Real-world effectiveness of different COVID-19 vaccines in Spain: a cohort study based on public electronic health records (BIFAP)

Study Information

Title	Real-world effectiveness of different COVID-19 vaccines in Spain: a cohort study based on public electronic health records (BIFAP)
Protocol version identifier	1.0
Date of last version of protocol	20 april 2021
EU PAS Register number	Study not yet registered
Active substance	1. Vacuna de ARNm frente a COVID‑19 (con nucleósidos modificados) 2. Vacuna de ARNm frente a COVID-19 (con nucleósidos modificados) 3. (ChAdOx1-S [recombinant]) made up of another virus (of the adenovirus family) 4. Recombinant, replication-incompetent adenovirus type 26 (Ad26) vectored COVID-19 vaccine (Ad26.COV2-S [recombinant])
Medicinal product	1. Comirnaty concentrado para dispersión inyectable 2. COVID-19 Vaccine Moderna, dispersión inyectable 3. VAXZEVRIA suspension inyectable 4. COVID-19 Vaccine Janssen suspension for injection
Product reference	-
Procedure number	-
Marketing authorisation holder(s)	1. Biontech Manufacturing GmbH 2. Moderna Biotech Spain 3. Astrazeneca 4. Janssen
Research question and objectives	To evaluate the effectiveness of COVID-19 vaccines (i.e. comirnaty, Moderna's, Vaxzevria and Janssen and subsequent authorised vaccines) in preventing COVID-19 diagnosis.
Country(-ies) of study	Spain
Author	Elisa Martín Merino The current protocol is based on recommendations included in the following protocol template: Layton B, Garcia X. et al. Real-world effectiveness of <<COVID-19 vaccine product>> in Europe: a protocol template for a cohort study based in existing health care data sources from the ACCESS project

Marketing authorisation holder(s)

Marketing authorisation holder(s)	NA
MAH contact person	NA

Contributions

Name & Organisation	Date	Contribution	Version of Document
Elisa Martin Merino. PhD Pharm. Epidemiologist. Division of Pharmacoepidemiology and Pharmacovigilance. Spanish Agency for Medicines and Medical Devices. AEMPS	22 February 2021	Creation of first protocol proposal including study design (time-varying cohort study controlling confusion and bias through restriction, stratification and adjustment), methodology and remaining sections of the protocol.	V1
	30 March 2021	Update study design: Matched comparative cohorts.	V2
	12 April 2021	Circulating with changes	V3
	20 April 2021	Final protocol	V3
Enrique Seco Meseguer. MD, Hospital La Paz Madrid	22 February 2021	Review of Clinical trials and advice for definitions for cut-off periods, confounders, outcomes and clinical restrictions.	V1
	18 March 2021	Comments on Updated study design	V2
Belén Castillo Cano. PhD candidate Mathematician. Division of Pharmacoepidemiology and Pharmacovigilance. Spanish Agency for Medicines and Medical Devices. AEMPS	30 March 2021 9 April 2021	Probabilities estimations	V1 V2
Aurora Limia Sánchez Carmen Olmedo Lucerón (SG Promoción, Prevención y Calidad; Dirección General de Salud Pública, Ministerio de Sanidad)	16 March 2021 9 April 2021	Comments to first protocol.	V1 V2
Susana Monge Corella (CCAES; Dirección General de Salud Pública, Ministerio de Sanidad)	16 March 2021	Comments to first protocol.	V1
Amparo Larrauri Clara Mazagatos Ateca	16 March 2021	Comments to first protocol. Agreement to changes provided in V2 protocol	V1

Trademarks

Brand Name	Generic Name	Trademark Holder
Comirnaty	Vacuna de ARNm frente a covid	Biontech Manufacturing GmbH
COVID-19 Vaccine Moderna	Vacuna de ARNm frente a covid	Moderna Biotech Spain
COVID-19 Vaccine Oxford	(ChAdOx1-S [recombinant]) made up of another virus (of the adenovirus family)	Astrazeneca
COVID-19 Vaccine Janssen	Recombinant, replication-incompetent adenovirus type 26 (Ad26) vectored	Janssen
Any other covid-19 vaccine approved during the study period		

1 Table of Contents

2 List of Abbreviations

AEMPS	Spanish Agency for Medicines and Medical Devices
BIFAP	Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria
CI	confidence interval
COVID-19	illness caused by the SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) virus
eHR	electronic health records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU PAS Register	European Union Electronic Register of Post-Authorisation Studies
GPP	Good Pharmacoepidemiology Practices
GVP	Good Pharmacovigilance Practices
HCU	health care utilisation
HR	hazard ratio
ICPC-2	*International Classification of Primary Care, 2nd Edition*
ICD-9	*International Classification of Diseases, 9th Revision*
ICD-10	*International Classification of Diseases, 10th Revision*
ICU	intensive care unit
IRR	incidence rate ratio
ISPE	International Society for Pharmacoepidemiology
PASS	postauthorisation safety study
RD	risk difference
RR	risk ratio
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology

3 Responsible Parties

[Principal investigator institution name]
Spanish Agency for Medicines and Medical Devices
Calle Campezo 1, Edificio 8, 1ª planta, Madrid
Elisa Martín Merino, PhD Pharm, Epidemiologist
Belen Castillo Cano, PhD Candidate, Mathematician
Ana Llorente-García, Computer Scientist

[Sponsor name]
[Address]
NA

Collaborating Institutions	**Study Sites**
SG Promoción, Prevención y Calidad; Dirección General de Salud Pública, Ministerio de Sanidad	Madrid
CCAES; Dirección General de Salud Pública, Ministerio de Sanidad	Madrid
Departamento de Enfermedades Transmisibles. Centro Nacional de Epidemiología. CIBER Epidemiología y Salud Pública (CIBERESP) Instituto de Salud Carlos III	Madrid

4 Abstract

Title: Real-world effectiveness of different COVID-19 vaccines in Spain: a cohort study based in primary health care (BIFAP) and linked data sources

Rationale and background: Multiple vaccine candidates for COVID-19 are progressing through development and testing with anticipated widespread adoption and use after approval. The real-world effectiveness of these vaccines must be evaluated in populations in Spain.

Research question and objectives: To evaluate the effectiveness of all COVID-19 vaccines administered in Spain in reducing the medically attended diagnosis of COVID-19 in any setting (confirmed through test and regardless the prognosis and clinical phenotype)

Secondary objectives of this study will be splitting the effectiveness

- by the following outcomes:
  - asymptomatic (defined as a positive test (whether PCR or antigens) without recorded symptoms in primary care +-14days of the date the sample was taken). Note: In epidemiological vigilance protocols, among asymptomatic patients, isolation must be maintained till 10 days after the date when the sample was taken.
  - symptomatic (defined as a positive test (whether PCR or antigens) with recorded symptoms in primary care or Hospitalisations/Intensive care unit (ICU) admissions +-14days of the date the sample was taken (or at anytime linked with the test result))
  - Hospitalisations/Intensive care unit (ICU) admissions for COVID-19 at anytime linked with the test result or within +-14days of the date the sample was taken (if hospitalisation was recorded in primary care clinical profile)
  - All-cause mortality as linked with the test result or within +-14days of the date the sample was taken (if hospitalisation was recorded in primary care clinical profile)
- over clinically meaningful subgroups (old people; patients with cardiovascular disease, diabetes, or chronic pulmonary disease, chronic renal impairment, treated cancer, patients with a history of transplantation, Down Syndrome found linked with the prognostic of the covid-19 infection; as well as smoking, males and obesity suspected linked with the prognosis in some studies)
- along the time since complete vaccination (3 months, 6 months and ever after).

Study design: This is an observational cohort study to compare the occurrence of covid-19 infection among unvaccinated and vaccinated individuals.

Population: Individual aged ≥18 years with at least 1 year of record with their primary care physician from December 2020 till April 2021.

Variables:

- COVID-19 vaccinations will be identified from a vaccination registry linked with each patient's primary care clinical history.
- Medically attended diagnosis of COVID-19 in any setting will be identified through linked registries of 1) COVID positive test results, 2) hospital COVID diagnosis or discharged COVID diagnosis.
- Covariates will be collected from primary care eHR. Covariates will include demographic information, morbidities, medication use, health care utilisation, markers of current disease status at time zero, and markers of frailty.

Data sources: Base de datos para la Investigacion farmacoepidemiológica en Atencion Primaria (BIFAP) and linked registries of 1) COVID positive test results, 2) hospital COVID diagnosis or 3) discharged COVID diagnosis

Study size: BIFAP database includes up-to-standard information for around 7.7 million patients aged ≥18 years.

Data analysis: Characteristics of the vaccinated and unvaccinated groups at baseline will be described. Incidence rates of COVID-19 outcomes will be calculated and compared between the two groups, and vaccine effectiveness measures will be estimated by 1 minus the hazard ratios (HR; 95%CI). Additional time period–specific effect estimates (e.g., at 3, 6, or >6 months after vaccination) will also be estimated.

Stratified analysis will separately estimate vaccine effectiveness in clinically meaningful subgroups and calendar periods. Sensitivity analyses will evaluate the robustness of the study approach across variations of the study methodology.

Milestones: Evaluation of the availability of data will start during 1-15 June but timelines may be impacted by the date the data are received from the different regions and BIFAP common data model is ready for evaluation (End of July).

5 Amendments and Updates

None to date.

6 Milestones and Timeline

Milestone	Date
Approvals by ethics and scientific review bodies	March 2021
Start of data collection[a]	End of June 2021
End of data collection[b]	July 2021
<Study progress report(s) 1>	August 2021
<Study progress report(s) 2>	September 2021
<Study progress report(s) 3>	December 2021
<Interim report 1>	
<Interim report 2>	
<Interim report 3>	
Registration in the EU PAS Register	April 2021
Final report of study results	April 2022

EU PAS Register = European Union Electronic Register of Post-Authorisation Studies.

Note: Timelines may be impacted by approvals of ethics and scientific review bodies, and availability of data and staff at research institutions once contracts and approvals are finalised.

[a] Start of data collection is "the date from which information on the first study subject is first recorded in the study data set or, in the case of secondary use of data, the date from which data extraction starts" [1].

[b] End of data collection is "the date from which the analytical data set is completely available" [1].

7 Rationale and Background

COVID-19, the disease caused by the novel coronavirus SARS-CoV-2, has become a global pandemic, affecting countries throughout Europe and the world. At the time of writing of this protocol, three vaccines are been administrated in Spain (i.e. Comirnaty and COVID-19 vaccine Moderna) and two approved (Astrazeneca-Oxford and Janssen's) after conditional marketing authorisation by EMA (after granting of the EU Commission on 21st December, 6th January, 29th January, and 11th March 2021 respectively). The real-world effectiveness of these vaccines must be evaluated in Spanish population.

This is an observational cohort study, including patient-level real-world evaluation of the effectiveness of the different COVID-19 vaccines administrated in Spain (not ecological evaluation). The secondary data included in the Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP; the Spanish Database for Pharmacoepidemiological Reseach in Primary Care) will be used [65, 67].

