinuation or treatment switch.
3) To compare the overall survival of patients with locally advanced or metastatic NSCLC under each immunotherapy to that of chemotherapy (reference cohort).

Table 7.1: Primary and secondary research questions and objective

A. Objective 1

Objective:	To characterise patients at the time of initiating the first-line therapy as well as to characterise treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences.

Hypothesis:	N/A
Population ***(mention key inclusion-exclusion criteria):***	Patients aged 18 or above with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies of interest between 01/01/2016 and 31/12/2022 will be included.
Exposure:	Initiation of the following treatments after diagnosis of locally advanced or metastatic NSCLC: pembrolizumab, nivolumab, atezolizumab, cemiplimab, durvalumab, Ipilimumab, and chemotherapy (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine).
Comparator:	N/A
Outcome:	N/A
Time ***(when follow up begins and ends):***	The Index date will be the start of the first NSCLC treatment following the first locally advanced or metastatic NSCLC diagnosis and follow-up to the date of death, loss to follow-up, or end of follow-up period, whichever occurs first.
Setting:	Inpatient records from CDWBordeaux [France] and IMASIS [Spain], and cancer registry records from NCR [The Netherlands].
Main measure of effect:	Proportions of patients on treatment types, combinations and sequences.

B. Objective 2

Objective:	To estimate the overall survival rates of patients with locally advanced or metastatic NSCLC who initiated treatment with Pembrolizumab, Nivolumab, Atezolizumab, Cemiplimab, Durvalumab, Ipilimumab, Chemotherapies (reference cohort for comparisons; cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine) given as monotherapy or in combination and as first line of treatment, regardless of treatment discontinuation or treatment switch.
Hypothesis:	N/A
Population ***(mention key inclusion-exclusion criteria):***	Patients aged 18 or above with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies of interest between 01/01/2016 and 31/12/2022 will be included. One cohort will be created for each treatment of interest, which could be given as monotherapy or in combination. The specific cohorts will be decided based on results from objective 1, with considerations on sufficient sample size.
Exposure:	Initiation of the following first-line treatments (index date) after diagnosis of locally advanced or metastatic NSCLC given as monotherapy or in combination: pembrolizumab, nivolumab, atezolizumab, cemiplimab,

	durvalumab, Ipilimumab, and chemotherapy (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine).
Comparator:	N/A
Outcome:	Overall, 1-, 2-, and 3-year survival since start of first-line therapy of locally advanced or metastatic NSCLC
Time *(when follow up begins and ends):*	Participants will be followed in each cohort from therapy initiation date until date of death, loss to follow-up or end of data availability in each database.
Setting:	Inpatient records from CDWBordeaux [France] and IMASIS [Spain], and cancer registry records from NCR [The Netherlands].
Main measure of effect:	Proportions and probability of survival

C. Objective 3

Objective:	To compare the overall survival of patients with locally advanced or metastatic NSCLC under each immunotherapy to that of chemotherapy (reference cohort).
Hypothesis:	N/A
Population *(mention key inclusion-exclusion criteria):*	Patients aged 18 or above with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies of interest between 01/01/2016 and 31/12/2022 will be included. Cohorts defined in objective 2 will be included.
Exposure:	Initiation of the following first-line immunotherapy treatments (index date) after diagnosis of locally advanced or metastatic NSCLC: pembrolizumab, nivolumab, atezolizumab, cemiplimab, durvalumab, ipilimumab.
Comparator:	Initiation of the following first-line chemotherapy treatments (index date) after diagnosis of locally advanced or metastatic NSCLC: cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine.
Outcome:	Overall survival since the start of first-line therapy of locally advanced or metastatic NSCLC
Time *(when follow up begins and ends):*	Participants will be followed in each cohort from therapy initiation date until date of death, loss to follow-up or end of data availability, whichever occurs first, in each database.
Setting:	Inpatient records from CDWBordeaux [France] and IMASIS [Spain], and cancer registry records from NCR [The Netherlands].

Main measure of effect:	Kaplan-Meier plots and cumulative incidence plots, Hazard Ratios and 95% confidence intervals.

8. RESEARCH METHODS

8.1 Study type and Study Design

Table 8.1. Description of Potential Study Types and Related Study Designs

STUDY TYPE	STUDY DESIGN	STUDY CLASSIFICATION
Comparative Effectiveness Studies	New User Cohort	Complex

An interim report will be developed upon completion of the first objective to characterise treatments in the target patient population. Based on the results, the specific cohorts for objectives 2 and 3 will be established.

To address objectives 2 and 3 a target trial emulation approach will be used. The estimand of the target trial is defined as per the following attributes:

- Population: Patients with locally advanced or metastatic NSCLC.
- Treatments:
    - Pembrolizumab
    - Nivolumab
    - Atezolizumab
    - Cemiplimab
    - Durvalumab
    - Ipilimumab
    - Chemotherapies (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine)

  given as monotherapy or in combination (as per the label) and as first line of treatment.
- Variable/outcome: overall survival, i.e. time from initiation of treatment to death from any cause.
- Summary measure: The hazard ratio will be used for comparison between the immunotherapy and chemotherapy treatment groups.
- Intercurrent events: treatment discontinuation and treatment switch. Both will be dealt with a 'treatment policy strategy', i.e. the interest lies on effectiveness of the above treatments regardless of treatment discontinuation and treatment switch.

Based on the above attributes, the estimand (precise research question of interest) targeted with objectives 2 and 3 can be described as: what is hazard ratio of time to death from any cause in selected immunotherapies given as first line of treatment compared to chemotherapies given as first line of treatment regardless of treatment discontinuation or switch?

The statistical analysis will be based on a Cox regression model. All time at risk of patients in the cohorts will be used regardless of treatment discontinuation and switch.

In a supplemental estimand, the summary measure will be changed to restricted mean survival time (RMST) to 1, 2 and 3 years. Comparison with the comparator cohort will be based on the difference in RMST.

8.2 Study Setting and Data Sources

This study will be conducted using routinely collected data from 3 databases in 3 European countries. The selection of databases for this study was performed based on the ability to identify patients with locally advanced or metastatic NSCLC as well as cancer treatments and date of death. All databases were previously mapped to the OMOP CDM.

At the time of writing the present study protocol, these are the databases for this study identified from the current network of data partners of DARWIN EU:

1. Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France
2. Institut Municipal Assistencia Sanitaria Information System (IMASIS), Spain
3. Netherlands Cancer Registry (NCR), The Netherlands

Information on selected data sources are described in a **Table 8.2**.