New emerging data on covid infections and vaccinations will not be validated since we will be working with the gold-standard data considered today (i.e. lab results registries for covid infection and vaccinations respectively).

BIFAP has been previously validated for research in pharmacoepidemiology, including the estimations of the precision of several clinical outcomes [65] that will be covariates of the current study (such as cardiovascular/digestive/endocrinological outcomes). Also, vaccination recorded in BIFAP's Table of Vaccines has also been validated [66] that will include the linked official covid vaccinations. The algorithm for maximizing the precision and sensitivity will be utilised for the covariates collection.

As receipt of a newly licenced vaccine may be strongly associated with risk status (e.g., health care workers or essential workers), health status (e.g., at high risk for complications), or lifestyle factors (e.g., adherence to recommendations regarding vaccination, hygiene, self-quarantine, social distancing, and/or mask-wearing), confounding is likely the strongest threat to the validity of an observational study [3-9].

We intent to address confusion through study design (e.g., restriction to scheduled people by calendar time, identifying exchangeable comparison groups/moments) and analysis (e.g., statistical adjustment and stratification).

This protocol is based on a template that outlined a general approach to designing and implementing a comparative study of the effectiveness of a newly approved COVID-19 vaccine using existing health care databases [68].

8 Research Question and Objectives

This study addresses the research question of whether vaccinations with new licenced COVID-19 vaccines, (Comirnaty, Moderna, AstraZeneca's and Janssen and subsequently

approved vaccines in the EU during the data collection), are effective in reducing the burden of COVID-19 in Spain in comparison with no vaccination periods.

Primary objective is to evaluate the effectiveness of each COVID-19 vaccine in reducing the covid-19 infections (confirmed through test and regardless the prognosis and clinical phenotype (i.e. whether symptomatic or asymptomatic, mild or severe).

Secondary objectives of this study will be splitting the effectiveness

- by the following outcomes:
  - asymptomatic (defined as a positive test (whether PCR or antigens) without recorded symptoms in primary care +-14days of the date the sample was taken). Note: In epidemiological vigilance protocols, among asymptomatic patients, isolation must be maintained till 10 days after the date when the sample was taken.
  - symptomatic (defined as positive test (whether PCR or antigens) with recorded symptoms in primary care or Hospitalisations/Intensive care unit (ICU) admissions +-14days of the date the sample was taken (or at anytime linked with the test result))
  - Hospitalisations/Intensive care unit (ICU) admissions for COVID-19 at anytime linked with the test result or within +-14days of the date the sample was taken (if hospitalisation was recorded in primary care clinical profile)
  - All-cause mortality as a recorded death with COVID 19 listed as an underlying cause of death on cause-of-death registers, or a death during a hospitalisation, with COVID-19 as a listed diagnosis, or a death within a specified time period (e.g., 28 days) after a diagnosis or hospitalisation for COVID 19.
- over clinically meaningful subgroups (old people; patients with cardiovascular disease, diabetes, or chronic pulmonary disease, chronic renal impairment, treated cancer, patients with a history of transplantation, Down Syndrome found linked with the prognostic of the covid-19 infection; as well as smoking, males and obesity suspected linked with the prognosis in some studies)
- along the time since complete vaccination (up-to 3 months, 6 months and ever after).

A feasibility evaluation will be conducted before the start of the study to evaluate the availability of key data elements and the appropriateness of the study design.

9 Research Methods

9.1 Study Design

An observational cohort study of people prioritized for covid-19 vaccination will be implemented to allow for the estimation of effect measures on the absolute scale

(incidence difference) (e.g., time point–specific difference) [10] and relative scale (e.g., time point–specific risk ratios, hazard ratios, and vaccine effectiveness measures). The occurrence of covid-19 infection among unvaccinated and vaccinated patients will be compared as defined below.

This study primarily describes an approach to the retrospective collection of data to using BIFAP data (and linked registry by patient's level) after the necessary data have accumulated (real-time; expected every 2 months starting in 21st December 2021).

The choice of an exchangeable comparator group is a key factor in ensuring a valid study design and addressing confounding. In the current protocol, individuals not yet receiving COVID 19 vaccines will be pooled for a random selection among those with matched criteria to the vaccinated individuals. We have discarded to select the following control groups due to the reason explained below:

- Identified at receipt of a different vaccine such as influenza: the period of influenza vaccination (September-December) seems to be fewer circulation covid-19 since fewer incidences were reported. Thus, they may not be comparable
- During a general practitioner visit: will not be as frequent or representative or a normal non-pandemic situation when people are able to visit the health setting without any restriction. Thus, they may not be comparable
- Choosing a fixed time unit (weeks for instance) to create a series of nested cohorts, each starting at each new time unit [20,21]. Similar to the selected option, this one is valid and although more precise it is more complicated to implement. Also, the variance estimation would need to be adjusted accordingly (e.g., by bootstrapping) because individuals are allowed to contribute to multiple study cohorts [22].

The inclusion and exclusion criteria for the study cohort and thus for the selection of comparative groups are described in the following sections.

9.1.1 Study Feasibility Stage

The feasibility of conducting a vaccine effectiveness study in BIFAP (electronic health records (eHR; i.e. secondary data) will be evaluated regarding:

1. data availability and data validity (Sections 9.1.1.1, 9.1.1.2, and 9.1.1.3)
2. valid study design (Section 9.1.1.4) during final protocol development.

9.1.1.1 Availability of Required Data Elements

We will determine that all data elements for the implementation of the study, including co-variates (see Section 9.3), inclusion and exclusion criteria (Section 9.2), vaccinations, covid-outcomes, and potential confounders (Sections 9.1.1.5 and 9.3.4) are available.

9.1.1.2 Validity of COVID-19 Test results and Diagnosis Coding

In this study, COVID-19 outcomes will be identified from the lab test results linked with PC eHR by the regional parties before sending to BIFAP to create its common data model.

COVID-19 Test results (PCR, antigens, etc.)

Test results for coronavirus will be linked with PC eHR. For effectiveness analysis, cases will be positive covid infections recorded in that registry that may be confirmed through positive PCR, Antigens or any other confirmatory criteria established by clinical protocols (that definition is out of the scope of the current study).

Laboratory-confirmed infection is considered gold-standard to identify cases for the main analysis, thus no predictive values will be estimated. As explained below, identification of so-called 'possible' covid cases recorded only through diagnosis coding (without lab information) will be performed in a Sensitivity analysis (Section 9.7.4.2).

Also, an estimation of the concordance between the observed incidences in BIFAP and those 'expected' will be performed (i.e. external validation) as a proxy for the accurateness to identify cases in BIFAP and discussion of potential reasons if differences would be present.

The 'expected' cumulative incidences will be calculated by dividing the COVID cases reported daily to RENAVE (base de datos de vigilancia epidemiológica; by calendar period, region and age) by the population reported in the Spanish National Statistics Institute for that year [73]. Access to RENAVE will be available by one of the institution collaborating in the current study [74].

9.1.1.3 Validity of Vaccine Exposure

The precision and sensitivity of COVID-19 vaccine records included in the BIFAP's Table of Vaccine are not known. However, high precision and sensitivity have been estimated for other vaccinations recorded in that table

- after manual revision (against free text comment for human papillomavirus vaccination)(2) or
- validated against expected coverages (as in the Spanish recommended schedules) or external sources (such as other European countries with public

vaccination programs) for vaccinations against human papillomavirus, pertussis, influenza or meningococcus (3).

The information about the COVID vaccinations included in that Table is the result of linking all vaccination registries with the public PC eHR. That link has been strongly recommended (and it is compulsory) by the Ministry of Health and it is used by the Ministry to provide the official coverages since the vaccination campaigns started.

Vaccination data for each patient will include the vaccination date, dose, name of the vaccine, brand and batch for each jab.

Even though, the vaccinations records of the initial priority population for vaccination (i.e. frontline health and social care workers, residents in a care home and patients with special needs) will be available in BIFAP, we will not study them since we expect not enough quality criteria of the baseline consultations to the PCP. Thus, as aforementioned the current study will be restricted to populations where both, vaccination status and baseline and updated clinical history, can be accurately identified.

9.1.1.4 Definition of Time Zero

In studies of vaccinated versus unvaccinated persons, researchers must determine the appropriate time point at which to identify patients and begin follow-up. Aligning the evaluation of eligibility criteria, covariate assessment, exposure assignment, and beginning of follow-up (time zero) avoids selection bias and immortal person-time bias and addresses a well-defined study question [13].

Time zero is the time when the vaccination status is assigned; all eligibility criteria must be fulfilled at that point, and COVID outcomes must start to be counted at that point [14].

Time zero in the vaccinated group will be the date when the 1st dose of COVID-19 vaccine is administered (i.e. recorded). That date will be also used for the unvaccinated pairs as Time zero.

The aim of the **Main analysis** will be the estimation of the effectiveness ≥15 days after full vaccination series whether 1- or 2-doses regimens (and regardless of delayed completion of 2-dose vaccine series ,i.e. recorded dates will be accepted as real world data even though the time between doses are not fully adjusted to recommended intervals). Analysis will be performed by vaccine brand.

In order to properly observe the moment when the effect of the vaccination starts, the period between 1st Dose and 14 days after the 2nd dose will be divided and compared with unvaccinated patients. Thus, at least 5 cut-of periods will be identified along each patient's follow-up contribution:

1) Non-exposed period among Unvaccinated patients (Reference for Effectiveness estimation)