When it comes to assessing the reliability of data sources, the data partners are asked to describe their internal data quality process on the source data as part of the DARWIN EU onboarding procedure. To further ensure data quality, we utilize the Achilles tool, which systematically characterises the data and presents it in a dashboard format that is inspected. The generated data characteristics such as age distribution, condition prevalence per year, data density, measurement value distribution can be compared against expectations for the data. Additionally, the data quality dashboard (DQD) provides more objective checks on plausibility consistently across the data sources. In terms of relevance, a more general-purpose diagnostic tool, CohortDiagnostics, was developed. This package evaluates phenotype algorithms for OMOP CDM datasets, offering a standard set of analytics for understanding patient capture including data generation. It provides additional insights into cohort characteristics, record counts and index event misclassification. Furthermore, timeliness is guarded by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it is important to have clear understanding of the time period covered by each released database, as this can vary across different domains. To facilitate this, the CdmOnboarding (and Achilles) packages contain a 'data density' plot. This plot displays the number of records per OMOP domain on a monthly basis. This allows to get insights when data collection started, when new sources of data were added and when until when data was included.

Table 8.2. Description of the selected Data Sources.

Country	Name of Database	Justification for Inclusion	Health Care setting (e.g. primary care, specialist care, hospital care)	Type of Data (EHR, claims, registries)	Number of active subjects	Data lock for the last update
France	CDWBordeaux	Covers secondary care setting, database has information on cancer treatment, mortality and other outcomes for in-house patients.	Secondary care (in and outpatients)	EHR	1.9 million	12/2023
Spain	IMASIS	Covers secondary care setting, database has information on cancer treatment, mortality and other outcomes for in-house patients.	Secondary care (in and outpatients)	EHR	0.6 million	12/2023
The Netherlands	NCR	Cancer registry data with high quality information on cancer diagnoses, mortality, and cancer treatment.	Cancer registry	Registry	3.5 million	10/01/2024 with incident cancer patients included up to 31/12/2022

CWDBordeaux = Clinical Data Warehouse of Bordeaux University Hospital, IMASIS = Institut Municipal Assistencia Sanitaria Information System, NCR = Netherlands Cancer Registry.

Institut Municipal Assistència Sanitària Information System (IMASIS), Spain

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation. Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, which are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information more than 1 million patients with at least one diagnosis and who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient in years is 6.37 (SD±6.82). IMASIS-2 is the anonymized relational database of IMASIS which is used for mapping to OMOP including additional sources of information such as the Tumours Registry. It contains structured data related to diagnosis, procedures, drug administration, and laboratory tests and clinical annotations in a free-text format.

Clinical Data Warehouse of Bordeaux University Hospital (CDWBordeaux), France

The clinical data warehouse of the Bordeaux University Hospital comprises electronic health records on more than 2 million patients with data collection starting in 2005. The hospital complex is made up of three main sites and comprises a total of 3,041 beds (2021 figures). The Bordeaux University Hospital serves as the primary public health facility for the entire population of the Bordeaux metropolitan area. Additionally, it functions as a referral and expertise center for the Nouvelle Aquitaine region. The database currently holds information about the person (demographics), visits (inpatient and outpatient), conditions and procedures (billing codes), drugs (outpatient prescriptions and inpatient orders and administrations), measurements (laboratory tests and vital signs) and dates of death (in and out-hospital deaths)(9). Deaths in this database have two sources. First, this database retrieves in-hospital deaths. Second, patient records are regularly linked to data from the national death registry (every six months) using probabilistic algorithms based on search engines and machine learning strategies, with satisfactory results (16). Still, some deaths could go undetected, thus producing an underestimation of the event, given that records are not matched by using a common identifier between the two data sources, such as the social security number.

Netherlands Cancer Registry (NCR), The Netherlands

The NCR compiles clinical data of all individuals newly diagnosed with cancer in the Netherlands. Cancer registration clerks register newly diagnosed cancer patients since 1989 on a national basis, with 3 million patients included. Over the past 35 years, this registry has provided clinicians and researchers with a wealth of clinical data (e.g., patient and tumour characteristics, primary treatment, outcome) on cancer patients of all ages. Specifically, it also comprises information on tumour staging (according to the TNM-classification developed and maintained by the Union for International Cancer Control (UICC)), tumour site (topography) and morphology (histology) (according to the WHO International Classification of Diseases for Oncology (ICD-O-3)), co-morbidity at diagnosis and treatment received directly after diagnosis (first line). Overall, patients are followed-up for less than one year, with the exception of death which is collected any time after diagnosis. See https://iknl.nl/en for more information.

8.3 Study Period

The study period will start on 01/01/2016 and will end on the 31/12/2022 for the inclusion of cases, with follow-up extended until last date of data availability in each database.

8.4 Follow-up

Participants will be followed in each cohort from therapy initiation date until date of death (from any cause), loss to follow-up, end of study date, or end of data availability, whichever occurs first, in each database. Clinical information recorded prior to the start of follow-up will be used to model large-scale propensity scores at index date. A minimum time of data availability of 30 days post treatment initiation will be applied to allow time to capture treatments. And a minimum of one year of potential follow-up time will be required for objectives 2 and 3.

Table 8.3: Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type	Diagnosis position	Incident with respect to…	Measurement characteristics / validation	Source of algorithm
Adult patients with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies of interest	Date of initiation of the first-line treatments after diagnosis of locally advanced or metastatic NSCLC	Single entry	Incident	[-∞, ID]	IP and OT	RxNorm	N/A	Specific medication	N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

8.5 Study Population with inclusion and exclusion criteria

Patients aged 18 or above with locally advanced or metastatic NSCLC who initiated first-line treatment with any of the therapies of interest between 01/01/2016 and 31/12/2022 will be included. One cohort will be created for each immunotherapy treatment (target cohorts) as well as a cohort of new users of chemotherapies (comparator cohort). The specific treatment cohorts will be determined based on the results from the first objective.

Inclusion criteria

Inclusion criteria are the following:

- Individuals aged 18 or above.
- Primary diagnosis of NSCLC, including the following morphological types: adenocarcinoma, squamous cell carcinoma, large-cell carcinoma, and all subtypes.
- Initial presentation with locally advanced or metastatic disease based on AJCC/UICC TNM classification (IIIB or IV stage).
- Patients with specific drug treatments for NSCLC registered within the 6 months following NSCLC diagnosis
- A minimum of 30 days of follow-up post-diagnosis of locally advanced or metastatic NSCLC

Preliminary code lists for locally advanced and metastatic NSCLC are available in Appendix I of this protocol (**Table 2**).