2) D1-induction time: up-to 14 days after 1st dose

3) D1-effectiveness: 15 days after 1st dose to 2nd dose

4) D2-induction time: up-to 14 days after 2nd dose

5) D2-effectiveness in the short run: 15-90 days after 2nd dose

6) D2-effectiveness in the medium run: 91-180 days after 2nd dose

7) D2-effectiveness in the long run: ≥181 days after 2nd dose

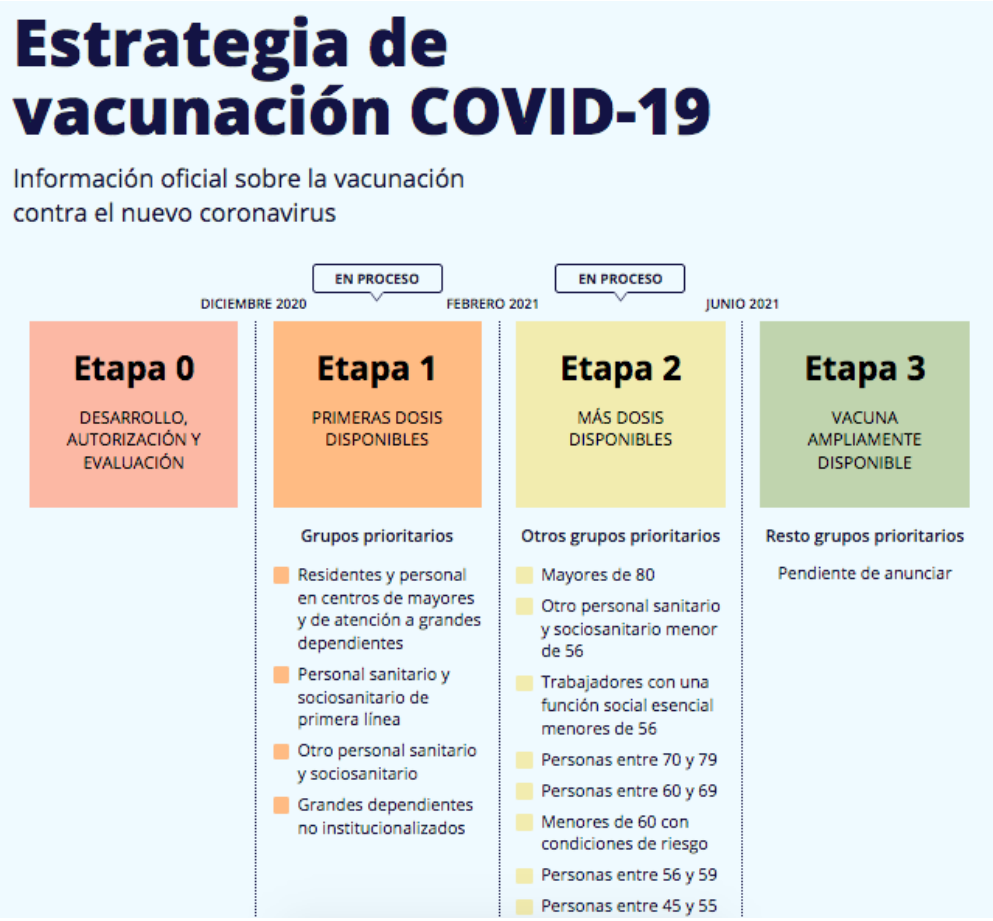
9.1.1.5 Assess Potential for Unmeasured Confounding

Unmeasured confounding is a serious threat to the estimation of vaccine effectiveness using observational data [3-9]. There may be situations where the initial recommendations for vaccination, the vaccine distribution strategy, and the uptake of the vaccine may heavily influence the magnitude and direction of confounding. Since unmeasured and unknown confusion may be present we will perform the following strategies in order to minimise then:

- unvaccinated comparator will be matched by time zero of the vaccinated pair, birth year cohort (year; it will be broad till 2-years if not enough patients to select), sex, region of the primary care practice. If available information, patients living at nursing home will be matched). Several other confounder will be controlled for through adjustment in the Cox model.
- Since multiple COVID-19 vaccines are available, recipients of different COVID-19 vaccines will be also compared with the same unvaccinated comparator and among them in a head-to-head fashion. This choice would greatly diminish, if not eliminate, confounding by indication.

9.2 Setting

This study will be conducted in Spain where the 1st COVID-19 vaccine was included in public campaigns after approval by the European Commission from 21st December 2020. The eligibility criteria will change over time according to prioritised vaccination groups as follows in Figure 1:



Due to the potentially changing incidence of COVID-19 over time and prioritised vaccination groups, comparisons within the study periods and prioritised criteria will be established as much as possible.

According to the 1st prioritized groups (see Figure above), the **study period** will begin on the 21st of December 2020. The end of study period will be the last date of data received at the start of the data collection. Since the vaccines start to request the vaccination for young people (for example: Pfizer requested to the FDA to expand use of its Covid-19 vaccine to adolescents ages 12 to 15 on 9th April 2021), no age restriction will be applied for the source population.

No sex restriction will be applied.

In BIFAP, as per quality and quantity criteria for research, and in order to ensure precise anamnesis, only those clinical histories considered up-to-standard are included in pharmacoepidemiological studies.

9.3 Variables

The measurement of exposure status, eligibility criteria, outcomes, and covariates defined below will be defined in the data source(s) and subsequently used in analysis.

9.3.1 Exposure Assessment and Time At Risk

Data will be collected from BIFAP database with data from Spain (described in Data Sources).

Inclusion criteria to the Source population: up-to-standard, >=1 year baseline information in January 2021 in BIFAP. The last date of filling all the inclusion criteria will be the inclusion date. No restriction to age or sex will be applied.

Exclusion criteria: Patients with prevalent COVID vaccinations or infections ever before (i.e. washout period) the inclusion criteria will be excluded.

Stop date: the last moment of the following criteria will define the stop date to contribution: covid-19 infection, death, loss to follow-up and end of study period (the last date available in the data source at data collection).

Exposure assessment: The 1st vaccination recorded after the inclusion date will be the time zero for patients included in the vaccinated group, as explained in section 'Definition of Time Zero' 9.1.1.4.

From the source population, 10 individuals unvaccinated on or before time zero of the vaccinated pair (10:1) will be randomly selected among those with similar birth cohort, sex and region (matched criteria). Patients included in the comparison group, may be vaccinated later during their follow-up period and then contribute to the vaccinated group (from the date of vaccine administration ahead; i.e. time zero for the vaccinated group).

*Note: All individuals vaccinated during their follow-up contribution will be included in the vaccinated group. While, a random selection of the unvaccinated individuals will be included as unvaccinated group.

The date and type of vaccine administrated will be collected from the **Table of Vaccine** from BIFAP data model. In particular, Covid-19 vaccinations will be ascertained from official covid-19 vaccination registries and provided by the regional governments. The coding system it not known yet. The feasibility analysis will include an evaluation of the coding type, whether or not vaccines are recorded according to standardised system or not. We expect that either ATC07 (J07????) or National Code (CNF [Código Nacional de

Farmacia]) specific for pharmaceuticals in Spain is available for all vaccines administered.

For the vaccines currently available, those with a 2-dose regimen, the second dose must be administrated 21/28 days after the 1st dose. In order to identify potential data **entry errors** (e.g., two vaccination records too closed or 3 doses recorded, may indicate double recording of the same vaccination) and cut-off points, a description of interval between doses will be performed beforehand.

Current vaccines are intended to give protection at least for 21 and 28 days after the 2nd dose, however the protection in the median-long run is unknown. Thus, in order to estimate the duration of the effectiveness, the follow-up period after 2nd dose will be divided as aforementioned.

The **follow-up** period will be the time between time zero and stop date.

9.3.2 Eligibility Criteria

No additional eligibility criteria are needed to define the cohort.

9.3.3 Outcome Assessment

Outcomes will be assessed during the follow-up period (Section 9.3.1). Multiple effectiveness outcomes associated with vaccination for COVID-19 are proposed in this protocol, as the severity of COVID-19 and its complications may vary widely. The severity of COVID 19 illness has been shown to vary from asymptomatic to life-threatening.

9.3.3.1 Laboratory-confirmed COVID-19

As Main outcome: cases will be incident **laboratory-confirmed covid-19** infections recorded in a covid-19 registry (regardless the symptoms, severity or the result of the COVID infection). Infection may be confirmed through positive PCR, Antigens or any other confirmatory criteria established by clinical protocols (that definition is out of the scope of the current study). That registry will be linked with BIFAP's primary care electronic health records (PC eHR). The date of the COVID cases will be the 1st date the sample was taken for testing as recorded in the lab registries linked to PC eHR.

9.3.3.2 Symptomatic or Asymptomatic Diagnosis of COVID-19

In a secondary analysis, in order to inform about the effectiveness to protect against asymptomatic and symptomatic infections, outcomes from the main analysis will be divided according to the symptoms recorded around the date of confirmed covid-19 in the patient PC eHR. Two effectiveness parameter, i.e. to protect against asymptomatic and asymptomatic infection, will be estimated separately for each vaccine. The date of the COVID cases will be the 1st date the sample was taken for testing.

Cases of **Asymptomatic** infection will be defined as a positive test without recorded symptoms in primary care +-14days, hospitalisations/Intensive care unit (ICU) or hospital diagnosis of the date the sample was taken). In asymptomatic analysis, symptomatic infections will participate as non-cases stopping at covid date.

Cases of **Symptomatic** infection will be defined as s positive test with recorded symptoms in primary care +-14days of the date the sample was taken, or linked Hospitalisations/Intensive care unit (ICU) admissions. In symptomatic analysis, asymptomatic infections will participate as non-cases stopping at covid date.

Note: In epidemiological vigilance protocols, among asymptomatic patients, isolation must be maintained till 10 days after the date when the sample was taken. For a sensitive approach we will broad till 14 days in order to allow for delayed records in PC eHR.

*Note: some misclassification may be present since patients with mild disease (mild symptoms) may never seek care that will be assumed as asymptomatic.

9.3.3.3 Severe COVID-19 Hospitalisation for COVID-19

In a secondary analysis, the outcome will be incident covid-19 infections resulting in an admission or discharge diagnosis to hospital (inpatient record) or intensive care unit as the setting of care. Admissions will be identified through linked (at anytime) with the test result or within +-14days of the date the sample was taken (if hospitalisation was recorded in primary care clinical profile). Those type of identification will depend on the region providing the data to BIFAP. If not available for all regions, restriction to regions will be performed.

The date of the outcome will be the admission date.

In that analysis, remaining non-severe covid-19 infections will participate as non-cases stopping at covid date.

9.3.3.4 COVID-19–Specific Mortality

The data availability on the cause of death will depend on the region providing the data to BIFAP, thus the primary outcome of COVID-19–specific mortality will have the following hierarchical definition

- A recorded death with COVID-19 listed as an underlying cause of death on cause-of-death registers
- Death during a hospitalisation, with COVID-19 as a listed diagnosis
- Death within a specified time period (e.g., up-to 8 weeks [73]) after a diagnosis or hospitalisation for COVID-19

The date of the outcome will be the date of death.

9.3.3.5 All-Cause Mortality

Cause-of-death information in BIFAP may be lacking depending on the region providing data; therefore, all-cause mortality will be used as a secondary outcome. All-cause mortality will be defined as a recorded patient death in primary care, or hospital records during the outcome ascertainment period of up-to 8 weeks after the date of the COVID date for main analysis (the 1st date the sample was taken for testing as recorded in the lab registries linked to PC eHR if available). The date of the outcome will be the date of death.

9.3.4 Covariate Assessment

In order to control for measurable confounders in the analysis, the following variables will be collected: lifestyle characteristics (BMI, smoking alcohol use), comorbidities, comedications, and health care utilisation prior to or at time zero and updated every 21 days and at every change in the exposure period (i.e. when receiving D1, D2) at the patient level. Given that risk factors for COVID-19 infection are not currently well understood [30], we will measure covariates that are classically controlled for in pharmacoepidemiologic research and leaves room for the inclusion of risk factors yet to be determined.

As explained in Data source, clinical events will be collected by using the International Classification of Primary Care (ICPC) and the International Classification of Diseases 9th Revision (ICD-9) medical terms (6,7).

The following variables will all be assessed as last status before time zero:

- Smoking status
- Body mass index
- Alcohol use

The following comorbidities (and others that may be shown to be associated with COVID-19 prognosis) will be assessed ever before time zero and updated:

- Diabetes mellitus (types 1 and 2)
- Hypertension record
- Cardiovascular disease
- Cerebrovascular disease
- Chronic respiratory disease
- Chronic kidney disease
- Chronic liver disease
- Cancer

- Immunodeficiencies including Human immunodeficiency virus and other immunosuppressing conditions
- Autoimmune disorders

Comedication use will be assessed through pharmacy dispensing and primary care physician prescribing records by ATC. Considered comedications may be indicative of comorbidities placing patients at higher risk or markers of health care–seeking behaviour and utilisation and may serve as markers of patients' history of other infections or be risk factors for severe COVID-19 disease themselves. The following comedications will be assessed during the year before time zero and updated:

- Antibiotics
- Antiviral medications
- Corticosteroids
- Non-steroidal anti-inflammatory drugs
- Other Analgesic
- Psychotropics
- Statins
- Immunosuppressant medication use

Health care utilisation in the year before time zero and updated will be evaluated as measures of health care–seeking behaviour, overall health status, and access to health care. Additionally, short-term health care utilisation in the 2 weeks before and including time zero and updated will be recorded separately, as short-term markers of current health status may influence individuals' vaccination decisions. Considered variables will include the following:

- Primary care utilisation (number of visits to PC)
- Influenza vaccination (number in the previous 5 years)
- Other vaccinations (number in the previous 5 years)
- COVID-19 tests (total number)

Frailty has been demonstrated as a confounder of vaccine-outcome associations in older adults [5,6,9]. Additional personal characteristics demonstrated to be associated with frailty [6,31,32] should be included as confounding factors, and may include the following:

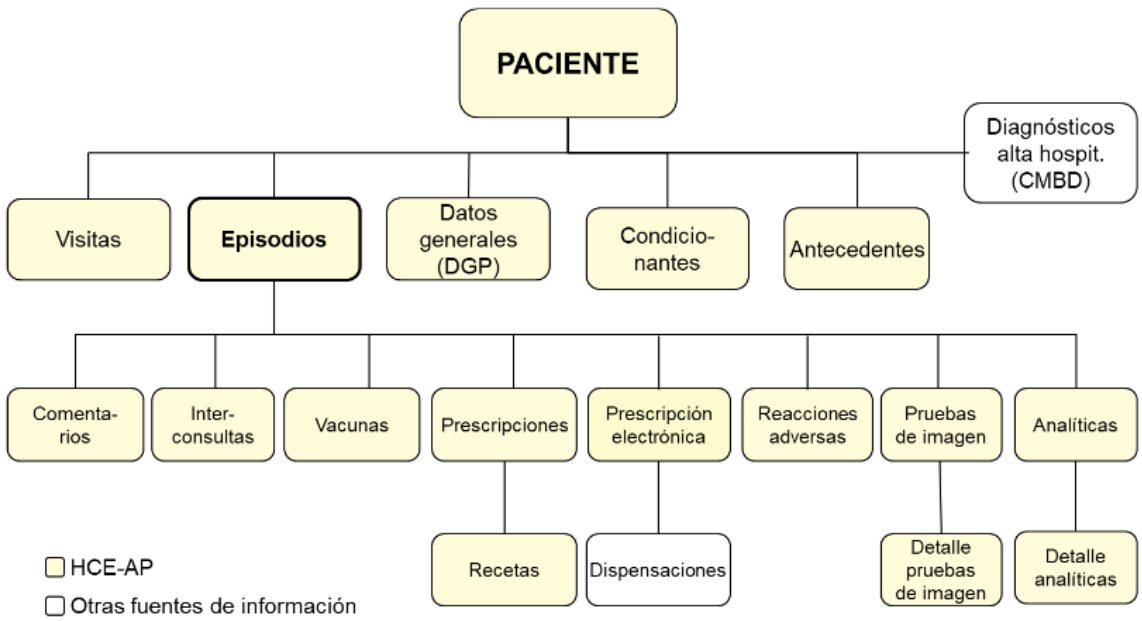
- Wheelchair use
- Home hospital bed
- Paralysis
- Parkinson's disease
- Skin ulcer

- Weakness
- Stroke/brain injury
- Ambulance transport
- Dementia
- Difficulty walking
- Home oxygen
- Rehabilitation care
- Psychiatric illness
- Sepsis
- Heart failure
- Podiatric care
- Bladder continence
- Diabetes complications
- Arthritis
- Coagulation deficiencies
- Vertigo
- Lipid abnormalities

9.4 Data Source

Data on patients will be obtained from primary care outpatient anonymised clinical records, dispensation in pharmacies of primary care prescriptions and diagnosis at hospital discharge from the Spanish public National Health System (SNS). Those data banks are unified in the data source so-called "Base de datos para la investigación Farmacoepidemiológica en Atención Primaria" (BIFAP) (8,9).

BIFAP includes information about patients' age, sex, life-style factors (i.e. body mass index (BMI), smoking status, alcohol consumption), clinical events (recorded by using



the International Classification of Primary Care (ICPC) and the International Classification of Diseases 9th Revision (ICD-9) medical terms (6,7)), anonymised primary care physicians' (PCPs) free-text notes, specialist referrals and discharge letters, prescriptions issued in primary care and their dispensations in pharmacies, vaccinations, laboratory test results and diagnosis at hospital discharge (recorded through the RAE-CMBD system) of around 13.7 million patients attended in a public primary care setting (7.4 of them aged ≥18 years). BIFAP covers 7 out of 19 districts in Spain and 57% of their population overall (as described in the Spanish Statistical Office (10)). In addition, currently, data on covid-19 infections resulting in an admission or discharge diagnosis to hospital (inpatient record) or intensive care unit as the setting of care are being incorporating to BIFAP. Figure 2:

BIFAP is fully funded by the Spanish Agency on Medicines and Medical Devices (AEMPS), belonging to the public Department of Health, and is maintained with the collaboration of the participant regions (8).

The records on the results of the COVID-19 laboratory test will be linked with BIFAP primary care profiles. Also, all vaccinations administrated against COVID-19 recorded in any setting will be linked with the primary care history in regions providing data. Particularly, vaccinations will be entered in the BIFAP's Table of Vaccines that in 2019 included around 61 million of recorded vaccinations.

BIFAP has been previously validated for research in pharmacoepidemiology, including the estimations of the precision of several clinical outcomes (9,11–13) that will be covariates of the current study (such as cardiovascular/digestive/endocrinological outcomes). Also, vaccination recorded in BIFAP's Table of Vaccines has also been confirmed valid through formal precision studied or against external comparable data sources (3,4).

9.5 Study Size

All patients meeting the eligibility criteria in BIFAP will be included as source population. BIFAP includes information of around 7.7 million patients aged ≥18 years (as last updated version in 2018), which distribution by sex and age is as follows:

	TOTAL	18 - 19	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45-49	50 - 54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-110
TOTAL	7720843	252143	424740	465933	526002	647688	747020	717918	702323	658868	569616	493770	454787	350359	317844	246643	145189
Women	3991612	122805	209807	232836	265565	324008	369475	357547	352106	330858	290309	255848	242742	194252	187908	154500	101046
Men	3729231	129338	214933	233097	260437	323680	377545	360371	350217	328010	279307	237922	212045	156107	129936	92143	44143

Taking into account the estimated COVID vaccine effectiveness reported in clinical trials [75-78], varying assumptions of vaccine unexposed to exposed ratio, anticipated global sample sizes of patients aged ≥18 years in BIFAP, a cumulative positive symptomatic COVID cases of 60,57 per 100,000 people in 14-days in Spain (as reported on 22 March 2021 [when started the vaccination for general people, i.e. apart from health workers and nursing home residents] in www.cnecovid.isciii.es/covid19/), [33], the different probabilities that the upper limit of the 95% CI of the risk ratio (RR) being below 1.00 (a correlate of the lower bound of the vaccine effectiveness estimate being above 0.00 [Section 9.7.3.1] demonstrating a protective effect of vaccination) was estimated and displayed in Table 1:

Table 1. Study Size Precision Estimates overall ages

Type of vaccine	Expected Vaccine Effectiveness[a]	Expected Ratio of Vaccine Unexposed to Exposed	Expected Sample Size (Exposed + Unexposed)	Expected Risk of the Outcome in the Unexposed[b]	Probability of the Upper Limit of the 95% CI to Be Below 1.00
Pfizer/Biontech (BNT162b2)	95% (90.3-97.6) → RR= 0.05	1:1	7,720,843 people	0.0006057	1
Pfizer/Biontech (BNT162b2)	95% (90.3-97.6) → RR=0.05	2:1	7,720,843 people	0.0006057	1

Moderna (mRNA-1273)	94.1% (89.3-96.8) → RR=0.059	1:1	7,720,843 people	0.0006057	1
Moderna (mRNA-1273)	94.1% (89.3-96.8) → RR=0.059	2:1	7,720,843 people	0.0006057	1
Oxford/Astrazeneca (ChAdOx1 nCoV-19)	70.4% (54.8-80.6)→ RR=0.296	1:1	7,720,843 people	0.0006057	1
Oxford/Astrazeneca (ChAdOx1 nCoV-19)	70.4% (54.8-80.6)→ RR=0.296	2:1	7,720,843 people	0.0006057	1
Janssen (we use the 14-days effectiveness against mild-severe for 1 dose)	66.9% (59.0-73.4)→ RR=0.331	1:1	7,720,843 people	0.0006057	1
Janssen (we use the 14-days effectiveness against mild-severe for 1 dose)	66.9% (59.0-73.4)→ RR=0.331	2:1	7,720,843 people	0.0006057	1

CI = confidence interval; RR = relative risk.

Source: Rothman, K. Episheet: spreadsheets for the analysis of epidemiologic data. 2015. Available at: http://www.krothman.org/episheet.xls. Accessed March 2021.

[a] Vaccine effectiveness measured as 1 minus the RR, where the RR compared the risk of the outcome in vaccine-exposed versus that in unexposed individuals (i.e., an RR below 1 indicates a protective effect of the vaccine, corresponding to positive vaccine effectiveness).

[b] Cumulative symptomatic COVID cases of 60,57 per 100,000 people in 14-days in Spain (as reported on 22 March) expressed as cases per person in population without scaling.