Exclusion criteria

Exclusion criteria are the following:

- Primary diagnosis of small-cell lung cancer.
- Patients with any record of major lung surgery 6 months before the primary diagnosis of NSCLC (index date) and until the start of initial IO/chemotherapy treatment.
- Stage I-IIIA lung cancer.
- Any cancer diagnosis prior to date of NSCLC diagnosis.
- Patients with no drug treatment registered within the 6 months following NSCLC diagnosis

A total period of 42 days following initial treatment for locally advanced or metastatic NSCLC, following initial diagnosis of locally advanced or metastatic NSCLC will be applied to allow time to capture all first-line treatments. For objectives 2 and 3, a minimum of one year of **potential follow-up time** will be required. This ensures that we have a minimum time of data availability to identify outcomes.

Operational definitions of Inclusion and Exclusion Criteria are provided in **Tables 8.4** and **8.5.**

Table 8.4. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Age	Age in years defined by (time 0 – year of birth)/365	Before	N/A	N/A	Years	N/A	All individuals within the selected databases	N/A	N/A
Locally advanced or metastatic NSCLC	See definition of locally advanced, metastasis and NSCLC in section 8.5	Before	2016-2022	IP and OT	SNOMED and TNM	N/A		N/A	N/A
Initiation of immunotherapy	Pembrolizumab, nivolumab, atezolizumab, cemiplimab, durvalumab, ipilimumab	After	Anytime post index	IP and OT	RXNorm	N/A		N/A	N/A
Initiation of chemotherapy	Cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine	After	Anytime post index	IP and OT	RXNorm	N/A		N/A	N/A
Minimum time period	A minimum period of 30 days post-diagnosis of locally	After	30 days	IP and OT	N/A	N/A		N/A	N/A

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
	advanced or metastatic NSCLC will be applied to allow time to capture treatments								
Minimum potential follow-up time (objectives 2 and 3)	Only participants with a treatment initiated one year prior to end of data availability in the database will be included	After	1 year	IP and OT	N/A	N/A		N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

Table 8.5. Operational Definitions of Exclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
History of cancer diagnosis	Patients will be excluded if they had a diagnosis of small cell lung cancer or another primary tumour prior to their NSCLC diagnosis	Before	Any time prior to locally advanced or metastatic NSCLC diagnosis	IP and OT	SNOMED	N/A	All study population	N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

8.6 Variables

8.6.1. Exposures

Initiation of the following first-line treatments (index date) after diagnosis of locally advanced or metastatic NSCLC given as monotherapy or in combination (as per the label) and as first line of treatment, regardless of treatment discontinuation or treatment switch: pembrolizumab, nivolumab, atezolizumab, cemiplimab, durvalumab, ipilimumab, and chemotherapy (cisplatin, carboplatin, pemetrexed, paclitaxel, docetaxel, gemcitabine, and vinorelbine). Preliminary code lists for all exposures are available in Appendix I of this protocol.

The drug treatment of cancer is usually based on the administration of more than one antineoplastic drug administered during a time frame of a theoretical maximum of 21 days (i.e., one treatment cycle) repeated over several cycles, and is considered as the first-line treatment regimen. However, hospital-based databases rarely record first-line regimens as such, making it difficult to ascertain the initial treatment regimen from treatment switches or modifications due to disease progression. In turn, when first-line treatment is not recorded as such within databases, we will define first-line treatment regimens as all treatments that are started between the first treatment day (index date) and including all additional therapies started within the following 42 days (which is the time for two cycles of standard chemotherapy to be completed, and after which the initial assessment of treatment effect is usually performed) (See **Figure 1**).

Operational definitions of exposures are provided in **Table 8.6.**

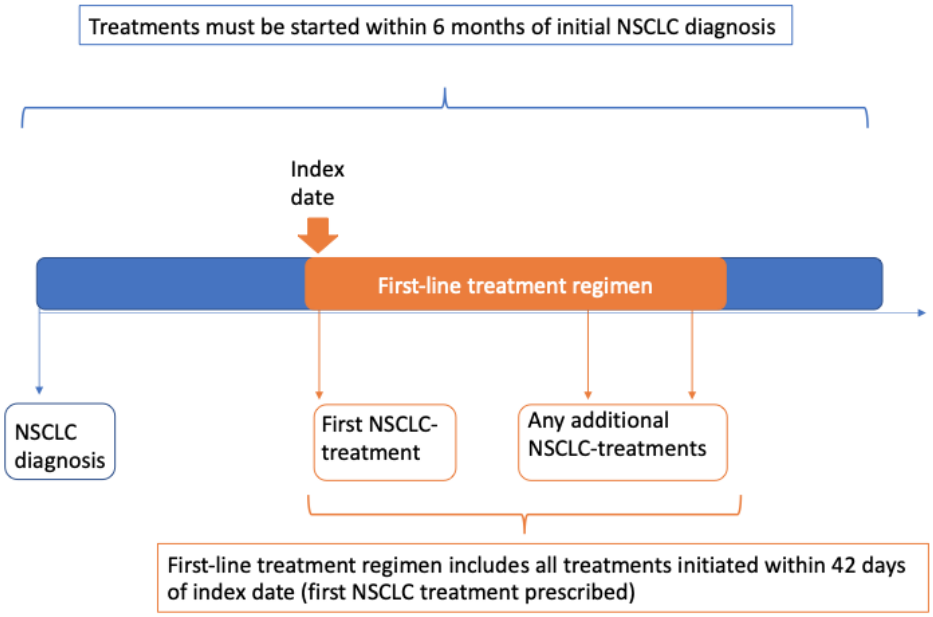

Figure 1. Timeline from NSCLC diagnosis, to index date, and first-line treatment definition.

Table 8.6. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type	Diagnosis position	Applied to study populations:	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Immunotherapies cohorts	Preliminary code lists provided in Table 1 of Appendix I	N/A	Anytime post index	IP and OT	RxNorm	N/A	All study population	Locally advanced or metastatic NSCLC	N/A	N/A
Chemotherapies cohorts	Preliminary code lists provided in Table 1 of Appendix I	N/A	Anytime post index	IP and OT	RxNorm	N/A	All study population	Locally advanced or metastatic NSCLC	N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

8.6.2. Outcome

The primary outcome of interest will be overall survival since start of first-line treatment for locally advanced or metastatic NSCLC, which will be calculated based on date of death (**Table 8.7**). Individuals will contribute with survival time as per the follow-up described in section 8.4.