Similarly, by age group the probabilities estimated were 1 (Table 2):

Table 2. Study Size Precision Estimates by age categories

Type of vaccine	Expected Vaccine Effectiveness[a]	Expected Ratio of Vaccine Unexposed to Exposed	Expected Sample Size (Exposed + Unexposed)	Expected Risk of the Outcome in the Unexposed[b]	Probability of the UL95% CI to Be <1.00
Pfizer/Biontech 16-55 years	95.6% (89.4-98.6) → RR= 0.044	1:1	4,483,767 people	0.0723	1
Pfizer/Biontech >=55 years	93.7% (80.6-98.8) → RR= 0.063	1:1	3,237,076 people	0.0562	1
Pfizer/Biontech 16-55 years	95.6% (89.4-98.6) → RR= 0.044	2:1	4,483,767 people	0.0723	1
Pfizer/Biontech >=55 years	93.7% (80.6-98.8) → RR= 0.063	2:1	3,237,076 people	0.0562	1
Moderna 18-65 years	95.6% (90.6-97.9) → RR=0.044	1:1	5,712,251 people	0.0697	1
Moderna >=65 years	86.4% (61.4-95.2) → RR=0.136	1:1	2,008,592 people	0.0531	1
Moderna 18-65 years	95.6% (90.6-97.9) → RR=0.044	2:1	5,712,251 people	0.0697	1
Moderna >=65 years	86.4% (61.4-95.2) → RR=0.136	2:1	2,008,592 people	0.0531	1
Janssen* 18-59 years	63.7% (53.9-71.6)→ RR=0.363	1:1	5,142,635 people	0.0710	1
Janssen >=60 years	76.3% (61.6-86.0)→ RR=0.237	1:1	2,578,208 people	0.0543	1
Janssen 18-59 years	63.7% (53.9-71.6)→ RR=0.363	2:1	5,142,635 people	0.0710	1
Janssen >=60 years	76.3% (61.6-86.0)→ RR=0.237	2:1	2,578,208 people	661,162/12,186,162=0.0543	1

The probabilities were calculated based on the following effectiveness reported in the clinical trials by age groups:

Pfizer/Biontech (BNT162b2) [75]:

1. 16-55 years: 5 cases (0,05%) de 9897 vaccinated and 114 cases (1,145%) among 9955 in comparison group. Efficacy: 95,6% (89,4–98,6)
2. >55 years: 3 cases (0,04%) de 7500 vaccinated and 48 cases (0,636%) among 7543 in comparison group. Efficacy: 93,7% (80,6– 8,8)
3. ≥65 years: 1 case (0,025%) de 3848 vaccinated and 19 cases (0,489%) among 3880 in comparison group. Efficacy: 94,7% (66,7– 99,9)

4. ≥75 years: 0 cases among 774 vaccinated and 5 cases (0,636%) among 785 in comparison group. Efficacy: 100% (-13,1 – 100)

Moderna (mRNA-1273) [76]:

1. ≥18-65 years: 7 cases (0,066%) de 10551 vaccinated and 156 cases (1,482%) among 10521 in comparison group. Efficacy: 95,6% (90,6-97,9).
2. ≥65 years: 4 cases (0,111%) de 3583 vaccinated and 29 cases (0,816%) among 3552 in comparison group. Efficacy: 86,4% (61,4-95,2)

Janssen (efficacy 14 days to COVID-19 mild-severe, 1 dose) [78]:

1. ≥18-59 years: 95 cases among 12750 vaccinated and 260 cases among 12782 in comparison group. Efficacy: 63,7% (53,9-71,6).
2. ≥60 years: 21 cases among 6764 vaccinated and 88 cases among 6762 in comparison group. Efficacy: 76,3% (61,6-86,0).

*Note: for Oxford/Astrazeneca vaccine no efficacy was published by age groups. The published efficacy referred mainly to individuals aged 18-55 years although also 12% of participants were ≥56y [77].

Regarding children and adolescents, 1.6 million patients aged 0-17 years are registered in the data source distributed by age and sex as follows:

	TOTAL	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
TOTAL	**1598371**	**131318**	**72389**	**77656**	**80082**	**81668**	**81753**	**86036**	**89054**	**90738**	**92347**	**95206**	**91962**	**90963**	**89196**	**88696**	**88560**	**85602**	**85145**
Women	**777196**	**63920**	**35333**	**37770**	**38976**	**39282**	**39612**	**41726**	**43396**	**44264**	**44783**	**46103**	**44623**	**44141**	**43534**	**42939**	**43432**	**41749**	**41613**
Men	**821175**	**67398**	**37056**	**39886**	**41106**	**42386**	**42141**	**44310**	**45658**	**46474**	**47564**	**49103**	**47339**	**46822**	**45662**	**45757**	**45128**	**43853**	**43532**

Since no information about the efficacy has been provided yet for younger than 16 or 18 years (according to the vaccines), probability of the upper limit of the 95% confidence interval to be < 1.00 will be calculated ah-hoc in case of approval vaccinations among children and adolescents.

9.6 Data Management

The necessary data to implement this study from each included data source will be extracted by BIFAP's IT through programming and/or epidemiologists through BIFAP specific applications.

Previous to that the different registries will be linked to primary care histories by the local government teams of regions participating in BIFAP. Those pseudonymised individual-level will be sent to BIFAP IT and held at the AEMPS research sites. The Governance of data use is publicly available [68].

STATA will be used as statistic analytic program.

9.7 Data Analysis

9.7.1 Descriptive Characteristics

The distributions of baseline characteristics by exposure group/periods will be calculated to describe the study cohorts and illustrate differences between the groups compared. For continuous variables, means, standard deviations, medians, and interquartile ranges will be estimated. For categorical variables, counts and proportions will be estimated. The missingness of variables such as BMI, smoking status and alcohol use among other will also be described.

9.7.2 Incidence Analysis

The crude incidence rates of the COVID-19 outcomes will be estimated after time zero for the vaccinated and unvaccinated groups for each brand.

Additionally, each vaccination strategy (e.g., fully completed on time, partially completed, delayed completion; see Section 9.3.1 and **¡Error! No se encuentra el origen de la referencia.**), for the same brand will be considered a separate exposure group and compared to unvaccinated.

Crude incidence rates (with 95% CI) will be estimated using the number of outcome events as the numerator and the number of person-years of follow-up as the denominator for the study periods defined.

Additionally, cumulative incidence curves (1 minus the Kaplan–Meier risk) for outcomes will be estimated to visualise the occurrence of COVID-19 outcomes over time since time zero [38].

9.7.3 Comparative Analysis

9.7.3.1 Estimation of Vaccine Effectiveness

Cox proportional hazards regression, yielding a hazard ratio (HR; 95%CI), accounting for censoring, changing incidence over time, and competing risks (Section 9.7.3.3) will be estimated for the defined periods: i.e. 0 to 14 days after D1, 15 after D1 until D2, 0 to 14 days after D2 and 15-90 days, 91-180 days, and ≥181 days after D2 in comparison with similar followed periods from time zero in the unvaccinated group.

This method will calculate a single vaccine effectiveness measure for each period of observation (assumed constant over those defined periods) in vaccinated versus unvaccinated groups. Crude vaccine effectiveness will be calculated for all the outcomes as 1 minus HR.

The assumption of proportionality of the survival curves in the vaccinated and unvaccinated groups will be evaluated.

9.7.3.2 Adjustment for Confounding

Backward stepwise selection method will be used to identify variables associated with each outcomes according to a level of significance (p-exit≥0.1, p-entry<0.05) for inclusion or exclusion, that could confound the effectiveness estimates and thus adjust the final models.

Confounders will be measured at baseline and updated before each vaccine dose is administrated or every 28 days (in order to update the collection of confounders during the follow-up of both, vaccinated and unvaccinated patients). The adjusted vaccine effectiveness for all the outcomes will be estimated as 1 minus the adjusted HR.

Adjusted time-specific RDs (e.g., RD at 3, 6, 9, 12 months after time zero) will also be tested [38]

Patients receiving one type of vaccine who receive later a dose of another COVID-19 vaccine, will be considered switcher from the date that the last dose was received and analysed as switchers, since such censoring/switch can be informative. Among switchers, each combination of vaccines will be analysed separately.

9.7.3.3 Competing Risk Analysis

Events (e.g., death due to causes other than COVID-19) that may make it impossible for the event of interest to occur (e.g., hospitalisations for COVID-19) act as competing events in the analysis of certain outcomes. The relevance of competing events increases when the study population is old or sick (a plausible target population of the COVID-19 vaccine) or when the follow-up is long. The use of specific methodology to handle competing events can help in the accurate interpretation of the study results.

Thus, competing risk in survival analysis will be taken into account in an additional HR estimation.

9.7.4 Subgroup, Sensitivity, and Exploratory Analyses

9.7.4.1 Conducting Subgroup Analyses

The primary analyses will be stratified by clinically meaningful subgroups found or suspected linked with the prognostic of the covid-19 infection, which will include the following:

- Old patients (i.e. aged ≥80 years)
- Patients with cardiovascular disease,
- Patients with diabetes,
- Patients with chronic pulmonary disease (COPD, asthma, etc);
- Patients with cancer
- Patients who are immunocompromised
- Pregnant women
- Smoker patients (suspected linked with the prognosis in some studies)
- Patients with obesity (BMI for obese category)
- Patients vaccinated against Influenza in the fall-winter season 2020-2021 this analysis seeks to reduce unmeasured confounding by health care access, adherence to recommendations, and preventive health care behaviours [9,25].
- Geographic region (if sufficient sample size in all regions)

Also, the primary analyses will be stratified by calendar months in order to observe in periods with stronger confusion due to individuals prioritisation (i.e. channelling of atypical patients in the period immediately following approval [44]). For instance, confusion could be strongest when the vaccinations occurred only for health care workers and nurse home residents. Background anamnesis may not be systematically recorded in primary care for those groups and we do not have information about the job or nursing home residing for selecting non-vaccinated individuals based on those criteria (or matching by it), thus the compared unvaccinated cohort may have less probability of vaccination or outcome, respectively.

Stratifying by months, we could observe differences in effectiveness that could be affected by a higher confusion in months previous to the date when vaccination was recommended for a broad general population.

9.7.4.2 Varying Outcome Definitions

The validity and precision of COVID-19 diagnosis coding is currently unknown, thus those recorded codes will be extracted for comparison to lab test as gold-standard of the infection.

If the number of patients with recorded codes for COVID-19 infection (so-called 'possible cases') are far from those identifies through lab test (so-called 'confirmed cases'), a sensitivity analysis for vaccine effectiveness estimation will be performed including all of them.

COVID-19 Diagnosis Coding

COVID-19 diagnosis coding has been introduced recently into the code systems included in BIFAP and linked registries (International Classification of Primary Care (ICPC)(19), International Classification of Diseases (20), 9th and 10th Revision (ICD-9 and ICD-9)) as follows:

ICPC	Infección debida a coronavirus NE	A77.01
ICD9MC	Infección por coronavirus asociada a SARS	079.82
ICD10 ES	COVID-19 (*)	U07.1

(*) Nota: hasta jun-2020 la equivalencia estuvo establecida a "Infección debida a coronavirus, no especificada" (código B34.2)

Those codes will be distinguished from subcodes for confirmed or suspected cases [70-71]:

ICPC	Contactos/exposición a otras enfermedades víricas transmisibles	A23.19
CIE9MC	Contacto/exposición a otras enfermedades víricas	V01.79
CIE10 ES	Contacto y (sospecha de) exposición a otras enfermedades víricas transmisibles	Z20.828

Those codes were introduced in 2020 (19,20) and ready for codification from the start of the vaccination period.

Also, in order to estimate the validity and precision of COVID recorded diagnosis, the following parameters will be estimated to further used in research studies in BIFAP in comparison with lab results as gold-standard:

- Positive Predictive Value (i.e. PPV=confirmed covid/diagnosis code)
- Sensitivity (S=diagnosis code/confirmed covid).

9.7.4.3 Correcting for Differential Outcome Misclassification

Since no validation studies are planned for the outcome (identified through lab results considered gold-standard in the clinical practice), quantitative bias analyses will not be performed to evaluate the potential impact of differential misclassification of the outcome on the observed study estimate. As a proxy, external comparison of incidences (overall, by region, calendar dates and age) will be performed as aforementioned. This may be mentioned as limitation although lab result are even clinically used as confirmatory infection.

9.7.5 Missing Data

Missing data for the information about COVID-19 vaccinations, i.e. date of administration, type/brand of vaccine, dose order (whether 1st or 2nd dose), may happen due the workload. However, due the importance of knowing the effect of the vaccinations in the pandemic we expect the health professional properly record all required data.

The distribution and pattern of missingness (i.e. missing completely at random or missing at random) will be assessed [49]. The pattern missing not at random (MNAR) may be derived from external data on the doses distributed by region and among prioritised groups. If amount of missing is relevant, single stochastic imputation will be performed based on determinant of both, missing data and the value of the variable with missing data. Mainly, calendar month or week, region and prioritised group will be tested as determinant of missing and will established for the imputation. Multiple imputation correcting by the variance is a more complex approach (even more using a time-varying data source) and may not be worthy if simple determinants are identified.

No imputation are planned for missing data in potential confounders (i.e. BMI or smoking use). In order to all individual contribute to the analysis, a missing category will be established for patients with missing data. If those variables suggest poor confusion in the regression models, they will be removed as confounders for the final models [72].

9.8 Quality Control

Quality control, rules for secure and confidential data storage, methods to maintain and archive project documents, quality-control procedures for programming, standards for writing analysis plans, will be performed according to the BIFAP governance that can be consulted in the following linked of BIFAP website [68].

9.9 Limitations of the Research Methods

Although this protocol addresses many design considerations to avoid common biases of vaccine effectiveness research, studies of COVID-19 vaccines may be subject to many of the limitations common to non-randomised studies based in existing health care data.

Patient's test-seeking behaviour may be associated with both severity of infection symptoms and personal health-seeking behaviour, which may introduce selection bias or confounding [28,29].

Confounding of the relationship between vaccine receipt and COVID-19 outcomes may be very likely, and spurious relationships between vaccine receipt and implausible outcomes have been noted in other studies comparing vaccine recipients to non-recipients. The use of eligibility criteria to define comparable exposure groups, covariates

based on updated subject-matter knowledge will allow investigators to adequately address confounding through the design and analysis.

The baseline clinical history of individuals in the initial priority groups (i.e. health care workers and people living in nursing home) may not be systematically included in BIFAP eHR. Thus, many of them may not be up-to-standard for research as explained above and power limitation may be expected.

Even though efforts will be done for proper control of bias and multivariate adjustment minimizing confusion, no information about the job and type of residence will be available. Consequently, confusion may still be present due to the higher probability of infection among them versus other social groups. That aspect would direct towards a reduction in the effectiveness estimations.

If data about determinants, such as personal health-seeking behaviour, would not be recorded in the database, the selection of patients included in the compared groups and/or their dates to start the comparison could be biased, i.e.

- Patients prone to infections or to develop a severe/hospitalised infection or those with active respiratory infection/disease (prevalent active pneumonia, COPD, etc) decided to attend to receive the vaccination earlier/more than those less prone to it, we could observe patients more affected by covid among the vaccinated group. That aspect would direct towards a biased lower effectiveness estimation due to selection bias).

- The opposite could also be true, i.e. patients more adhered to general prevention measured (personal health-seeking behaviour such as applying social distance, wearing mask, volunteer confinement, etc) were more vaccinated we could be observing more covid infections in the unvaccinated group. That aspect would direct towards a biased higher effectiveness estimation due to selection bias.

Selective recruitment into the study of vaccinated/unvaccinated subjects from those recorded in the database with quality criteria (up-to-standard information) that are not representative of the general vaccinated/unvaccinated subjects respectively in the source population could produce selection bias. For instance:

- If we were losing people having died from coronavirus disease even though they were vaccinated (i.e. non effective vaccinations), selection bias would be present. Similarly, if more vaccinated participants were survivors of previous covid infections than vaccinated non-participants (example, they died by covid) because those attend the PCP (having information and minimum anamnesis in the database required to participate) while vaccinated non-participants do not.

- Also, If vaccinated participants were recorded in the database because they are more closely surveyed by the PCP/nurse due to their predisposition to complicated COVID infection (i.e. baseline health conditions), while vaccinated

non-participants were not recorded in the database because they do not seek healthcare, we would be including in the study patients with more probability to severe infection than the real overall vaccinated individuals

The selection bias could direct towards any direction the effectiveness estimates.

if an individual's personal health beliefs and behaviours increase the likelihood of both COVID-19 vaccines receipt and seeking health care for milder disease, then the analysis of this outcome may be particularly subject to confounding by health care–seeking behaviour.

The validity of coding for future COVID-19 vaccines—including the ability to distinguish between separate products—or COVID-19 outcomes is unknown at this time, and the likelihood of missing data on specific vaccine brand may be high. Misclassification of vaccine exposure, outcome status, or covariates is possible in existing health care data not collected for research purposes. Information about COVID-19 infection (through test results or codes) may not be captured reliably in databases. Additionally, COVID-19 testing is not systematic, and laboratory confirmation of case status may not be available for many patients. Currently, bias may be minimised since COVID 19 testing becomes near universal and repeated with individuals, with similar testing capacity and practice across regions, though. Since we will not have any better gold-standard for case status than the information about confirmed covid-19 infection through lab results (used for case definition in the current study) its precision will not be evaluated but with the expected external incidences published by the Health authorities.

Also, we consider that if COVID misclassification occurred it would be similar in both, vaccinated and unvaccinated groups, however surveillance or detection bias (different likelihood of screening or testing for COVID between the groups) cannot be discarded. If vaccinated individuals had less likelihood of screening or test than unvaccinated, we would artificially observe more cases among unvaccinated, directing toward a biased higher vaccine effectiveness estimation.

Data from multiple regions and sources will be merged in BIFAP and included in the study, thus, there may be variation in the capture and recording of various clinical elements. Additionally, different types of data sources will be used (e.g., records from general practice and lab results from other registers) as well as different coding systems, then variables defined in different data sources may not exactly represent the same concept across data sources. Thus, heterogeneity of vaccine effectiveness estimates across data sources may be due to the underlying heterogeneity of confounding control, misclassification, or other data source factors rather than true differences in vaccine effectiveness.

Patterns of routine health care delivery and utilisation may be disrupted during the COVID-19 pandemic as patients and providers forgo or delay routine preventive, elective, or non-emergency care. These disruptions in health care may result in under-

ascertainment of important patient comorbidities in existing health care databases during periods of disruption.

In order to conduct a valid study and address potential confounding, the eligibility criteria of the study will be restricted relative to the approved indications of the vaccines. Therefore, the results of this study may not be generalisable to the general population but to the restricted population.

With the introduction of a new vaccine, there is the potential for rapidly changing herd immunity in the population. This study is not designed to assess overall and indirect effects of vaccination with COVID-19 vaccines.

Comparative analyses may not be possible in every setting if confounding is deemed to be insurmountable. However, descriptive information about vaccine recipients and crude incidence rates may still be informative and meaningful, even without calculation of vaccine effectiveness measures.

The capture of over-the-counter medications, potentially indicative of short-term disease status (e.g., painkillers, cough medicines, and fever reducers) may not be captured reliably.

9.10 Other Aspects

This study protocol will be sent to the BIFAP scientific advisory board and Ethical Committee for Research with Medical products (CEIm regional de la Comunidad de Madrid)

10 Protection of Human Subjects

This is a non-interventional study using secondary data collection and does not pose any risks for patients.

The investigators will have access to secondary use of only fully anonymised data.

We will apply for an independent ethics committee review according to local regulations (i.e. Ethical Committee for Research with Medical products (CEIm regional de la Comunidad de Madrid)).

Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants as reported in the BIFAP governance document [69].

11 Management and Reporting of Adverse Events/Adverse Reactions

For studies in which the research team uses only data from automated health care databases, according to the International Society for Pharmacoepidemiology (ISPE) [51] *Guidelines for Good Pharmacoepidemiology Practices (GPP*),

> *"Aggregate analysis of database studies can identify an unexpected increase in risk associated with a particular exposure. Such studies may be reportable as study reports, but typically do not require reporting of individual cases. Moreover, access to automated databases does not confer a special obligation to assess and/or report any individual events contained in the databases. Formal studies conducted using these databases should adhere to these guidelines."*

For non-interventional study designs that are based on secondary use of data, such as studies based on medical chart reviews or electronic health care records, systematic reviews or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of adverse events/adverse drug reactions should only be summarised in the study report, where applicable [52].

According to the European Medicines Agency (EMA) *Guideline on Good Pharmacovigilance Practices (GVP), Module VI – Management and Reporting of Adverse Reactions to Medicinal Products* [52],

> *"All adverse events/reactions collected as part of [non-interventional post-authorisation studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarised in the interim safety analysis and in the final study report."*

Module VIII – Post-Authorisation Safety Studies, echoes this approach [1]. European Union legislation further states that for certain study designs such as retrospective cohort studies, particularly those involving electronic health care records, it may not be feasible to make a causality assessment at the individual case level.

12 Plans for Disseminating and Communicating Study Results

In its *Guidelines for Good Pharmacoepidemiology Practices (GPP)*, ISPE contends that "there is an ethical obligation to disseminate findings of potential scientific or public health importance" [51].

Study results will be published following guidelines, including those for authorship, established by the International Committee of Medical Journal Editors [53]. When

reporting results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist will be followed [54].

Communication via appropriate scientific venues will be considered.

The study progress reports as included in the section 6-Milestones and Timeline will be circulated among the participants of the collaborating public institutions for communication and review.

13 Other Good Research Practice

This study will adhere to the *Guidelines for Good Pharmacoepidemiology Practices (GPP)* [51] and has been designed in line with the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) *Guide on Methodological Standards in Pharmacoepidemiology* [55]. The *ENCePP Checklist for Study Protocols* [56] will be completed (see Annex 1).

The study is a postauthorisation study of vaccine effectiveness and will comply with the definition of the non-interventional (observational) study referred to in the International Conference on Harmonisation tripartite guideline *Pharmacovigilance Planning E2E* [57] and provided in the EMA *Guideline on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies* [1], and with the 2012 European Union pharmacovigilance legislation, adopted 19 June 2012 [58]. The study will comply with the study reporting requirements specified in Module VIII section VIII.B.6.3.2. "Final study report" of the *Guideline of Good Pharmacovigilance Practices* [1].

The study will be registered in the EU PAS Register[1] [59] before study implementation commences.