Table 8.7. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Washout window	Care Settings[1]	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics/ validation	Source of algorithm
Overall survival	Overal survival since date of start of first-line treatment	Yes	Time	N/A	IP and OP	Date of Death	N/A	All study participants	N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

8.6.3. Other covariates, including confounders, effect modifiers and other variables

Demographics: age at the index date following the first treatment for locally advanced or metastatic NSCLC and sex (male/ female) will be described. The following age grouping will be used: 18-64; 65-80; 80 and over.

Health conditions pre-index date: history of the co-morbidities will be identified over three time periods prior to the index date: 1) 30 days prior to one day prior index date, 2) 365 days prior to one day prior index date, 3) all available days observed up to one day prior to index date. This information will be used for summary characterisation and calculation of large-scale propensity scores. A range of health conditions will be assessed using the time windows above.

Medications pre-index date: Pre-existing medication use will be identified using 2-time windows, defined as 365 days to one day prior to index date, and 30 days to 1 day prior to index date, and will be used to provide summary characterisation for patients and calculation of large-scale propensity scores.

Other information: Eastern Cooperative Oncology Group (ECOG)/WHO Performance Scale (concept_ids 4190931, 4161577, 4161578, 4162590, 4161579) and programmed death-ligand 1 (PD-L1) expression (concept_id 718584) at index date is available in NCR only.

Operational definitions of covariates are provided in **Table 8.8.**

Table 8.8. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Demographics	Age at index date and sex	Numeric, binary	All history	IP and OT	N/A	N/A	All study population	N/A	N/A
Health conditions	Conditions of interest prior to index date	Binary	[-365,-1], [-30 –1], All history prior to index date	IP and OT	SNOMED	N/A		N/A	N/A
Medication use	Drug prescriptions prior to index date	Binary	[-365,-1], [-30 –1]	IP and OT	RxNorm	N/A		N/A	N/A
Other information	ECOG/WHO Performance Scale and PD-L1 expression will be available in NCR only	Categorical, binary	At NSCLC diagnosis	IP and OT	N/A	N/A		N/A	N/A

[1] IP = inpatient, OT = other, n/a = not applicable

8.7 Study size

For each database, all individuals that satisfy the eligibility criteria for a study cohort will be included.

For objective 1, based on exploratory feasibility counts, the total number of subjects across the three data sources will be around 253,000 subjects.

For objective 3, **Table 8.9** provides some information on the precision (based on the width of the 95% confidence intervals (CI)) for the effect size comparing treatment groups for mortality. Because there is no closed-form sample size formula available related to the width of a CI for a hazard ratio (HR), calculations are done for the mortality rate ratio (MRR) as proxy for the HR. These calculations are based on assumptions of a MRR comparing immunotherapy versus chemotherapy of 0.74 (17) and a more conservative MRR of 0.85 which might be expected in real-world data. With a median survival in the chemotherapy group of 1.17 years (14 months) and assuming an exponential distribution of survival times, the mortality rate (MR) in the chemotherapy group is set to 0.59. The relative precision is defined as the difference between the upper limit of the 95% CI of the MRR and the estimated MRR, as percentage of the estimated MRR. Person-years needed in the immunotherapy group are calculated (18) and subsequently numbers of deaths. The sample size needed for the immunotherapy group is calculated based on participants surviving contributing on average 2.5 years of follow-up and participants dying contributing on average 1.42 years (17 months) of follow-up. The total sample size needed is twice the sample size in the immunotherapy group because the two comparison groups will be of equal size, as a result of the 1:1 LSPS matching.

Which numbers will be in the comparison groups can only be determined after the first phase of this study (Objective 1 and 2) has finished, so no statement on the CI widths can be given beforehand.

Table 8.9 Sample size needed for different levels of precision for assumed mortality rate ratios (MRR).

	Mortality Rates		95% CI							
Assumed MRR	Chemo therapy	Immunot herapy	Lower limit	Upper limit	Ratio	Relative precision (%)	Person-years in immuno	Deaths in immuno	Sample size in immuno	Sample size total
0.74	0.59	0.44	0.70	0.78	1.12	5.7	4,923	2,164	2,907	5,814
0.74	0.59	0.44	0.65	0.84	1.30	13.8	904	397	534	1,068
0.74	0.59	0.44	0.60	0.91	1.52	23.3	346	152	205	410
0.74	0.59	0.44	0.55	1.00	1.81	34.5	173	76	103	206
0.85	0.59	0.51	0.80	0.90	1.13	6.3	3,829	1,934	2,370	4,740
0.85	0.59	0.51	0.75	0.96	1.28	13.3	898	453	556	1,112
0.85	0.59	0.51	0.70	1.03	1.47	21.4	373	188	231	462
0.85	0.59	0.51	0.65	1.11	1.71	30.8	196	99	122	244

8.8 Analysis

All analyses will be conducted separately for each database, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data.

First, we will run cohort diagnostics to evaluate data availability and data quality in terms of identification of locally advanced or metastatic NSCLC as well as recording of cancer treatments of interest.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources and quality control checks will be performed. After all the tests are passed (see section 10 Quality Control), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

Data partners will locally execute the analytics against the OMOP-CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency.

All analyses will be reported by database, overall and stratified by age and sex when possible (minimum cell count reached). Results from objective 1 will further be stratified by calendar year.

8.8.1 Patient privacy protection

A minimum cell count of 5 will be used when reporting results, with any smaller counts reported as "<5" to comply with privacy protection regulations.

8.8.2 Descriptive statistics

For each analysis, summary descriptive analyses will be conducted including age, sex, key variables for matching and conditions and medication pre-index date for characterisation.

8.8.3 Main analysis per study objective

Objective 1: To characterise patients at the time of initiating the first-line therapy as well as to characterise treatments received by patients with locally advanced or metastatic NSCLC, including treatment combinations and sequences.

Large-scale patient-level characterisation will be conducted (objective 1). Age and sex at the time of NSCLC diagnosis will be described for each generated study cohort. The index date will be the date of first-line treatment initiation for NSCLC, following each patient's initial diagnosis of locally advanced or metastatic NSCLC. Medical history will be assessed for any time –and up to 365 days before the index date, between 365 to 31 days before the index date, between 30 to 1 day before the index date, and at the index date. Medication use history will be reported for the period between 365 and 31 days before the index date, between 30 and 1 day before, and at the index date. We will also report medication use for 1 to 30, 1 to 90, and 1 to 365 days post-index date. These time windows are defined based on the options currently available in the standard analytical tools that will be used for this project. Covariates to be presented in a summary baseline characteristics table will be pre-defined as described in section 8.6.3. and **Table 8.8**.