The research team will adhere to the general principles of transparency and independence in the ENCePP Code of Conduct [60] and the ADVANCE Code of Conduct [50]. There is no sponsor in the current study.

The research team will apply for the ENCePP Study Seal [61].

14 References

1. EMA: European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies (EMA/813938/2011 Rev 3). 13 October 2017. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf. Accessed 17 January 2020.

[1] EU PAS Register = European Union electronic register of post-authorisation studies.

2. Lopez Bernal JA, Andrews N, Amirthalingam G. The use of quasi-experimental designs for vaccine evaluation. Clin Infect Dis. 2019 May 2;68(10):1769-76.

3. Jackson ML, Yu O, Nelson JC, Naleway A, Belongia EA, Baxter R, et al. Further evidence for bias in observational studies of influenza vaccine effectiveness: the 2009 influenza A(H1N1) pandemic. Am J Epidemiol. 2013 Oct 15;178(8):1327-36.

4. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006 Apr;35(2):337-44.

5. Jackson LA, Nelson JC, Benson P, Neuzil KM, Reid RJ, Psaty BM, et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. Int J Epidemiol. 2006 Apr;35(2):345-52.

6. Zhang HT, McGrath LJ, Wyss R, Ellis AR, Stürmer T. Controlling confounding by frailty when estimating influenza vaccine effectiveness using predictors of dependency in activities of daily living. Pharmacoepidemiol Drug Saf. 2017 Dec;26(12):1500-6.

7. Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA. Impact of influenza vaccination on seasonal mortality in the US elderly population. Arch Intern Med. 2005 Feb 14;165(3):265-72.

8. Simonsen L, Taylor RJ, Viboud C, Miller MA, Jackson LA. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. Lancet Infect Dis. 2007 Oct;7(10):658-66.

9. McGrath LJ, Ellis AR, Brookhart MA. Controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. Am J Epidemiol. 2015 Jul 1;182(1):17-25.

10. Murray EJ, Caniglia EC, Swanson SA, Hernández-Díaz S, Hernán MA. Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. J Clin Epidemiol. 2018 Nov;103:10-21.

11. Lund JL, Horvath-Puho E, Komjathine Szepligeti S, Sorensen HT, Pedersen L, Ehrenstein V, et al. Conditioning on future exposure to define study cohorts can induce bias: the case of low-dose acetylsalicylic acid and risk of major bleeding. Clin Epidemiol. 2017;9:611-26.

12. WHO: World Health Organization. COVID-19 coding in ICD-10. 25 March 2020. Available at: https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf?ua=1. Accessed 19 June 2020.

13. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov;79:70-5.

14. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016 Apr 15;183(8):758-64.

15. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. Eur J Epidemiol. 2017 Jun;32(6):495-500.

16. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615-25.

17. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007 Mar;16(3):241-9.

18. García-Albéniz X, Hernán MA, Logan RW, Price M, Armstrong K, Hsu J. Continuation of annual screening mammography and breast cancer mortality in women older than 70 years. Ann Intern Med. 2020 Mar 17;172(6):381-9.

19. Brookhart MA. Counterpoint: the treatment decision design. Am J Epidemiol. 2015 Nov 15;182(10):840-5.

20. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat Methods Med Res. 2013 Feb;22(1):70-96.

21. García-Albéniz X, Hsu J, Bretthauer M, Hernán MA. Effectiveness of screening colonoscopy to prevent colorectal cancer among Medicare beneficiaries aged 70 to 79 years: a prospective observational study. Ann Intern Med. 2017 Jan 3;166(1):18-26.

22. Hernán MA, Robins JM. Target trial emulation (chapter 22). In: Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020.

23. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010 May;21(3):383-8.

24. Butler AM, Layton JB, Krueger WS, Kshirsagar AV, McGrath LJ. Assessing residual bias in estimating influenza vaccine effectiveness: comparison of high-dose versus standard-dose vaccines. Med Care. 2019 Jan;57(1):73-8.

25. Zhang HT, McGrath LJ, Ellis AR, Wyss R, Lund JL, Stürmer T. Restriction of pharmacoepidemiologic cohorts to initiators of medications in unrelated preventive drug classes to reduce confounding by frailty in older adults. Am J Epidemiol. 2019 Jul 1;188(7):1371-82.

26. McGrath LJ, Cole SR, Kshirsagar AV, Weber DJ, Stürmer T, Brookhart MA. Hospitalization and skilled nursing care are predictors of influenza vaccination among patients on hemodialysis: evidence of confounding by frailty. Med Care. 2013 Dec;51(12):1106-13.

27. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data. New York: Springer; 2009.

28. Ainslie KEC, Shi M, Haber M, Orenstein WA. On the bias of estimates of influenza vaccine effectiveness from test-negative studies. Vaccine. 2017 Dec 19;35(52):7297-301.

29. Jackson ML, Phillips CH, Benoit J, Kiniry E, Madziwa L, Nelson JC, et al. The impact of selection bias on vaccine effectiveness estimates from test-negative studies. Vaccine. 2018 Jan 29;36(5):751-7.

30. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis. 2020;94:91-5.

31. Faurot KR, Jonsson Funk M, Pate V, Brookhart MA, Patrick A, Hanson LC, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. Pharmacoepidemiol Drug Saf. 2015;24(1):59-66.

32. Cuthbertson CC, Kucharska-Newton A, Faurot KR, Stürmer T, Jonsson Funk M, Palta P, et al. Controlling for frailty in pharmacoepidemiologic studies of older adults: validation of an existing Medicare claims-based algorithm. Epidemiology (Cambridge, Mass). 2018;29(4):556-61.

33. Rothman KJ, Greenland S. Planning study size based on precision rather than power. Epidemiology. 2018 Sep;29(5):599-603.

34. VAC4EU. Operations of VAC4EU study network. 2019. Available at: https://vac4eu.org/operations-of-vac4eu-study-network/. Accessed 22 October 2020.

35. Gini R, Sturkenboom MCJ, Sultana J, Cave A, Landi A, Pacurariu A, et al. Different strategies to execute multi-database studies for medicines surveillance in real-world setting: a reflection on the European model. Clin Pharmacol Ther. 2020 Apr 3.

36. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009;28(25):3083-107.

37. Yang D, Dalton JE: SAS Institute Inc. A unified approach to measuring the effect size between two groups using SAS. Cary, NC; 2012. Available at: https://support.sas.com/resources/papers/proceedings12/335-2012.pdf. Accessed 22 October 2020.

38. Hernan MA, Robins JM. Causal survival analysis. In: Causal inference. Boca Raton: Chapman & Hall/CRC; 2020. Available at: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/. Accessed 22 October 2020.

39. Hanquet G, Valenciano M, Simondon F, Moren A. Vaccine effects and impact of vaccination programmes in post-licensure studies. Vaccine. 2013;31(48):5634-42.

40. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol. 2006 Feb 1;163(3):262-70.

41. Hernán MA, Robins JM. G-methods for time-varying treatments (chapter 21). In: Causal inference: what if. Boca Raton: Chapman & Hall/CRC; 2020.

42. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med. 2020 Apr 15;39(8):1199-236.

43. Geskus RB. Data analysis with competing risks and intermediate states. CRC Press Taylor & Francis Group; 2016.

44. Gagne JJ, Bykov K, Willke RJ, Kahler KH, Subedi P, Schneeweiss S. Treatment dynamics of newly marketed drugs and implications for comparative effectiveness research. Value Health. 2013 Sep-Oct;16(6):1054-62.

45. Jonsson Funk M, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. Curr Epidemiol Rep. 2014;1(4):175-85.

46. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. Am J Epidemiol. 1993 Dec 1;138(11):1007-15.

47. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M, Brown JS. Confounding adjustment in comparative effectiveness research conducted within distributed research networks. Med Care. 2013 Aug;51(8 Suppl 3):S4-10.

48. Weibel D, Sturkenboom M, Black S, de Ridder M, Dodd C, Bonhoeffer J, et al. Narcolepsy and adjuvanted pandemic influenza A (H1N1) 2009 vaccines—multi-country assessment. Vaccine. 2018 Oct 1;36(41):6202-11.

49. Greenland S, Rothman KJ. Fundamentals of epidemiologic data analysis (chapter 13). In: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. Philadelphia: Lippincott Williams & Williams; 2008.

50. Kurz X, Bauchau V, Mahy P, Glismann S, van der Aa LM, Simondon F, et al. The ADVANCE Code of Conduct for collaborative vaccine studies. Vaccine. 2017 Apr 4;35(15):1844-55.

51. ISPE: International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). Revision 3. June 2015. Available at: https://www.pharmacoepi.org/resources/policies/guidelines-08027/. Accessed 17 January 2020.

52. EMA: European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). 22 November 2017. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf. Accessed 16 June 2020.

53. ICMJE: International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. December 2019. Available at: http://www.icmje.org/recommendations/. Accessed 22 October 2020.

54. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344-9.

55. ENCePP: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Guide on methodological standards in pharmacoepidemiology (EMA/95098/2010 Rev. 7). July 2018. Available at: http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml. Accessed 22 October 2020.

56. ENCePP: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. ENCePP checklist for study protocols (revision 4). 15 October 2018. Available at: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml. Accessed 22 October 2020.

57. ICH: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Pharmacovigilance planning. E2E. 2004. Available at: https://database.ich.org/sites/default/files/E2E_Guideline.pdf. Accessed 22 October 2020.

58. European Commission. Commission implementing Regulation (EU) No 520/2012 of 19 June 2012 on the performance of pharmacovigilance activities provided for in Regulation (EC) No 726/2004 of the European Parliament and of the Council and Directive 2001/83/EC of the European Parliament and of the Council. 20 June 2012. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:159:0005:0025:EN:PDF. Accessed 22 October 2020.

59. ENCePP: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. The European Union electronic register of post-authorisation studies (EU PAS Register). 20 December 2018. Available at: http://www.encepp.eu/encepp_studies/indexRegister.shtml. Accessed 22 October 2020.

60. ENCePP. The ENCePP code of conduct for scientific independence and transparency in the conduct of pharmacoepidemiological and pharmacovigilance studies (Revision 4). 15 March 2018. Available at: http://www.encepp.eu/code_of_conduct/. Accessed 17 January 2020.

61. ENCePP: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. The ENCePP seal. 08 September 2020. Available at: http://www.encepp.eu/encepp_studies/index.shtml. Accessed 22 October 2020.

62. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):237-42.

63. García-Albéniz X, Hsu J, Hernán MA. A challenge of real world data: how to assign individuals to a treatment strategy when their data are consistent with several treatment strategies at baseline. Pharmacoepidemiol Drug Saf.28(S2):234.

64. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. BMJ. 2018 Feb 1;360:k182.

65. Maciá-Martínez M-A, Gil M, Huerta C, Martín-Merino E, Álvarez A, Bryant V, et al. Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria (BIFAP): A data resource for pharmacoepidemiology in Spain. Pharmacoepidemiol Drug Saf. 2020;29(10):1236–45.