The number and percentage of patients receiving each of the pre-specified NSCLC treatment/s as monotherapy/combinations will be described at index date and including all treatments up to 42 days following index date, which will represent the first-line treatment. Additionally, sunburst plots and Sankey diagrams will be used to describe treatment combinations and sequences over time when available. This will

also be used to inform and evaluate sample size for the conduction of objectives 2 and 3. The results of this analysis will be reported in an interim report which will be used to inform feasibility of objectives 2 and 3. These results will also be included in the final report.

Please note that treatment sequences will not be reported in NCR since only first-line treatments are available in this database.

Objective 2: To estimate the overall survival rates of patients who initiated treatments of interest given as monotherapy or in combination and as first line of treatment, regardless of treatment discontinuation or treatment switch.

Overall survival from the time of therapy initiation will be estimated for each study cohort using data on time at risk of death from any cause and the Kaplan-Meier method. Results will be reported as plots of the estimated survival curves as well as the overall and median estimated probability of survival with 95% confidence intervals at years 1, 2 and 3. If sample size allows, we will stratify the results by histology and PD-L1 expression (in NCR only).

Objective 3: To compare the overall survival under each immunotherapy to that of chemotherapy (reference cohort).

The analysis results for objective 1 will be available in the interim report and will provide additional information to inform the feasibility and extent of analysis for objective 3. We will use a propensity score-matched cohort design, where target and comparator cohort participants will be matched to 1:1 based on propensity scores, and exact-matched on year of birth and calendar year of index date. Propensity score matching will proceed using nearest neighbour matching with nearest neighbour matching with a calliper width of 0.2.

Large-scale propensity scores will be estimated as the probability of exposure (target cohort) conditional on all covariates available in the data with a prevalence >1% (19), which will have been previously described in the results of cohort diagnostics as described above. LSPS will be estimated using Lasso regression. Hazard Ratios (HR) and 95% confidence intervals will be estimated using Cox proportional hazards models comparing the target vs comparator (reference) cohorts after LSPS matching. Kaplan-Meier plots and/or cumulative incidence plots will be used to summarise survival over time. Log-log plots will be visually inspected to identify potential violation of the proportional hazards assumption and will be reported...

In addition, the RMST at 1, 2 and 3 years will be calculated and reported, as well as differences in RMST between the target and reference cohorts.

Sensitivity analysis

Table 8.10. Description of Study Types and Type of analysis

STUDY TYPE	STUDY CLASSIFICATION	TYPE OF ANALYSIS
Comparative Effectiveness Studies	Complex	New cohort design: - Large-scale characterisation of participants in the target and comparator cohorts - Large-scale propensity scores (LSPS) will be estimated - Incidence rate/s of each of the outcomes of interest in the target and comparator cohorts - Diagnostic/s: Covariate balance, Equipoise, Power, residual confounding/systematic error (optional) - Rate Ratios or Hazard Ratio/s and 95% confidence intervals using Poisson or Cox models respectively

Table 8.11: Primary, secondary, and subgroup analysis specification

A. Primary analysis

Hypothesis:	Objective 1, 2 and 3: not applicable
Exposure contrast:	Objective 1: not applicable, descriptive treatment characterisation. Objective 2: not applicable. Objective 3: immunotherapy vs. chemotherapy (reference cohort)
Outcome:	Objective 1: not applicable Objective 2: overall survival Objective 3: overall survival
Analytic software:	R
Model(s): ***(provide details or code)***	Objective 1: Not applicable Objective 2: Overall survival from the time of therapy initiation will be estimated for each study cohort using data on time at risk of death from any cause and the Kaplan-Meier method. Objective 3: Hazard Ratios (HR) and 95% confidence intervals for overall survival will be estimated using Cox proportional hazards models comparing the target vs comparator (reference) cohorts after LSPS matching.
Confounding adjustment method	
	Objective 1-2: not applicable Objective 3: Among those participants in the target and comparator cohorts who met the inclusion criteria, target participants will be matched 1:1 to a comparator

	participant, based on year of birth, calendar year, and large-scale propensity scores using the nearest neighbour matching, with caliper width 0.2 standard deviations. Large-scale propensity scores (LSPS) will be estimated as the probability of exposure (target cohort) conditional on all covariates available in the data with a prevalence >1%, which will have been previously described in the results of cohort diagnostics as described above. LSPS will be estimated using Lasso regression.
Subgroup Analyses	
	All analyses will be reported by database, overall and stratified by age and sex when possible (minimum cell count reached). Results from objective 1 will further be stratified by calendar year.

8.9 Evidence synthesis

We will report analyses separately for each database. Additionally, we will combine the effect estimates across databases using random effect meta-analyses. Forest plots will be used to show results from meta-analyses.

9 DATA MANAGEMENT

9.1 Data management

All databases are mapped to the OMOP common data model. This enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM is developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org

The analytic code for this study will be written in R. Each data partner will execute the study code against their database containing patient-level data and will then return the results set which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report.

9.2 Data storage and protection

For this study, participants from various EU member states will process personal data from individuals which is collected in national/regional electronic health record databases. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All databases used in this study are already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and

adequate privacy control are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be run, which generate non-identifiable aggregate summary results.

The output files are stored in the DARWIN Digital Research Environment. These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures in order to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

10 QUALITY CONTROL

General database quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, it is expected that data partners will have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

When defining metastatic or locally advanced NSCLC, a systematic search of possible codes for inclusion will be identified using CodelistGenerator R package (https://github.com/darwin-eu/CodelistGenerator). This software allows the user to define a search strategy and using this will then query the vocabulary tables of the OMOP CDM so as to find potentially relevant codes. The codes returned will be then reviewed by clinical epidemiologists to consider their relevance. In addition, the CohortDiagnostics R package (https://github.com/OHDSI/CohortDiagnostics) will be run to assess the use of different codes across the databases contributing to the study and identify any codes potentially omitted in error. This will allow for a consideration of the validity of the defined study cohorts in each of the included databases, and inform decisions around whether multiple definitions are required.

The study code will be based on four R packages previously developed to (1) characterise demographic characteristics of study cohorts (PatientProfiles), (2) characterise treatments received by patients and their sequences (TreatmentPatterns), (3) estimate the overall survival rates of patients who initiate different treatments (CohortSurvival), and (4) estimate differences in overall survival between the different study cohorts (CohortMethod). These packages will include numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package for data analyses will be made publicly available via GitHub.

11 LIMITATIONS OF THE RESEARCH METHODS

The study will be informed by routinely collected health care data and so data quality issues must be considered. We will use LSPS to minimize measured confounding, and NCOs to assess potential residual confounding. However, given the observational nature of our data, we cannot rule out remaining confounding, which could partially account for findings in this study.

This study will be carried out following several assumptions that must be considered. First, it is important to note that we will rely on the assumption that patients treated with these therapies fall within the indication as per the label. Additionally, while the nature of the data may not be able to reflect patient characteristics to ascertain the adequateness of each treatment, there is interest in describing any patient with locally advanced or metastatic NSCLC treated with these therapies. Lastly, while excluding patients with specific mutations from this analysis for which target treatments exist (such as EGFR, ALK, ROS-1) would be ideal, it is also worth noting that such patients are usually treated with target therapies instead of immunotherapies, thus unlikely to be included within the patient sample.

Because some drugs are used for treatment for first- and second-line treatments, some residual misclassification between first-line and second-line therapy may be present. However, strict inclusion and exclusion criteria, together with the definition of first-line treatments, will likely mitigate this risk. Still, this should be taken into account when interpreting the results.

It should be noted that among Data Partners, there is different use of Concept IDs related to AJCC/UICC staging (6th, 7th and 8th versions). We will adopt site-specific staging classification versions used by each Data Partners. However, this may lead to possible misclassification issues related to the use of different versions, which should be taken into account.

There may be incomplete treatment exposures in NCR. This is because data on cancer treatments administered as part of a clinical trial cannot be shared for research in this database. Also, only first-line treatment is available in NCR; therefore, treatment sequences and combinations will not be reported for this database.

The recording of co-morbidities and medications pre-index may vary across databases. In NCR, no history of health conditions or non-cancer treatments is available. Therefore, the use of large-scale patient-level characterization and propensity score are not possible in this database. However, the ECOG/WHO performance status before start of cancer treatment and PD-L1 expression are available only in this database and will be taken into account in analysis.

In the IMASIS database, there is an established linkage between the electronic health records and the hospital Cancer Registry, facilitating the integration of pertinent data such as TNM staging, histology and dates of death. However, it should be noted that while in-hospital deaths are captured in IMASIS in real-time, this database currently lacks a direct connection with the population death registry which allows capturing out-of-hospital deaths. However, this is likely mitigated by the linkage between IMASIS and the hospital Cancer Registry, since the Cancer Registry is periodically linked to the death registry. Still, absence of direct linkage may result in potential delays in accurately reflecting mortality status updates for some patients within the system.

In CDWBordeaux, death is recorded through two sources. First, this database retrieves in-hospital deaths. Second, patient records are regularly linked to data from the national death registry (every six months) using probabilistic algorithms based on search engines and machine learning strategies, with satisfactory results (16). Still, some deaths could go undetected, thus producing an underestimation of the event, given that records are not matched by using a common identifier between the two data sources, such as the social security number.

Finally, the lack of pathology data for some patients might result in an underestimation of cases. However, the confirmation of tumour morphology is important to avoid misclassification of the study population.

12 MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions will not be collected or analysed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

Only in case of prospective data collection, there is a need to describe the procedures for the collection, management and reporting of individual cases of adverse events/adverse reactions.

13 GOVERNANCE BOARD ASPECTS

All data sources require approval from their respective IRB boards.

14 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

An interim report will be submitted to EMA by the DARWIN EU® CC upon completion of the first objective to characterise treatments in the target patient population. This will inform the feasibility of conducting objectives 2 and 3. If these are feasible, a final study report will be submitted including all study results.

An interactive dashboard incorporating all the results (tables and figures) will be provided alongside the PDF report. The full set of underlying aggregated data used in the dashboard will also be made available if requested.

Other dissemination activities to be undertaken will also include the creation of scientific publications and presentations at conferences.

15 OTHER ASPECTS

N/A

16 REFERENCES

1. Howlader N, Forjaz G, Mooradian MJ, Meza R, Kong CY, Cronin KA, et al. The Effect of Advances in Lung-Cancer Treatment on Population Mortality. N Engl J Med. 2020;383(7):640-9.

2. Mamdani H, Matosevic S, Khalid AB, Durm G, Jalal SI. Immunotherapy in Lung Cancer: Current Landscape and Future Directions. Front Immunol. 2022;13:823618.

3. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(17):1627-39.

4. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(2):123-35.

5. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387(10027):1540-50.

6. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389(10066):255-65.

7. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2016;375(19):1823-33.

8. Herbst RS, Giaccone G, de Marinis F, Reinmuth N, Vergnenegre A, Barrios CH, et al. Atezolizumab for First-Line Treatment of PD-L1-Selected Patients with NSCLC. N Engl J Med. 2020;383(14):1328-39.

9. Sezer A, Kilickap S, Gümüş M, Bondarenko I, Özgüroğlu M, Gogishvili M, et al. Cemiplimab monotherapy for first-line treatment of advanced non-small-cell lung cancer with PD-L1 of at least 50%: a multicentre, open-label, global, phase 3, randomised, controlled trial. Lancet. 2021;397(10274):592-604.

10. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, et al. Nivolumab plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer. N Engl J Med. 2019;381(21):2020-31.

11. Paz-Ares L, Ciuleanu TE, Cobo M, Schenker M, Zurawski B, Menezes J, et al. First-line nivolumab plus ipilimumab combined with two cycles of chemotherapy in patients with non-small-cell lung cancer (CheckMate 9LA): an international, randomised, open-label, phase 3 trial. Lancet Oncol. 2021;22(2):198-211.

12. Rizvi NA, Cho BC, Reinmuth N, Lee KH, Luft A, Ahn MJ, et al. Durvalumab With or Without Tremelimumab vs Standard Chemotherapy in First-line Treatment of Metastatic Non-Small Cell Lung Cancer: The MYSTIC Phase 3 Randomized Clinical Trial. JAMA Oncol. 2020;6(5):661-74.

13. Murteira R, Borges FC, Mendes GP, Ramos C, Ramos A, Soares P, et al. Real-world effectiveness of pembrolizumab in previously treated non-small cell lung cancer: A population-based cohort study. Pharmacoepidemiol Drug Saf. 2020;29(10):1295-302.

14. Matsumoto K, Shiroyama T, Tamiya M, Minami T, Kinehara Y, Tamiya A, et al. Real-world outcomes of nivolumab plus ipilimumab and pembrolizumab with platinum-based chemotherapy in advanced non-small cell lung cancer: a multicenter retrospective comparative study. Cancer Immunol Immunother. 2024;73(1):4.

15. Goring SM, Waser N, Varol N, Penrod JR, Yuan Y, Wang S. PCN39 TREATMENT EFFECT MODIFICATION OF IMMUNOTHERAPY-BASED REGIMENS IN FIRST-LINE (1L) ADVANCED NON-SMALL CELL LUNG CANCER

(ANSCLC): A SYSTEMATIC REVIEW OF RANDOMIZED CONTROLLED TRIALS (RCTS). Value in Health. 2020;23:S29.

16. Cossin S, Diouf S, Griffier R, Le Barrois d'Orgeval P, Diallo G, Jouhet V. Linkage of Hospital Records and Death Certificates by a Search Engine and Machine Learning. JAMIA Open. 2021;4(1):ooab005.

17. Reck M, De T, Paz-Ares L, Yuan Y, Chaudhary A, Lee A, et al. 1026P Treatment switching adjustment of overall survival in the CheckMate 227 clinical trial of nivolumab plus ipilimumab versus chemotherapy in first-line treatment of patients with advanced non-small cell lung cancer. Annals of Oncology. 2022;33:S1024.

18. Rothman KJ, Greenland S. Planning Study Size Based on Precision Rather Than Power. Epidemiology. 2018;29(5):599-603.

19. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health. 2010;13(2):273-7.

17 ANNEXES

Appendix I: Preliminary list of codes

Table 1. Preliminary medication code list

Concept name	Concept id*
Pembrolizumab	45775965
Nivolumab	45892628
Atezolizumab	42629079
Cemiplimab	35200783
Durvalumab	1594034
Ipilimumab	40238188
Cisplatin	1397599
Carboplatin	1344905
Pemetrexed	1304919
Paclitaxel	1378382
Docetaxel	1315942
Gemcitabine	1314924
Vinorelbine	1343346

*Including all descendants

Table 2. Preliminary code list for advanced and metastatic NSCLC

Concept name*	Concept id**
Metastatic non-small cell lung cancer	36684857
Non-small cell carcinoma of lung, TNM stage 4	4308479
Non-small cell lung cancer	4115276
Squamous non-small cell lung cancer	37109576
Large cell carcinoma of lung	4110589
Adenocarcinoma of lung	4112738
Non-small cell carcinoma of middle lobe, lung	44501471
Non-small cell carcinoma of lower lobe, lung	44500188
Non-small cell carcinoma of upper lobe, lung	44499422
Cancer modifier concept name	**Cancer modifier concept id**
TNM stage 3B	1634175
TNM stage 4	1633987
M1A	45882500
M1B	45881618
M1C	45878386

*Restricted to its advanced or metastatic types using the following cancer modifier concepts

**Including all descendants

Table 3. Dates of EMA approval (any indication and study-site specific) and indications for selected immunotherapies for first-line treatment of locally advanced NSCLC in Europe.

EMA (https://www.ema.europa.eu/en/homepage)	**Pembrolizumab**	**Nivolumab**	**Ipilimumab**	**Atezolizumab**	**Cemiplimab**	**Durvalumab**	***Tremelimumab***
	KEYTRUDA	Nivolumab BMS / Opdivo	Yervoy	Tecentriq	Libtayo	Imfinzi	*Tremelimumab AstraZeneca / IMJUDO*
Inital approval for any indication (CHMP)	29-7-2015	23-4-2015	19-5-2011	20-7-2017	26-4-2019	26-7-2018	*15-12-2022*
Indication	KEYTRUDA as monotherapy is indicated for the treatment of advanced (unresectable or metastatic) melanoma in adults	Opdivo as monotherapy is indicated for the treatment of advanced (unresectable or metastatic) melanoma in adults	"Yervoy is indicated for the treatment of advanced (unresectable or metastatic) melanoma in adults who have received prior therapy"	"Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic urothelial carcinoma (UC) after prior platinum-containing chemotherapy or who are considered cisplatin ineligible (see section 5.1). Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic non-small cell lung cancer (NSCLC) after prior chemotherapy. Patients with EGFR activating mutations or ALK-positive tumour mutations should also have received targeted therapy before receiving Tecentriq."	Libtayo as monotherapy is indicated for the treatment of adult patients with metastatic or locally advanced cutaneous squamous cell carcinoma who are not candidates for curative surgery or curative radiation	Imfinzi as monotherapy is indicated for the treatment of locally advanced, unresectable non-small cell lung cancer (NSCLC) in adults whose tumours express PD-L1 on ≥ 1% of tumour cells and whose disease has not progressed following platinum-based chemoradiation therapy.	*IMJUDO in combination with durvalumab is indicated for the first line treatment of adults with advanced or unresectable hepatocellular carcinoma (HCC). Tremelimumab AstraZeneca in combination with durvalumab and platinum-based chemotherapy is indicated for the first-line treatment of adults with metastatic non-small cell lung cancer (NSCLC) with no sensitising EGFR mutations or ALK positive mutations.*

Approval for study site specific indication (CHMP) (if different)	23-6-2016	21/5/2015 - 30/11/2015	17-9-2020	20-7-2017	20-5-2021	26-7-2018	*15-12-2022*
Indication	"KEYTRUDA is indicated for the treatment of locally advanced or metastatic non-small cell lung carcinoma (NSCLC) in adults whose tumours express PD-L1 and who have received at least one prior chemotherapy regimen. Patients with EGFR or ALK positive tumour mutations should also have received approved therapy for these mutations prior to receiving KEYTRUDA."	Nivolumab BMS is indicated for the treatment of locally advanced or metastatic squamous non-small cell lung cancer (NSCLC) after prior chemotherapy in adults".	YERVOY in combination with nivolumab and 2 cycles of platinum-based chemotherapy is indicated for the first-line treatment of metastatic non-small cell lung cancer in adults whose tumours have no sensitising EGFR mutation or ALK translocation.	"Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic urothelial carcinoma (UC) after prior platinum-containing chemotherapy or who are considered cisplatin ineligible (see section 5.1). Tecentriq as monotherapy is indicated for the treatment of adult patients with locally advanced or metastatic non-small cell lung cancer (NSCLC) after prior chemotherapy. Patients with EGFR activating mutations or ALK-positive tumour mutations should also have received targeted therapy before receiving Tecentriq."	Libtayo as monotherapy is indicated for the first-line treatment of adult patients with non-small cell lung cancer (NSCLC) expressing PD-L1 (in ≥ 50% tumour cells), with no EGFR, ALK or ROS1 aberrations, who have: • locally advanced NSCLC who are not candidates for definitive chemoradiation, or • metastatic NSCLC.	Imfinzi as monotherapy is indicated for the treatment of locally advanced, unresectable non-small cell lung cancer (NSCLC) in adults whose tumours express PD-L1 on ≥ 1% of tumour cells and whose disease has not progressed following platinum-based chemoradiation therapy.	*IMJUDO in combination with durvalumab is indicated for the first line treatment of adults with advanced or unresectable hepatocellular carcinoma (HCC). Tremelimumab AstraZeneca in combination with durvalumab and platinum-based chemotherapy is indicated for the first-line treatment of adults with metastatic non-small cell lung cancer (NSCLC) with no sensitising EGFR mutations or ALK positive mutations.*
Approval for study objective specific indication (CHMP) (if different)	**15-12-2016**	**17-9-2020**	**17-9-2020**	**31-1-2019**	**20-5-2021**	**26-7-2018**	***15-12-2022***

Indication	KEYTRUDA as monotherapy is indicated for the first-line treatment of metastatic non-small cell lung carcinoma (NSCLC) in adults whose tumours express PD-L1 with a ≥50% tumour proportion score (TPS) with no EGFR or ALK positive tumour mutations"	OPDIVO in combination with ipilimumab and 2 cycles of platinum-based chemotherapy is indicated for the first-line treatment of metastatic non-small cell lung cancer in adults whose tumours have no sensitising EGFR mutation or ALK translocation.	YERVOY in combination with nivolumab and 2 cycles of platinum-based chemotherapy is indicated for the first-line treatment of metastatic non-small cell lung cancer in adults whose tumours have no sensitising EGFR mutation or ALK translocation.	Tecentriq, in combination with bevacizumab, paclitaxel and carboplatin, is indicated for the first-line treatment of adult patients with metastatic non-squamous non-small cell lung cancer (NSCLC). In patients with EGFR mutant or ALK-positive NSCLC, Tecentriq, in combination with bevacizumab, paclitaxel and carboplatin, is indicated only after failure of appropriate targeted therapies	Libtayo as monotherapy is indicated for the first-line treatment of adult patients with non-small cell lung cancer (NSCLC) expressing PD-L1 (in ≥ 50% tumour cells), with no EGFR, ALK or ROS1 aberrations, who have: • locally advanced NSCLC who are not candidates for definitive chemoradiation, or • metastatic NSCLC.	Imfinzi as monotherapy is indicated for the treatment of locally advanced, unresectable non-small cell lung cancer (NSCLC) in adults whose tumours express PD-L1 on ≥ 1% of tumour cells and whose disease has not progressed following platinum-based chemoradiation therapy.	*IMJUDO in combination with durvalumab is indicated for the first line treatment of adults with advanced or unresectable hepatocellular carcinoma (HCC). Tremelimumab AstraZeneca in combination with durvalumab and platinum-based chemotherapy is indicated for the first-line treatment of adults with metastatic non-small cell lung cancer (NSCLC) with no sensitising EGFR mutations or ALK positive mutations.*

Appendix II: ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® – Overall survival in patients with locally advanced or metastatic non-small cell lung cancer treated with selected immunotherapies as first line of treatment

EU PAS Register® number: EUPAS1000000112
Study reference number (if applicable): 1000000112

EU PAS Register® number: EUPAS1000000112
Study reference number (if applicable): 1000000112

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1. Does the protocol specify timelines for 1.1.1 Start of data collection[1] 1.1.2 End of data collection[2] 1.1.3 Progress report(s) 1.1.4 Interim report(s) 1.1.5 Registration in the EU PAS Register® 1.1.6 Final report of study results.	X			5. Milestones, 8.2 Data Sources

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain: 2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue) 2.1.2 The objective(s) of the study? 2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised) 2.1.4 Which hypothesis(-es) is (are) to be tested? 2.1.5 If applicable, that there is no *a priori* hypothesis?	X			7. Research question and objectives 8. Research methods

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	X			8.1 Study type and Study Design
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	X			8.2 Study Setting and Data Sources
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	X			8.8 Analysis

3.4	Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	X			8.8 Analysis
3.5	Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)			X	

Comments:

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	X			8.5 Study Population
4.2	Is the planned study population defined in terms of:	X			
	4.2.1 Study time period				8.3 Study Period
	4.2.2 Age and sex				8.6.3. Other covariates
	4.2.3 Country of origin				8.2 Study Setting and Data Sources
	4.2.4 Disease/indication				8.6.1. Exposures
	4.2.5 Duration of follow-up				8.4 Follow-up
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	X			8.5 Study Population with inclusion and exclusion criteria

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	X			8.6.1. Exposures
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)			X	
5.3 Is exposure categorised according to time windows?	X			8.6.1. Exposures
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	X			8.6.1. Exposures
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?			X	
5.6 Is (are) (an) appropriate comparator(s) identified?	X			8.8 Analysis

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	X			8.6.2. Outcomes
6.2 Does the protocol describe how the outcomes are defined and measured?	X			8.6.2. Appendix I

6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)			X	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)			X	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	X			8.8 Analysis
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)			X	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	X			8.8 Analysis

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	X			8.8 Analysis

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	X			8.6.1. Exposures
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	X			8.6.2. Outcomes
9.1.3 Covariates and other characteristics?	X			8.6.3. Other covariates
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	X			8.2 Study Setting and Data Sources
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	X			8.2 Study Setting and Data Sources
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	X			8.2 Study Setting and Data Sources
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	X			8.6.1. Exposures

9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	X			8.6.2. Outcomes
9.3.3 Covariates and other characteristics?	X			8.6.3. Other covariates
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)			X	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	X			8.8 Analysis
10.2 Is study size and/or statistical precision estimated?			X	
10.3 Are descriptive analyses included?	X			8.8.2 Descriptive statistics
10.4 Are stratified analyses included?	X			8.8 Analysis
10.5 Does the plan describe methods for analytic control of confounding?	X			8.8 Analysis
10.6 Does the plan describe methods for analytic control of outcome misclassification?	X			8.8 Analysis
10.7 Does the plan describe methods for handling missing data?			X	
10.8 Are relevant sensitivity analyses described?	X			8.8 Analysis

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	X			9. Data management
11.2 Are methods of quality assurance described?	X			10. Quality Control
11.3 Is there a system in place for independent review of study results?			X	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of: 12.1.1 Selection bias? 12.1.2 Information bias? 12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	X			11. Limitations of the research methods
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	X			Table 8.2. Description of the selected Data Sources.

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	X			13. Governance board aspects
13.2 Has any outcome of an ethical review procedure been addressed?			X	
13.3 Have data protection requirements been described?	X			9.2 Data storage and protection

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	X			4. Amendments and updates

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	X			14. Plans for disseminating and communicating study results
15.2 Are plans described for disseminating study results externally, including publication?	X			14. Plans for disseminating and communicating study results

Comments:

Name of the main author of the protocol: Talita Duarte-Salles

Date: 19/01/2024

Signature: _____