66. Martin-Merino E, Llorente-Garcia A, Montero-Corominas D, Huerta C. The recording of human papillomavirus (HPV) vaccination in BIFAP primary care database: A validation study. Pharmacoepidemiol Drug Saf. 2019 Feb;28(2):201–8.

67. Agencia Española del Medicamento y Productos Sanitarios. BIFAP: Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria [Internet]. [cited 2017 Jan 31]. Available from: http://bifap.aemps.es/

68. Layton B, García X. Real-world effectiveness of <<COVID-19 vaccine product>> in Europe: a protocol template for a cohort study based in existing health care data sources from the ACCESS project [Internet]. 2020 Oct. Available from: https://vac4eu.org/wp-content/uploads/2021/02/3f.-Protocol_ACCESS_COVID-19-EHR-Vaccine-Effectiveness-Protocol-Template.pdf

69. Agencia Española del Medicamento y Productos Sanitarios. Gobernanza del acceso a los datos de BIFAP (Base de datos para la Investigación Farmacoepidemiológica en Atención Primaria). [Internet]. 2021 [cited 2021 Mar 30]. Available from: http://bifap.aemps.es/docs/Gobernanza_acceso_datos_BIFAP_v2_2021.pdf

70. ICPC-2. International classification of primary care. Second Edition. Oxford: Oxford University Press; 1998.

71. World Health Organization. WHO IRIS: International classification of diseases : [9th] ninth revision, basic tabulation list with alphabetic index [Internet]. 1978 [cited 2017 Sep 26]. Available from: http://www.who.int/iris/handle/10665/39473

72. Martín-Merino E, Calderón-Larrañaga A, Hawley S, Poblador-Plou B, Llorente-García A, Petersen I, et al. The impact of different strategies to handle missing data on both precision and bias in a drug safety study: a multidatabase multinational population-based cohort study. Clin Epidemiol. 2018;10:643–54.

73. Instituto Nacional de Estadística. (National Statistics Institute) [Internet]. [cited 2021 16 April]. Available from: https://www.ine.es/en/

74. Ministerio de Sanidad, Consumo y Bienestar Social. Enfermedad por coronavirus, COVID-19. Informacion científica-técnica. [Internet]. [cited 2021 Mar 29]. Available from: https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/ITCoronavirus.pdf

75. BioNTech SE. A PHASE 1/2/3, PLACEBO-CONTROLLED, RANDOMIZED, OBSERVER-BLIND, DOSE-FINDING STUDY TO EVALUATE THE SAFETY, TOLERABILITY, IMMUNOGENICITY, AND EFFICACY OF SARS-COV-2 RNA VACCINE CANDIDATES AGAINST COVID-19 IN HEALTHY INDIVIDUALS [Internet]. clinicaltrials.gov; 2021 Jan [cited 2021 Jan 11]. Report No.: NCT04368728. Available from: https://clinicaltrials.gov/ct2/show/NCT04368728

76. ModernaTX, Inc. A Phase 3, Randomized, Stratified, Observer-Blind, Placebo-Controlled Study to Evaluate the Efficacy, Safety, and Immunogenicity of mRNA-1273 SARS-CoV-2 Vaccine in Adults Aged 18 Years and Older [Internet]. clinicaltrials.gov; 2020 Oct [cited 2021 Jan 11]. Report No.: NCT04470427. Available from: https://clinicaltrials.gov/ct2/show/NCT04470427

77. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. The Lancet. 2021 Jan 9;397(10269):99–111.

78. Zhang R, Hefter Y. FDA Review of Efficacy and Safety of the Janssen COVID-19 Vaccine Emergency Use Authorization Request. :75.

Annex 1.
ENCePP Checklist for Study Protocols

Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 4)

Adopted by the ENCePP Steering Group on 15/10/2018

For each question of the Checklist, the investigator should indicate whether or not it has been addressed in the study protocol. If the answer is "Yes", the section number of the protocol where this issue has been discussed should be specified. It is possible that some questions do not apply to a particular study (for example, in the case of an innovative study design). In this case, the answer 'N/A' (Not Applicable) can be checked and the "Comments" field included for each section should be used to explain why. The "Comments" field can also be used to elaborate on a "No" answer.

Study title: Real-world effectiveness of different COVID-19 vaccines in Spain: a cohort study based on public electronic health records (BIFAP)

EU PAS Register® number: Not-applicable
Study reference number (if applicable):

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	6
1.1.2 End of data collection[2]	☒	☐	☐	6
1.1.3 Progress report(s)	☒	☐	☐	6
1.1.4 Interim report(s)	☐	☒	☐	6
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	6
1.1.6 Final report of study results.	☒	☐	☐	6

Comments:

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.

[2] Date from which the analytical dataset is completely available.

Section 2: Research question		**Yes**	**No**	**N/A**	**Section Number**
2.1	Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	7,8
	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)	☒	☐	☐	7,8
	2.1.2 The objective(s) of the study?	☒	☐	☐	7,8
	2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	7,8
	2.1.4 Which hypothesis(-es) is (are) to be tested?	☒	☐	☐	7,8
	2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

2.1.5 NA: There is a priori hypothesis of the efficacy reported in trials and expected replicated in real world.

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)	☒	☐	☐	9.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	7, 9.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	9.7.2
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))	☒	☐	☐	9.7.3
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)	☐	☐	☒	

Comments:

To 3.5: NA since this a an observational effectiveness study using secondary data

Section 4: Source and study populations		**Yes**	**No**	**N/A**	**Section Number**
4.1	Is the source population described?	☒	☐	☐	9.2; 9.3.1
4.2	Is the planned study population defined in terms of:				
	4.2.1 Study time period	☒	☐	☐	9.2; 9.3.1
	4.2.2 Age and sex	☒	☐	☐	9.2; 9.3.1
	4.2.3 Country of origin	☒	☐	☐	9.2; 9.3.1
	4.2.4 Disease/indication	☒	☐	☐	9.3.1*
	4.2.5 Duration of follow-up	☒	☐	☐	9.2;9.3.1$^{\Psi}$

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	9.3.1

Comments:

The last date available in the data source at data collection will be the end of the Study time period.

*The study population is defined in terms of Exposure (instead of Disease/Indication)

ΨDuration of follow-up: Inclusion date and stop date to follow-up define the Duration of follow-up.

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)	☒	☐	☐	9.3.1
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☒	☐	☐	9.1.1.3
5.3 Is exposure categorised according to time windows?	☒	☐	☐	9.1.1.4
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	9.1.1.4
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☒	☐	☐	9.1.1.4
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	9.3.1;9.1.1.4*

Comments:

*We have discarded to select other control groups due to the reason explained in 9.1.

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?	☒	☐	☐	8; 9.3.3
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	8; 9.3.3
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)	☒	☐	☐	8; 9.1.1.2
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)	☐	☒	☐	

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)	☒	☐	☐	9.3.4
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	9.1*;9.1.1.4 Ψ
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	9.1.1.2; 9.1.1.4; 9.7.3.1; 9.9

Comments:

*The choice of an exchangeable comparator is described as well as the reasons to discard other potential reference groups, informing about the efforts to avoid Selection Bias. Ψ Also, in order to avoids selection bias, the date to start the contributed and compared time in each compared group is provided and justified. That date (Time zero) allows to align the evaluation of eligibility criteria, covariate assessment and exposure assignment.

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)	☒	☐	☐	9.7.4.1

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)	☒	☐	☐	9.3.1; 9.4
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	9.3.3; 9.4
9.1.3 Covariates and other characteristics?	☒	☐	☐	9.3.4; 9.4
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)	☒	☐	☐	9.3.1; 9.4
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	9.3.3; 9.4
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	9.3.4; 9.4
9.3 Is a coding system described for:				

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	9.3.1
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	9.3.3
9.3.3 Covariates and other characteristics?	☒	☐	☐	9.3.4
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☐	☒	☐	

Comments:

The coding of COVID and covid-19 vaccines will be assessed in feasibility analysis. Since COVID confirmation will came from lab result it easy that it will not adjust systematic classifications but original ones. Regarding vaccines, we expect that the either ATC07 (J07????) or National Code (CNF) specific for pharmaceuticals in Spain is available for all vaccinations.

For a particular patient, data are linked in the regions that provide the complete data to the Data Source (BIFAP) based on a unique identifier.

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	9.7.3.1
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	9.5
10.3 Are descriptive analyses included?	☒	☐	☐	9.7.1
10.4 Are stratified analyses included?	☒	☐	☐	9.7.4
10.5 Does the plan describe methods for analytic control of confounding?	☒	☐	☐	9.7.3.2
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☒	☐	9.7.4.3*
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	9.7.5
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	9.7.4.2

Comments:

* Quantitative bias analyses will not be performed to evaluate the potential impact of differential misclassification of the outcome on the observed study estimate. This is because no validation studies are planned for the outcome (identified through lab results considered gold-standard in the clinical practice).

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)	☐	☒	☐	9.8*
11.2 Are methods of quality assurance described?	☐	☒	☐	9.8*
11.3 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

*Data storage and quality assurance will be performed according to the BIFAP governance that can

be consulted in the website of BIFAP.

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?	☒	☐	☐	9.9¶
12.1.2 Information bias?	☒	☐	☐	9.9*
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).	☒	☐	☐	9.9Ψ
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)	☒	☐	☐	9.5

Comments:

¶Selective recruitment into the study of vaccinated/unvaccinated subjects from those recorded in the database with quality criteria (up-to-standard information) that are not representative of the general vaccinated/unvaccinated subjects respectively in the source population could produce selection bias. Some examples are described in the protocol '9.9 Limitations of the Research Methods' section.

* If vaccinated individuals had less likelihood of screening or test than unvaccinated (due to a feeling of induced immunoprotection among vaccinated), we would artificially observe more cases among unvaccinated, directing toward a biased higher vaccine effectiveness estimation. The opposite could also be true, i.e. if vaccinated individuals had more likelihood of screening because they have a health/test-seeking behaviour, we would artificially observe more cases among vaccinated, directing toward a biased lower vaccine effectiveness estimation.

ΨNo information about the job and type of residence will be available. Consequently, confusion may still be present due to the higher probability of infection among them versus other social groups. That aspect would direct towards a reduction in the effectiveness estimations.

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	10
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	Ψ
13.3 Have data protection requirements been described?	☐	☒	☐	*

Comments:

ΨNot yet sent to ethics committee review.

*Reference to the BIFAP governance document (www.bifap.aemps.es) have been provided regarding data protection.

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	5

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)?	☒	☐	☐	6,12*
15.2 Are plans described for disseminating study results externally, including publication?	☒	☐	☐	12

Comments:

*The study progress reports as included in the section 6-Milestones and Timeline will be circulated among the participants of the collaborating public institutions for communication and review.

Name of the main author of the protocol: Elisa Martín Merino

Date: 20/April/2021

Signature: 

[The most current revision of the ENCePP Checklist for Study Protocols at the time of protocol finalisation, available here: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml]