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)
Version identifier of the final study report	
Date of last version of the final study report	15 February 2023
EU PAS Register number	EUPAS41134
Active substance	1. COVID-19 mRNA Vaccine (nucleoside modified) 2. (ChAdOx1-S [recombinant]) made up of another virus (of the adenovirus family) 3. Recombinant, replication-incompetent adenovirus type 26 (Ad26) vectored COVID-19 vaccine (Ad26.COV2-S [recombinant])
Medicinal product	1. Comirnaty 2. COVID-19 Vaccine Moderna 3. VAXZEVRIA 4. COVID-19 Vaccine Janssen
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 COVID1-9 diagnosis.
Country(-ies) of study	Spain
Author	Elisa Martín Merino; Belén Castillo Cano 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

Table of Contents

1. 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

Keywords: COVID-19, effectiveness, vaccine, real-world, observational, cohort study

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

Research question and objectives: To estimate the effectiveness ≥7 (PF)/14 (MD, AZ, JA) days after full vaccination with COVID-19 vaccines in reducing the number of cases with

- SARS-CoV-2 infection, i.e. a positive test (whether PCR or antigens) regardless the symptoms and prognosis
- Hospitalisations/Intensive Care Unit (ICU) admissions with a diagnosis of COVID-19/SARS-CoV-2 or pneumonia from −30 to +120 days of a positive test result
- Death with COVID-19 in 56 days after a positive test result

by vaccine brand and along the time since complete vaccination (7/14, 60, 90, 120 days).

Interpretation focused on the effectiveness of full vaccination while the estimates for the 1st dose (or PF, MD or AZ) were used as evaluation of potential confounding and comparability.

Study design: This is an observational cohort study to compare the occurrence of COVID-19 outcomes among vaccinated individuals and unvaccinated matched controls.

Setting and study period: Four central and northern Spanish regions from March till October 2021. The study period began on the 1st March, i.e. when the first general population age group (≥80 years) was called to vaccinate. The end of study period was the last date of data received by each region, i.e. October (3 regions) or June (1 region).

Data sources: Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (BIFAP) including primary healthcare records and linked registries of 1) SARS-COV-2 positive test results, 2) hospital discharge COVID-19 diagnosis or 3) COVID-19 vaccinations

Subjects and study size: Individuals with up-to-standard and ≥2-year baseline information during the study period and without COVID-19 vaccinations or outcome ever before, were included. No restriction to age or sex was applied.

Exposure assessment: The 1st vaccination during the study period was the time zero for patients included in the vaccinated group. A person unvaccinated on or before time zero of the vaccinated pair (1:1) was randomly selected among those with similar year of birth, sex and region and previous COVID-19 (matched criteria for control selection). Controls may be vaccinated later and then contributed to the vaccinated group. Only pairs in which both individuals were still at risk to COVID-19 outcomes 7/14 days after the second dose, could participate in the estimation of the effectiveness of the complete vaccination.

Variables:

- COVID-19 vaccinations were identified from a vaccination registry linked with each patient's primary care clinical history.
- Medically attended diagnosis of COVID-19 in any setting was identified through linked registries of 1) SARS-CoV-2 positive test results, 2) hospital COVID-19 diagnosis or discharged COVID-19 diagnosis, 3) administrative death.
- Covariates were collected from primary care, included age, sex, region, morbidities, medication and health care utilisation.

Data analysis

Vaccine effectiveness (VE) of PF, MD, AZ and JA's vaccines were estimated by 1 minus the hazard ratios (HR; 95% CI) calculated in flexible parametrical (FPM) regression models, overall and by time periods (e.g., at 7/14, 60, 90, 120 days after initiation and complete vaccination) and age. Characteristics at time zero (baseline) and visits before 1st and 2nd dose were used to adjust the risk model. A post-hoc analysis removing hospitalisations for other primary reasons was performed.

Results

2,351,174 individuals were identified to analyse Pfizer effectiveness (1,588,039 pairs), 486,517 to ModeRNA (264,033 pairs), 326,038 to AstraZeneca (226,670 pairs), and 257,719 to Janssen (135,569 pairs) between March and October 2021 who were free of prior infection.

Effectiveness against SARS-CoV-2 infection

At complete vaccination the numbers of pairs at risk were 229,139 for Pfizer (PF) analysis, 28,817 ModeRNA (MD), 7,684 AstraZeneca (AZ) and 33,148 Janssen (JA).

SARS-CoV-2 infection among vaccinated and controls were 599 and 606 for PF, 34 and 54 for MD, 26 and 26 for AZ and, 327 and 122 for JA.

The VE for PF was 88% (71%-95%) for patients aged 12-17 years, 15% (2%-26%) for aged 18-59 years and 24% (non-significant: -2% to 43%) for aged ≥60 years. VE was no longer significant 90 days afterwards (33%; -1% to 55%). For MD and AZ, the VE was 57% (34%-72%) and 18% (not significant: -42% to 53%) respectively. Effectiveness was not observed for JA but an increased risk of risk among vaccinated (HR: 2.21; 95% CI: 1.77-2.76). A precision study showed a higher proportion of missing positive lab-test among vaccinated than controls that would be overestimating the VE.

Effectiveness against hospitalisation with COVID-19 from 30 days before to 120 days after

Pairs at risk were 229,859 for Pfizer (PF) analysis, 28,994 ModeRNA (MD), 7,727 AstraZeneca (AZ) and 33,262 Janssen (JA).

The admission 'with' COVID-19 among vaccinated and controls were 41 and 109 for PF, 4 and 9 for MD, 1 and 10 for AZ and, 33 and 26 for JA.

The VE for PF against hospitalisation 'with' COVID-19 was 79% (95% CI: 68-87%), being 92% (80-97%) for those aged 18-59 years and 64% (40-79%) for ≥ 60 years. For MD and AZ, the VE was 74% (16-92%) and 92% (43-99%) respectively. Effectiveness was not observed for JA (HR=0.96; 95% CI: 0.56-1.63).

After manual review, the hospitalisations for other reasons were 6 (15%) and 10 (9%) for PF, 3 (75%) and 1 (11%) for MD, none for AZ, and 14 (42%) and 6 (23%) for JA.

The VE for PF against hospitalisation 'for' COVID-19 was 77% (66-85%), being 91% (81-96%) for aged 18-59 years and 66% (41-81%) for ≥ 60 years. For MD, the EV was 93% (47-99%). Effectiveness was neither observed for JA (HR=0.69; 95% CI: 0.37-1.28).

Effectiveness against death with COVID-19 during 8 weeks before

Among pairs still at risk 7 days after the complete vaccination with Pfizer, 2 deaths with COVID-19 among vaccinated and 5 among controls were identified. All cases were ≥60 years-old and the VE was 87% (57%-96%) adjusted by influenza vaccinations in the 5 years before. Table 7.

Evaluation of comparability at initiation of the vaccination

We observed differences in the incidence of SARS-CoV-2 infection between vaccinated and controls during the induction period of the 1st dose: at baseline, adolescents vaccinated with PF were more protected than controls (57%) as well as adults vaccinated with AZ (47%), while adults vaccinated with PF or MD were at higher risk of SARS-CoV-2 infection

(>50% and 21%, respectively). That can not be attributed to an effect of the vaccines and suggests confusion.

We did not observe differences in the incidence of hospitalisation with COVID-19 during the induction period of 1st doses of PF, AZ or MD vaccines suggesting comparability between vaccinated and controls. On the contrary, during the induction period of Janssen vaccine, we observed 3 times higher risk of hospitalisation with COVID-19 among vaccinated than controls limiting their comparability.

Discussion

The effectiveness of complete vaccination against **hospitalisation with COVID-19** during alpha and delta predominance was moderate-high in three out of four vaccines, even though the simple size was low, and a few numbers of cases occurred. Good comparability of vaccinated and controls was suggested by comorbidity and comedication patterns and the lack of immediate effect after the initiation of the vaccinations. Hospitalisations for other primary reasons were found, which mainly implied that ModeRNA effectiveness was higher than estimated (in 19%).

The duration of the immunity could only be calculated for Pfizer's vaccine which was reduced or lost three months after the complete vaccination.

Janssen vaccine was not shown effective against hospitalisation with COVID-19, even though the VE estimation increased substantially after removing hospitalisations for other reasons. Unmeasured non-clinical confounders could be substantial for that single dose vaccine recommended mainly to people difficult to be contacted for a second dose. Also, the delta variant could explain part of the lack of effectiveness found.

The effectiveness of Pfizer against **death with COVID-19** was high (87%) for people aged ≥60 years. That VE was lower than other studies in Spain. Certain underestimation could be consequence of the broad outcome definition (since cause of death was not available, we evaluated deaths with a SARS-CoV-2 infection within 56 days), adjustment or varying characteristics of cohorts.

Regarding **SARS-CoV-2 infection**, vaccinated and controls were at different baseline risk of infection that could not be completely controlled for after matching, restriction and adjustment. Consequently, part of the VE estimated for PF (VE high among adolescents and very low among adults), MD (moderate VE) and not statistical effect of AstraZeneca could be due to unmeasured factors or differential behaviour (testing) after being or not vaccinated rather than the vaccines' effects.

The manual review of clinical histories with COVID-19 diagnosis in primary care confirmed some true episodes without a linked positive test, for different reasons. That could imply an overestimation of the VE and may explain some differences with other studies with potentially higher precision to capture that outcome.

Larger study populations are required to elucidate questions not answered in the current project.

2. List of Abbreviations

AEMPS	Spanish Agency of Medicines and Medical Devices
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público
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. Investigators

Principal investigator

Elisa Martín Merino, PhD Pharm, Epidemiologist,

Division of Pharmacoepidemiology and Pharmacovigilance. Spanish Agency of Medicines and Medical Devices.

Calle Campezo 1, Edificio 8, 1ª planta, Madrid

Co-Authors

Belén Castillo Cano, PhD Candidate, Mathematician, Spanish Agency of Medicines and Medical Devices. Protocol review, statistical analysis and report edition.

Collaborators

- Ana Llorente García, Computer Scientist, Spanish Agency of Medicines and Medical Devices. Programming of data extraction.
- Susana Monge Corella, CCAES; Dirección General de Salud Pública, Ministerio de Sanidad). Study protocol and statistical and epidemological advise.
- Aurora Limia Sánchez and Carmen Olmedo Lucerón (SG Promoción, Prevención y Calidad; Dirección General de Salud Pública, Ministerio de Sanidad). Reviewed the study protocol and study results.
- Amparo Larrauri and Clara Mazagatos Ateca (Departamento de Enfermedades Transmisibles. Centro Nacional de Epidemiología. CIBER Epidemiología y Salud Pública (CIBERESP) Instituto de Salud Carlos III). Reviewed the study protocol.
- Enrique Seco Meseguer. MD, Hospital La Paz Madrid. Participated in the study protocol.
- Beatriz Sánchez Delgado; Division of Pharmacoepidemiology and Pharmacovigilance. Spanish Agency of Medicines and Medical Devices (AEMPS). Participated in the validation study of cases of hospitalised COVID-19 and performed the corresponding manual reviews.
- Oliver Astasio González; Clinical Pharmacology Dept, Hospital Clínico San Carlos, IdISSC, Madrid (OA was affiliated to this hospital when the study was performed). Collaborated in the SARS-COV-2 validation study.
- Rosa Gini. Epidemiology unit, Agenzia Regionale di Sanità della Toscana, Florence, Italy. Collaborated in the SARS-COV-2 validation study.
- Fabio Riefolo Teamit Institute, Partnerships, Barcelona Health Hub, Barcelona, Spain. Collaborated in the SARS-COV-2 validation study.

4. 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].

5. 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 were 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 was 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 el Ámbito Público (BIFAP) were used [65, 67].

BIFAP has been previously validated for research in pharmacoepidemiology, including the estimations of the precision of several clinical outcomes [65] that were 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 was included the linked official COVID-19 vaccinations. The algorithm for maximizing the precision and sensitivity was 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 intended 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 was based on a template that outlined a general approach to designing and implementing a comparative study of the effectiveness of the newly approved COVID-19 vaccines during the study period using existing health care databases [68].

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

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

This study was approved by the Ethical Committee for Research with Medical products (CEIm regional de la Comunidad de Madrid) under the reference code BIFAP_02_2021).

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

6. Research Question and Objectives

This study addressed the research question of whether vaccinations with original licenced COVID-19 vaccines, (Comirnaty, ModeRNA, AstraZeneca's and Janssen approved vaccines

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

Primary objective was 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 were splitting the effectiveness

- by the following outcomes:
  - Hospitalisations/Intensive care unit (ICU) admissions for COVID-19 –30 to +120 days of a SARS-CoV-2 positive test result at any time linked with the test result.
  - Death up to 8 weeks of a SARS-CoV-2 positive test.
- over age subgroups (12-17, 18-59 and ≥60 years)
- along the time since complete vaccination (up-to 7/14, 60, 90, 120 days and ever after).

Assess and address misclassification of the hospitalisation outcome through estimation of the predictive values is critical to avoid biased effectiveness estimates [79]. Thus post hoc evaluations were performed to correct estimates and interpretations.

7. Amendments and Updates

The following amendments and updates were implemented:

- We observed that prioritised criteria were not available in BIFAP (i.e. no information about nurse home residents or social-healthcare professionals) and cohorts could not be made comparable. Consequently, we restricted the study period to the calendar moments and year of birth in which the general population were called to vaccinate in Spain.
- Algorithms to identify outcomes were applied according to the data received from the participants regions, that could be different among them, and changes in the recording system (clinical dictionary updates) in the real clinical practices (i.e. inclusion of codes).
- Results were not stratified by Symptomatic or Asymptomatic Diagnosis of COVID-19 due the complexity of symptoms and recorded dates and expected under recorded mild symptoms. Only Laboratory-confirmed COVID-19, Hospitalisation with COVID-19 or Death with COVID-19 were studied.
- Evaluation of the positive predictive values of the COVID-19 diagnosis in primary care against positive test as gold standard and identification of the cause of

hospitalisation among patients assumed hospitalised with COVID-19 were performed.

- Interpretation focused on the effectiveness of full vaccination while the estimates for the 1st dose were used as evaluation of potential confounding and comparability.
- Covariates were collected at the date of the 1st dose. Visits were also collected at the date of the 2nd doses.
- Flexible parametrical models were used instead of PH Cox models to allow for non-proportionality of hazards.
- Competing risk in survival analysis were not considered.
- Apart from elderly, no other subgroup analysis was conducted due mainly to having insufficient sample size.
- Post-hoc we used the date of the positive test (instead of the date of hospital admission) as outcome date.

8. Research Methods

9. Study Design

An observational cohort study of people prioritized for COVID-19 vaccination was 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 was compared as defined below.

This study primarily described 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 starting on 21st December 2020 and stopping in Octobre 2022.

The choice of an exchangeable comparator group was a key factor in ensuring a valid study design and addressing confounding. In the current protocol, individuals not yet receiving COVID-19 vaccines was 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: was not as frequent or representative or a normal non-pandemic situation when people were 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 was more complicated to implement. Also, the variance estimation would need to be adjusted accordingly (e.g., by bootstrapping) because individuals were 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.

10. *Study Feasibility Stage*

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

1. data availability and data validity (Sections 10.1.1.1, 10.1.1.2, and 10.1.1.3)
2. valid study design (Section 10.1.1.4) during final protocol development.

10.1.1.1. Availability of Required Data Elements

We determined that main data elements for the implementation of the study, including several co-variates (see Section 12), inclusion and exclusion criteria (Section 11), vaccinations, COVID-19 outcomes, and potential confounders (Sections 10.1.1.5 and 15) were available.

10.1.1.2. Precision and Validity of the information about COVID-19 outcomes

Assess and address misclassification of the outcome is critical to avoid biased effectiveness estimates [79].

Test results (PCR, antigens, etc.)

In this study, COVID-19 outcomes were identified from the lab test results linked with primary care electronic health records (PC eHR) by the regional parties before sending to BIFAP to create its common data model. Laboratory-confirmed infection was linked with PC eHR and considered to identify cases for the main analysis.

During the pandemic emergency, COVID-19 was a mandatory notifiable disease to the competent national authorities [80].

We estimated the precision of the positive test linked to BIFAP data through a manual review of a random sample of 100 anonymised clinical histories of individuals with COVID-19 diagnosis recorded in PC but no linked positive COVID-19 test (false positives). Complete methods are included in Annex–Section 1. This review was blinded to vaccination status. 'False positives' are also named 'false diagnoses' hereof to ease the interpretation

of the results. The aim of the review was to also find evidence to confirm or reject the diagnosis through PC physicians free text comments and explore the reasons for the lack of positive tests. Final estimations were used to precise the interpretation of the effectiveness estimated.

Also, this allowed to estimate validation parameters of the Covid-19 diagnosis recorded in primary care as informative to further outcome definitions (Annex-Section 3).

Hospitalisation with COVID-19

The potential misclassification of cases hospitalised for other reasons was assessed post-hoc by vaccination status and vaccine brand (Janssen-JA, ModeRNA-MD, Pfizer-PZ, AstraZeneca-AZ) (Annex-Section 2).

Death with COVID-19

The information about the administrative death was linked to the patients' PC clinical profiles and available to research. Cause of death was not available. A previous validation study in comparison with Spain mortality registry was performed proving high concordance (data not available).

10.1.1.3. Validity of Vaccine Exposure

Vaccination data for each patient included the vaccination date, dose, name of the vaccine, brand and batch for each jab. An algorithm was set to exclude entry error as reported in Exposure section and the overall trend of the coverage by age and month agreed with the expected according to the public health recommendations.

The precision and sensitivity of COVID-19 vaccinations records linked to BIFAP primary care profiles have been proved harmonised with ECDC aggregated coverages and among European data sources in different international studies [81].

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) was available in BIFAP, we could not identify those factors among unvaccinated people to control confounding. Also, as aforementioned the current study was restricted to populations where both, vaccination status and baseline and updated clinical history, could be accurately identified.

10.1.1.4. Definition of Time Zero

Time zero (time0) in the vaccinated group was the date when the 1st dose of COVID-19 vaccine was administered (i.e. recorded). That date was also used for the unvaccinated pairs as Time zero. All eligibility criteria were also fulfilled, and covariates were collected, at that point, and COVID-19 outcomes started to be counted at that point [14]. Aligning those moments intended to avoid selection bias and immortal person-time bias [13].

The aim of the **Main analysis** was the estimation of the effectiveness ≥7/14 days after full vaccination series whether 1- or 2-doses regimens. Second doses were included in this study if administrated ≥19 (PF), ≥25 (MD) and ≥21 (AZ) days after the first dose to allow for real-world delayed completion of 2-dose vaccine series. Analysis was performed by vaccine brand. Only pairs in which both individuals were still at risk to COVID-19 outcomes 7/14 days after the second dose, could participate in the estimation of the effectiveness of the full vaccination.

In order to properly observe the effect of the vaccination along the time, the following cut-of periods were identified after time zero along each patient's follow-up contribution:

- D1-induction time: up-to 13 days after time zero
- D1-effectiveness: 14 days after time zero to 2nd dose
- D2-induction time: up-to 6 (PF) or 13 (MD, AZ) days after 2nd dose
- D2-effectiveness in the short run: ≥7 (PF) or ≥14 (MD, AZ) days after 2nd dose
- D2-effectiveness in the medium run: ≥60 days after 2nd dose
- D2-effectiveness ≥90 days after 2nd dose
- D2-effectiveness in the long run ≥120 days after 2nd dose

10.1.1.5. 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 performed the following strategies in order to minimise then: unvaccinated comparator was matched by time zero of the vaccinated pair, birth year cohort (year; it was opened to age groups called to vaccinate if not enough patients to select), sex, region of the primary care practice and prior COVID-19. Several other confounders were controlled through adjustment in the risk model.

11. Setting

This study was 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 changed 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 were established as much as possible.

The 1st prioritized groups (see Figure above) was called to vaccinate on the 21st of December 2020. The study period began on the 1st of March, i.e. when the first general population age group (≥80 years) was called to vaccinate. The end of study period was the last date of data received by each region, i.e. October (3 regions) or June (1 region). 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 were included in this study.

12. Subjects, Exposure Assessment and Time At Risk

Data was collected from BIFAP database from Spain (described in Data Sources).

Inclusion criteria to the Source population: up-to-standard, ≥2-year baseline information on 27th December 2020 in BIFAP. The last date of filling all the inclusion criteria was the inclusion date. No restriction to age or sex was applied since the vaccination for

young people was requested during this protocol development (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).

Exclusion criteria: Patients with prevalent COVID-19 vaccinations or infections ever before (i.e. washout period) the inclusion criteria were excluded as well as vaccination not fulfilling the quality algorithm regarding distance between doses, brand information and dates.

Exposure assessment: The 1st vaccination recorded after the inclusion date was 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, a person unvaccinated on or before time zero of the vaccinated pair (1:1) was randomly selected among those with similar birth cohort, sex and region and previous COVID-19 (matched criteria). Patients included in the comparison group, may be vaccinated later during their follow-up period and then contributed 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 was included in the vaccinated group. While a random selection of the unvaccinated individuals was included as unvaccinated group.

The date and type of vaccine administrated was collected. In particular, COVID-19 vaccinations were ascertained from official COVID-19 vaccination registries and provided by the regional governments.

For the vaccines available, those with a 2-dose regimen, the second dose must be administrated ≥19 days after the 1st dose of Comirnaty, 25 days of Spikevax and 21 days of Vaxzervria taken into account guidelines recommendations and intervals used in trials [82-86]. Also, in order to identify potential data entry errors (e.g., two vaccination records too close or 3 doses recorded, may indicate double recording of the same vaccination) and cut-off points, a checking of interval between doses was performed beforehand.

Studied vaccines were intended to give protection at least for 21 and 28 days after the 2nd dose, however the protection in the median-long run was unknown. Thus, in order to estimate the duration of the effectiveness, the follow-up period after 2nd dose was divided as aforementioned whenever follow-up time was long enough.

Stop date: the last moment of the following criteria defined the stop date to contribution: COVID-19 outcome, death, loss to follow-up, vaccination of the control and end of study period (the last date available in the data source at data collection).

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

13. *Eligibility Criteria*

Study cohort was restricted as explained in Setting section.

14. *Outcome Assessment*

Outcomes was assessed during the follow-up period (Section **¡Error! No se encuentra el origen de la referencia.**). Multiple effectiveness outcomes associated with vaccination for COVID-19 were 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.

14.1.1.1. Laboratory-confirmed COVID-19

As Main outcome: cases was incident **laboratory-confirmed COVID-19** infections recorded in a COVID-19 registry (regardless the symptoms, severity or the result of the COVID-19 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 was linked with BIFAP's primary care electronic health records (PC eHR). The date of the COVID-19 cases was the 1st date of a laboratory confirmed COVID-19 as recorded in the lab registries linked to PC eHR. If the date of admission to hospital/ICU with COVID-19 was earlier, this was used instead for analysis.

14.1.1.2. Hospitalisation with COVID-19

In a secondary analysis, the outcome was hospitalisation with diagnosis of SARS-CoV-2 infection, COVID-19 or pneumonia, with a positive COVID-19 test from 30 before to 120 days after of admission to hospital (inpatient record) or intensive care unit as the setting of care. Those type of identification was restricted to the region providing hospital data.

The date of the outcome was the admission date.

In that analysis, remaining non hospitalised COVID-19 infections participated as non-cases. The stop date for those was death, loss to follow-up or end of study period whichever came earlier.

A manual review, blinded to vaccination status, of patient clinical notes, was performed to identify any alternative primary reasons for the hospitalisation. This allowed the analysis of 1) all automatic cases identified and 2) excluding admissions for other reasons.

14.1.1.3. Death with COVID-19

The cause of death was not available. Thus, death up-to 8 weeks [73] after a positive SARS-CoV-2 or hospitalisation for COVID-19 was utilized as a third outcome. The date of the outcome was the date of death.

15. *Covariate Assessment*

To control for measurable confounders in the analysis, the following variables was collected: comorbidities and comedications as reported in Tables 1, and primary health care utilisation prior to or at time zero at the patient level, and health care utilisation was also collected when receiving the 2nd dose. Given that risk factors for COVID-19 infection were not well understood [30] at this protocol development, we measured covariates that are classically controlled for in pharmacoepidemiologic research.

As explained in Data source, clinical events were collected by using the International Classification of Primary Care (ICPC) and the International Classification of Diseases 9th Revision (ICD-9) medical terms [70, 71] mapped to Systematized Nomenclature of Medicine (SNOMED).

The following comorbidities was assessed ever before time zero:

- 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 was 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 was assessed during the 180 days before time zero and updated:

- Antibiotics
- Antiviral medications
- Corticosteroids
- Non-steroidal anti-inflammatory drugs
- Other Analgesic
- Psychotropics

- Statins
- Immunosuppressant medication use

Primary health care utilisation (visits) in the year before time zero and 2nd dose was evaluated as measures of health care–seeking behaviour, overall health status, and access to health care. Additionally, visits in the week before and including time zero and date of 2nd dose were recorded separately, as short-term markers of current health status may influence individuals' vaccination decisions.

Considered variables included the following:

- Influenza vaccination (number in the previous 5 years at time zero)
- Other vaccinations (number in the previous 5 years at time zero)
- COVID-19 negative tests (number before time zero and date of 2nd dose; available in 3 out of 4 regions)

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] were included as confounding factors, and included the following:

- Parkinson's disease
- Stroke
- Dementia
- Sepsis
- Heart failure
- Diabetes complications
- Coagulation deficiencies

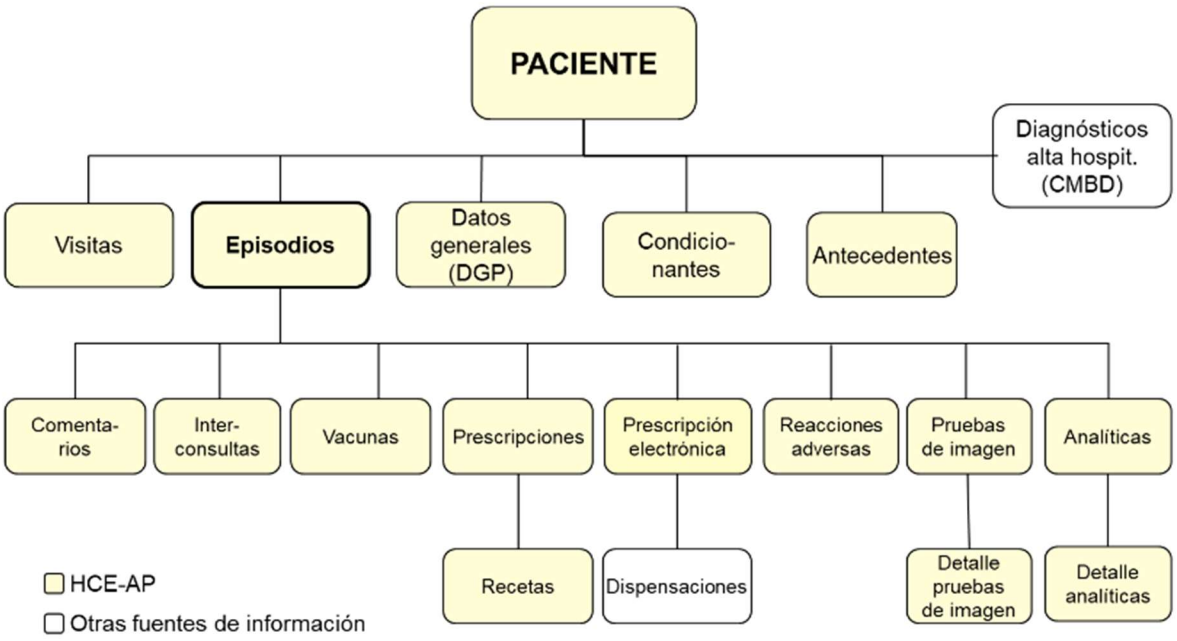
16. Data Source

Data on patients was 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 were unified in the data source so-called "Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público" (BIFAP) [65, 67]

.

BIFAP includes information about patients' age, sex, clinical events (recorded by using the International Classification of Primary Care (ICPC) and the International Classification of Diseases 9th Revision (ICD-9) medical terms [70, 71], 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 (*Porcentaje de población cubierta por el sistema sanitario publico. Encuesta Nacional de Salud. MSSSI/INE. Indicadores clave*, 2018)). In addition, 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 were being incorporating to BIFAP. Figure 2:



BIFAP is fully funded by the Spanish Agency of Medicines and Medical Devices (AEMPS), belonging to the public Department of Health, and is maintained with the collaboration of the participant regions [67].

The records on the results of the COVID-19 laboratory test were linked with BIFAP primary care profiles. Also, all vaccinations administrated against COVID-19 recorded in any setting was linked.

BIFAP has been previously validated for research in pharmacoepidemiology, including the estimations of the precision of several clinical outcomes [65] that were covariates of the current study (such as cardiovascular/digestive/endocrinological outcomes). Also, other vaccinations recorded in BIFAP have been confirmed valid through formal precision studied or against external comparable data sources (Braeye et al. Vaccine. 2020; Martín-Merino et al. PDS 2019).

17. Study Size

Taking into account the estimated COVID-19 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-19 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 21.1.1.1] 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-19 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 9,897 vaccinated and 114 cases (1.145%) among 9,955 in comparison group. Efficacy: 95.6% (89.4–98.6)
2. >55 years: 3 cases (0.04%) de 7,500 vaccinated and 48 cases (0.636%) among 7,543 in comparison group. Efficacy: 93.7% (80.6– 8.8)
3. ≥65 years: 1 case (0.025%) de 3,848 vaccinated and 19 cases (0.489%) among 3,880 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 10,551 vaccinated and 156 cases (1.482%) among 10,521 in comparison group. Efficacy: 95.6% (90.6-97.9).
2. ≥65 years: 4 cases (0.111%) de 3,583 vaccinated and 29 cases (0.816%) among 3,552 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 12,750 vaccinated and 260 cases among 12,782 in comparison group. Efficacy: 63.7% (53.9-71.6).
2. ≥60 years: 21 cases among 6,764 vaccinated and 88 cases among 6,762 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 were 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 was provided 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 was not calculated.

18. Data Management

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

Previous to that, the different registries were linked to primary care histories by the local government teams of regions participating in BIFAP. Those pseudonymised individual-level was sent to BIFAP IT and held at the AEMPS research sites. The Governance of data use is publicly available [69].

STATA 15.1 was used as statistic analytic program.

19. Data Analysis

20. *Descriptive Characteristics*

The distributions of baseline characteristics by exposure group/periods were calculated to describe the study cohorts and illustrate differences between the groups compared. For categorical variables, counts and proportions were estimated. Variables with missingness (BMI, smoking status and alcohol use) was not available.

21. *Comparative Analysis*

21.1.1.1. Estimation of Vaccine Effectiveness

Flexible parametrical (FPM) regression models, yielding a hazard ratio (HR; 95%CI)**¡Error! No se encuentra el origen de la referencia.** was estimated for the defined periods: i.e. 0 to 13 days after D1, 14 days after D1 until D2, ≥7 or 14 days after D2 and ≥60 days, ≥90 days, and ≥120 days after D2 in comparison with similar followed periods from time zero in the unvaccinated group.

The assumption of proportionality of the survival curves in the vaccinated and unvaccinated groups was evaluated for proper specification of the model.

21.1.1.2. Adjustment for Confounding

Factors associated with the outcomes with a p-value <0.05 were used to adjust a preliminary risk model, and later excluded if not confounding the VE estimate (to reach parsimonious models). FPM calculated a single vaccine effectiveness measure for each period of observation (assumed constant over those defined periods) in vaccinated versus unvaccinated groups. Adjusted vaccine effectiveness was calculated for all the outcomes as 1 minus HR (and 1 minus the 95% UCI or LCI). [38].

21.1.1.3. Conducting Subgroup Analyses

The primary analyses were stratified by age categories.

21.1.1.4. Correcting for Differential Outcome Misclassification

For COVID-19 infection (i.e. positive SARS-CoV-2) a validation study was performed to estimate the number of COVID-19 cases without a linked positive test that could be true infections. Validation included a comparison with the COVID-19 diagnosis recorded in primary care and a manual review to confirm the diagnosis and the reasons of missing linked test result. Methods and results are provided in Annex-Section 1.

Quantitative bias analyses were performed to evaluate the potential impact of differential misclassification of hospitalisation with COVID-19 on the observed study estimate. Methods and results are provided in Annex-Section 2.

22. *Missing Data*

An algorithm of quality was applied to identify the vaccinations linked to BIFAP. In summary, records with missing date or brand of the COVID-19 vaccine administrated were excluded from the study, that represented less than 6 missing every 100,000 of total recorded vaccinations.

Records with missing 'dose' were imputed according to the sequential order/dates of the vaccinations recorded in a particular person.

SARS-CoV-2 negative lab-results were missing in one of the participant regions so we could not control by that factor.

No missing data were allowed in other variables, i.e. comorbidity and comedication were considered yes/not the condition was present; age or other continuous variable was always complete.

23. Post-hoc analyses

We performed two post-hoc analyses:

In the first post-hoc analysis, we removed the hospitalisations for other primary reasons (different from COVID-19) as identified during manual review of clinical histories.

In the second post-hoc analysis, we used the date of the SARS-CoV-2 positive test as the outcome date for cases of hospital/ICU with COVID-19 (instead of using the date of admission) and as stop date for people with non-hospitalised SARS-CoV-2 infection. In this analysis, patients with prior SARS-CoV-2 infection were removed.

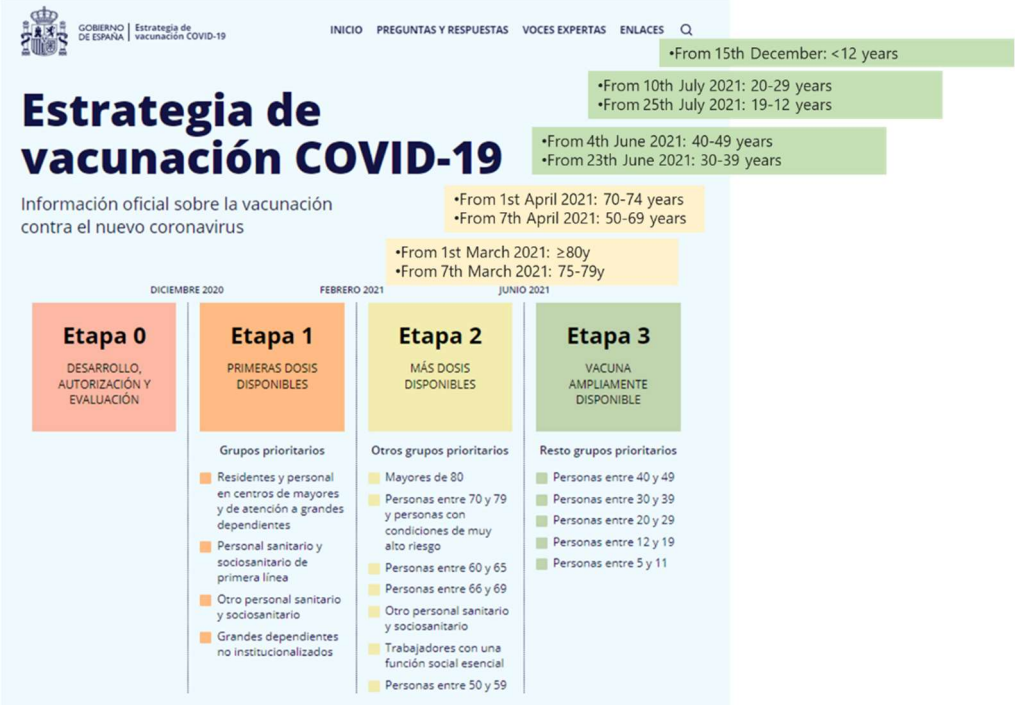
24. Quality control

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, was performed according to the BIFAP governance that can be consulted in the following linked of BIFAP website [69].

25. Results

26. Participants

Identification of the study population was according to the general population age groups called to vaccinate and calendar moment as reported in Figure 3.



In the current study, the same patient could contribute to different pairs, as vaccinated and control, according to its periods of observations.

Thus, to estimate effectiveness, 2,351,174 individuals were identified to Pfizer analysis (1,588,039 pairs), 486,517 to ModeRNA (264,033 pairs), 326,038 to AstraZeneca (226,670 pairs), and 257,719 to Janssen (135,569 pairs) analysis between March and October 2021 who were free of prior infection. Table 3.

27. Descriptive data

The number of pairs still-at-risk at 7/14 days after complete vaccination, that contributed information to the full VE estimation is included below for each outcome of interest.

Baseline comorbidity and comedication were more frequent among vaccinated than controls, particularly among patients aged ≥60 years. The distributions of baseline characteristics among vaccinated and controls contributed to effectiveness against infection is reported in Tables 4-5 for Pfizer (stratified by age), ModeRNA, AstraZeneca and Janssen.

28. Outcome data

The number of cases automatically identified among people vaccinated and controls is reported in the following Main results section per each analysis and outcome.

29. Main results

30. *Effectiveness against SARS-CoV-2 infection*

Table 6 shows the VE against SARS-CoV-2 infection by vaccine.

Among pairs still at risk 7 days after complete vaccination with Pfizer (N=229,139), 599 cases of SARS-CoV-2 infection occurred among vaccinated and 606 among controls, in a maximum of 216 days of follow-up, resulting in a VE of 22% (95% CI: 12%-32%), being 88% (71%-95%) for patients aged 12-17 years, 15% (2%-26%) for aged 18-59 years and 24% (non-significant: -2% to 43%) for aged ≥60 years. Overall, VE was no longer significant 90 days afterwards (33%; -1% to 55%).

The variables remaining in the adjusted model were: any other vaccination during the prior 5 years (for people aged 12-17 years), hypertension and visits to primary care the year before the complete vaccination (2nd dose) (for aged 18-59 years) or antibiotics prescriptions and visits to primary care the year before the complete vaccination (2nd dose) (for aged ≥60 years).

Among pairs still at risk 14 days after complete vaccination with ModeRNA (N=28,817), 34 cases of SARS-CoV-2 infection occurred among vaccinated and 54 among controls in a maximum of 199 days of follow-up, resulting in a VE of 57% (34%-72%) adjusted by opioids prescriptions and visits to primary care the year before the complete vaccination. The simple size was not sufficient to estimate the VE by age groups or time windows after complete vaccination.

Among pairs still at risk 14 days after complete vaccination with AstraZeneca (N=7,684), 26 cases of SARS-CoV-2 infection occurred among vaccinated and 26 among controls in a maximum of 116 days of follow-up, resulting in a VE of 18% (not significant: -42% to 53%) adjusted by visits to primary care the year before the complete vaccination. The simple size was not sufficient to estimate the VE by age groups or time windows after complete vaccination.

Among pairs still at risk 14 days after complete vaccination with Janssen (N=33,148), 327 cases of SARS-CoV-2 infection occurred among vaccinated and 122 among controls in a maximum of 178 days of follow-up, resulting in an increased risk of SARS-CoV-2 infection among vaccinated of 2.21 (95% CI: 1.77-2.76) adjusted by obesity, antibiotic prescriptions or prescription of anxiolytic, hypnotic, sedative or antidepressant drugs and visits to primary care the year before starting the vaccination. No effectiveness of the vaccine was observed, so we did not split by age group or time windows after complete vaccination.

*Note: The methods and results to validate the precision of the SARS-CoV-2 information are reported in the Annex-Section 1. In summary, during the review of a sample of the people, we observed a higher proportion of missing positive lab-test among vaccinated than controls and the reasons behind. That would overestimate the VE against infection in the current study.

31. *Effectiveness against hospitalisation with COVID-19 (admitted to hospital or ICU)*

Table 7 shows the VE against hospitalisation with COVID-19 (admitted to hospital or ICU) by vaccine.

Among 229,859 pairs still at risk 7 days after complete vaccination with Pfizer, 41 cases of hospitalisation with COVID-19 were identified among vaccinated and 109 among controls, resulting in a VE of 79% (68%-87%), being 92% (80%-97%) for people aged 18-59 years and 64% (40%-79%) for aged ≥60 years. The VE was significant for the first three months, then it decreased to no significant (90 days of vaccination VE was 58%; -34% to 87%). No cases of hospitalisation with COVID-19 were identified among people aged 12-17 years.

The variables remaining in the adjusted model were COPD, prescription of analgesic drugs before initiating the vaccination, prescription of anticoagulants, and visits to primary care the year before the complete vaccination among people aged ≥60 years, and hypertension, prescription of antibiotics and visits to primary care the year before the complete vaccination among aged 18-59 years.

In a manual review of primary care clinical histories blinded to vaccination status (Annex-Section 2), we observed that COVID-19 was not the reason to hospital admission in 10 (9%) controls and 6 (15%) vaccinated individuals, according to physicians' free text comments. In a post-hoc analysis we removed all of them from the group of cases hospitalisation with COVID-19 resulting in a VE of 77% (66%-85%), being 91% (81%-96%) for people aged 18-59 years and 66% (41-81%) for aged ≥60 years. Table 8.

Among 28,994 pairs still at risk 14 days after complete vaccination with ModeRNA, 4 cases of hospitalisation with COVID-19 were identified among vaccinated and 9 among controls, resulting in an adjusted VE of 74% (16%-92%) adjusted by prescriptions of anticoagulants and visits the year before the full vaccination. The simple size was not sufficient to estimate the VE by age groups or time windows after complete vaccination. After manual review (Annex-Section 2), we observed that COVID-19 was not the reason to hospital admission in 3 out of 4 vaccinated (75%) and 1 out of 9 controls (11%). In a post-hoc analysis we removed all of them from the group of cases hospitalisation with COVID-19 increasing the VE estimation to 93% (47%-99%). Table 8.

Over 7,727 pairs still at risk 14 days after complete vaccination with AstraZeneca, 1 case of hospitalisation with COVID-19 were identified among vaccinated and 10 among controls, resulting in a VE of 92% (95% CI: 43%-99%) adjusted by visits. The simple size was not sufficient to estimate the VE by age groups or time windows after complete vaccination. After manual review (Annex-Section 2), we did not observe that hospitalisations were for other reasons, thus the VE did not change.

Among 33,262 pairs still at risk 14 days after complete vaccination with Janssen, 33 cases were identified among vaccinated and 26 among controls, showing no effectiveness (HR=0.96; 95% CI: 0.56-1.63) after final adjustment by prescriptions of opioids, anticoagulants, obesity and visits. No effectiveness of the vaccine was observed, so we did not split by age group or time windows after complete vaccination. After manual review (Annex-Section 2), we observed that COVID-19 was not the reason to hospital admission in double vaccinated (N=14; 42%) than controls (N=6; 23%). In a post-hoc analysis we removed all of them from the group of cases increasing the VE estimation to non-significant 31% (HR=0.69; 95% CI: 0.37-1.28). Table 8.

The above presented analyses were run among patients that could have a SARS-CoV-2 positive test between the dates of 1st and 2nd doses (but not before the 1st dose). We replicated the analysis removing all patients with prior SARS-CoV-2 infection, the VE varied in 1% for Pfizer (VE=78%; 65%-85%), since 5 hospitalisation with COVID-19 among controls were removed, VE varied in 6% for Janssen (HR=0.90; 95% CI: 0.53-1.55; since 2 cases among vaccinated were removed) and

VE remained similar for ModeRNA (74%; 16%-92%; where none case had infection before the 2nd dose) and for AstraZeneca (92%; 37%-99%; 1 cases among controls was removed).

To easy reading, table 8 displays the VE against hospitalisation with COVID-19 estimated in final model and post-hoc analyses.

32. *Effectiveness against death with COVID-19 during 8 weeks before*

Among pairs still at risk 7 days after the complete vaccination with Pfizer, 2 deaths with COVID-19 among vaccinated and 5 among controls were identified. All cases were ≥60 years-old and the VE was 87% (57%-96%) adjusted by influenza vaccinations in the 5 years before. Table 9.

33. *Evaluation of comparability at initiation and completeness of the vaccination*

We assume that had the matching reached comparable vaccinated and control groups (as randomisation does), we should not see an association during the induction period of the 1st dose. However, we observed differences in the incidence of COVID-19 during that period and after adjustment, which can not be attributed to an effect of the vaccines. That suggests baseline differences between vaccinated and controls of factors not controlled in the study.

For Pfizer, during the induction period of the 1st dose (13 days), adjusted VE against SARS-CoV-2 infection was statistically different between vaccinated and controls in some age groups, suggesting confusion and lack of comparability at time0 (date of 1st vaccination). Table 3:

- We observed a 57% (47%-65%) reduced risk of SARS-CoV-2 infection during 13 days after 1st dose among patients aged 12-17 years, meaning that vaccinated adolescents were more protected for reasons different from the vaccination (that were not considered in the study).
- We observed a 60% (46%-76%) higher risk of infection during 13 days after 1st dose among patients aged 18-59 years and 50% (from 10% to 2 times higher) among aged ≥60 years, meaning that vaccinated adults were at higher risk of SARS-CoV-2 infection at baseline that the study could not control for.

During the induction period of the 1st dose, the SARS-CoV-2 infections were 47% lower among vaccinated with AstraZeneca than controls (30%-59%) and higher among vaccinated with ModeRNA than controls (21%; 1%-44%).

For **hospitalisation with COVID-19**, we did not observe association during the induction period of PF, AZ or MD vaccines, suggesting comparability between compared groups from the vaccination initiation. On the contrary, during the induction period of Janssen vaccine, we observed 3 times higher risk of hospitalisation with COVID-19 among vaccinated than controls (with a confidence interval between 12% and 10 times higher).

Table 8 shows those differences after the 1st dose against SARS-CoV-2 infection and hospitalisation with COVID-19 by vaccine brands and age categories.

34. Discussion

35. Key results

In this observational study, the effectiveness of complete vaccination against hospitalisation with COVID-19 during alpha and delta predominance was observed moderate-high in three out of four vaccines, even though the simple size was low, and a few numbers of cases occurred.

The effectiveness of Pfizer was high for adults aged 18-59 years (92%) and moderate for people aged ≥60 years (64%) that was reduced or lost three months after the complete vaccination. The effectiveness was moderate (74%) for ModeRNA and high for AstraZeneca (92%). In those analysis, good comparability of the pairs was suggested by comorbidity and comedication patterns and the lack of immediate effect after the initiation of the vaccinations.

In some subgroups (i.e., aged ≥60 years with Pfizer vaccine) the estimation was lower than the calculated in randomised clinical trials and published observational studies. That could be consequence of our outcome definition or differential control of confounders among studies.

Regarding outcome definition, in post-hoc analysis, after removing hospitalisations for other primary reasons, an important underestimation was observed for ModeRNA (19%) in main analysis, little overestimation for Pfizer (2%) and VE for AstraZeneca was unaffected. That misclassification could explain differences with other studies.

That precision analysis could be performed thanks to the pseudonymised primary care physicians' free text comments available in BIFAP (that also included discharge letters among other secondary care information that were manually reviewed) when no such information was available from the hospital registries. As limitation, it is important to clarify that in a proportion of the cases, no information of the reason to hospitalisation was available so some misclassification could still be present. The effect of that misclassification on the VE would be more precise in bigger sample sizes.

Control of confounders could also be different among studies. It could be that patients living in nursing homes (and at-higher risk of infection or severe prognosis) were more

represented among vaccinated than controls in our study, underestimating the effectiveness, or other confounders of the real world (i.e. visits) not controlled for in other studies overestimating the effectiveness.

We did not observe effectiveness of Janssen vaccine against severe COVID-19, even though the VE estimation increased 27% after removing hospitalisations for other reasons. At the administration of the 1st and unique dose, baseline differences (regarding the risk of severe COVID-19) between vaccinated and controls were observed and controlled for, however unmeasured confounders could be substantial and not be controlled for after matching, restrictions and adjustment. Single dose Janssen vaccine was administrated mainly to people difficult to be contacted for a second dose with vaccines requiring two-dose schedules, which unmeasured non-clinical confounders (not available in BIFAP). Also, the cases of hospitalisation with COVID-19 in this Janssen analysis occurred in June-July with delta predominance that was associated, in the literature [87], with loss of effectiveness. So, the delta variant could explain part of the lack of effectiveness found.

The effectiveness of complete vaccination with Pfizer against death with COVID-19 among people aged ≥60 years was high (87%), but lower than other studies in Spain [87], that could be consequence of the broad study definition in our study (i.e. death with a positive test 56 days before, that may include death for other reasons), extra adjustment in the current study or different baseline characteristics among the participants (people with or without complete primary care profile, respectively).

The effectiveness of PF against SARS-CoV-2 infection (regardless the prognosis) was high (88%) among adolescents and very low (≤24%) among adults. As it is already well-known, the duration of the effectiveness was limited (lasting three months in our study). The effectiveness was moderate with MD (57%) and not observed for AstraZeneca (not significant 18%).

In those three analyses of SARS-CoV-2 infection (PF, MD, AZ) baseline confusion was observed, meaning that vaccinated and controls were at different baseline risk of SARS-CoV-2 infection that were not completely controlled for after matching, restriction and adjustment. Consequently, certain estimation could be due to other unmeasured factors.

Adults vaccinated with PF and MD shower higher baseline risk of SARS-CoV-2 infection than their unvaccinated pairs that suggested less protection against the contagious. As an example, among those aged ≥60 years for PF analysis, half of controls did never seek primary care the year before while most vaccinated (around 93%) did. Also, vaccinated had approximately 2-3 times more prevalence of co-morbidities and co-medication, and 42% of vaccinated, while 9% of unvaccinated, were immunised against influenza the years before. Suggesting other unmeasured confounders among vaccinated.
The opposite was true for adolescents (vaccinated with PF) and adults vaccinated with AZ, who showed a lower baseline risk of SARS-CoV-2 infection at the initiation of the vaccination than their unvaccinated pairs. They could be people more protected against transmission/infection (whether were they more careful or living/working in less contagious environments).

Furthermore, COVID-19 testing became universal along the time, thus laboratory confirmation of case status may not be available for some patients at the beginning of the study period. In the current study and according to the manual review of patients with COVID-19 diagnosis recorded in primary care, we observed that some infections did not have a linked positive test for different reasons. That could be higher in vaccinated than control that implied an overestimation of the VE in some extend. That could also explain some differences with other studies with higher precision to capture that outcome.

The estimation of the VE against SARS-CoV-2 infection could also be influenced by the (differential) testing frequency between vaccinated and controls, that makes analysis against infection to be less robust than analysis against hospitalisation with COVID-19.

Finally, when the date of testing was analysed instead of the date of admission to hospital among patients with hospitalisation with COVID-19, the HR and/or VE estimations were scarcely modified.

36. Limitations

Some limitations must be recognised. As aforementioned, after matching, restriction and adjustment, people vaccinated and their controls seemed to be comparable regarding the risk of hospitalisation with COVID-19, but not against all SARS-CoV-2

infections. Subsequently, it should be considered that potential confounding may also be present at complete vaccination against all SARS-CoV-2 infections.

For one of the participating regions, the test negative results were not available. Thus, overall results were not controlled by the potential effect of the vaccinated to be more tested that unvaccinated pairs (due for instance to be more vulnerable, more exposed or careful). Had we detected more infections among vaccinated due to more testing, the VE estimation would be underestimated. As a proxy, the adjustment by number of visits to primary care the year before could be indirectly (or partially) correcting that limitation. The opposite, i.e. mess frequent testing immediate after receiving the jab would overestimate the effectiveness. That have been observed in previous studies when lag-period for the induction of the immunity by the vaccination were not set.

Despite the high sample size reached in the study populations, the restriction to pairs similar in clinical subgroups or still-at-risk in time-windows reduced the sample sizes precluding the analysis of those subgroups of interest.

The analysis of paired cohorts is demanding as the couples are undone in the follow-up losing sample size. However, it maximises comparability of, for instance, calendar moment, which is quite relevant in the virus epidemiology.

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]. Confusion may still be present due to the higher probability of infection (and vaccination) among certain jobs (ex. healthcare workers) or type of residence (nurse home). That aspect would direct towards a reduction in the effectiveness estimations.

Selective recruitment into the study of vaccinated/unvaccinated subjects/moments 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 were more closely surveyed by the PCP/nurse due to their predisposition to complicated COVID-19 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.

Data from multiple regions and sources were 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 were used (e.g., records from general practice, hospital 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, some differences with other studies may be due to heterogeneity of the regional data, the underlying heterogeneity of confounding control, misclassification, or other data source factors that could not be 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.

37. Generalisability

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

38. Other Information

This study protocol was approved by the BIFAP scientific advisory board (reference number 02_2021) and Ethical Committee for Research with Medical products (CEIm regional de la Comunidad de Madrid; under the code number BIFAP_02_2021).

39. 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.

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. https://www.fda.gov/media/146267/download

79. Seeger J. Considerations of Misclassification and Confounding on COVID-19 Vaccines Effectiveness Studies- A Vaccine SIG Endorsed Symposium. ICPE Vol 67.; 2022. ICPE. 2022;67.

80. Sanidad M de. Boletín Oficial Del Estado Del 12 de Mayo de 2020. Orden SND/404/2020, de 11 de mayo, de medidas de vigilancia epidemiológica de la infección por SARS-CoV-2 durante la fase de transición hacia una nueva normalidad [Internet]. 2020. Available from: https://www.boe.es/boe/dias/2020/05/12/pdfs/BOE-A-2020-4933.pdf

81. Durán CE, Messina D, Gini R, Riefolo F, Aragón M, Belitser S, et al. Rapid Safety Assessment of SARS-CoV-2 Vaccines in EU Member States using Electronic Health Care Data Sources (COVID Vaccine Monitor-CVM study): Final Study Report for WP3 (electronic health record data) [Internet]. Zenodo; 2023. Available from: https://doi.org/10.5281/zenodo.10093613

82. CDC. Use of COVID-19 Vaccines in the United States [Internet]. Available from: https://www.cdc.gov/vaccines/covid-19/clinical-considerations/covid-19-vaccines-us.html

83. Ministerio de Sanidad. Guías Técnicas Vacunas COVID-19 [Internet]. Available from: https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/covid19/GuiasTecnicas_vacunaCOVID-19.htm

84. Ministerio de Sanidad. Vaxzevria, COVID-19 Vaccine (ChAdOx1-S [recombinant]) (europa.eu) [Internet]. Available from: https://www.mscbs.gob.es/gabinete/notasPrensa.do?metodo=detalle&id=5465 Vaxzevria, COVID-19 Vaccine (ChAdOx1-S [recombinant]) (europa.eu)

85. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med 2021; 384:403-416. DOI: 10.1056/NEJMoa2035389.

86. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2020 Dec 31;383(27):2603–15.

87. Grupo de Trabajo de Efectividad Vacunación COVID-19. ISCIII. 2º informe. Análisis de la efectividad de la vacunación frente a COVID-19 en España [Internet]. Available from: https://web.mscbs.gob.es/areas/promocionPrevencion/vacunaciones/covid19/docs/Efectividad_VacunacionCOVID-19_Espana.pdf

88. Astasio O, Castillo-Cano B, Delgado BS, Gini R, Riefolo F, Merino EM. The Validation Of Covid-19 information In The Pharmacoepidemiological Research Database for Public Health System by vaccination status [Internet]. Preprints; 2022 Dec [cited 2023 Dec 14]. Available from: https://www.authorea.com/users/569206/articles/614904-the-validation-of-covid-19-information-in-the-pharmacoepidemiological-research-database-for-public-health-system-by-vaccination-status?commit=c26574dcb3bf752a458bcb9c03347cc7212c3dcc

Appendices

Main result tables

Tables 4-5 include the description of baseline characteristics among vaccinated and controls still at risk after complete primary vaccination that contributed to the complete vaccination effectiveness analysis.

Table 4. Baseline characteristics of people fully vaccinated with Pfizer's vaccine and matched unvaccinated controls by age category.

	60+years				18-59 years				12-17 years			
	Controls		Vaccinated		Controls		Vaccinated		Controls		Vaccinated	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
	39346		39346		165625		165625		22948		22948	
Women	20648	52.5%	20648	52.5%	69229	41.8%	69229	41.8%	10703	46.6%	10703	46.6%
Birth year												
1900-1941	14270	36.3%	13772	35.0%	-	-	-	-	-	-	-	-
1942-1951	14870	37.8%	15611	39.7%	-	-	-	-	-	-	-	-
1952-1955	5020	12.8%	5069	12.9%	-	-	-	-	-	-	-	-
1956-1961	5186	13.2%	4894	12.4%	-	-	-	-	-	-	-	-
1962-1971	-	-	-	-	37311	22.5%	37490	22.6%	-	-	-	-
1972-1981	-	-	-	-	44528	26.9%	44514	26.9%	-	-	-	-
1982-1991	-	-	-	-	42986	26.0%	42821	25.9%	-	-	-	-
1992-2001	-	-	-	-	35231	21.3%	35205	21.3%	-	-	-	-
2002-2019	-	-	-	-	5569	3.4%	5595	3.4%	22948		22948	
Region												
2	16724	42.5%	16724	42.5%	4789	2.9%	4789	2.9%	-	-	-	-

3	5811	14.8%	5811	14.8%	26150	15.8%	26150	15.8%	1799	7.8%	1799	7.8%
7	12079	30.7%	12079	30.7%	73879	44.6%	73879	44.6%	9694	42.2%	9694	42.2%
14	4732	12.0%	4732	12.0%	60807	36.7%	60807	36.7%	11455	49.9%	11455	49.9%
Time0 (month)												
March	4755	12.1%	4755	12.1%	-	-	-	-	-	-	-	-
April	17531	44.6%	17531	44.6%	1069	0.6%	1069	0.6%	-	-	-	-
May	11546	29.3%	11546	29.3%	7209	4.4%	7209	4.4%	-	-	-	-
June	2190	5.6%	2190	5.6%	50416	30.4%	50416	30.4%	-	-	-	-
July	1402	3.6%	1402	3.6%	57672	34.8%	57672	34.8%	<5	0.0%	<5	0.0%
August	1248	3.2%	1248	3.2%	41994	25.4%	41994	25.4%	16049	69.9%	16049	69.9%
September	652	1.7%	652	1.7%	7100	4.3%	7100	4.3%	6838	29.8%	6838	29.8%
October	22	0.1%	22	0.1%	165	0.1%	165	0.1%	59	0.3%	59	0.3%
Month of full vaccination												
March	503	1.3%	503	1.3%	-	-	-	-	-	-	-	-
April	4326	11.0%	4326	11.0%	-	-	-	-	-	-	-	-
May	18044	45.9%	18044	45.9%	1076	0.6%	1076	0.6%	-	-	-	-
June	11244	28.6%	11244	28.6%	10556	6.4%	10556	6.4%	-	-	-	-
July	2033	5.2%	2033	5.2%	50308	30.4%	50308	30.4%	-	-	-	-
August	1323	3.4%	1323	3.4%	58090	35.1%	58090	35.1%	18	0.1%	18	0.1%
September	1349	3.4%	1349	3.4%	38986	23.5%	38986	23.5%	17682	77.1%	17682	77.1%
October	524	1.3%	524	1.3%	6609	4.0%	6609	4.0%	5248	22.9%	5248	22.9%
Flu vaccination in 5y prior												
No vaccination	35651	90.6%	22627	57.5%	161305	97.4%	156393	94.4%	22499	98.0%	21919	95.5%
<5 vaccinations	2978	7.6%	9149	23.3%	4096	2.5%	8375	5.1%	410	1.8%	911	4.0%
≥5 vaccinations	717	1.8%	7570	19.2%	224	0.1%	857	0.5%	39	0.2%	118	0.5%
Other vaccinations in 5y prior												

No other vaccinations	30379	77.2%	26400	67.1%	156657	94.6%	151551	91.5%	11312	49.3%	5346	23.3%
≥1 vaccination	8967	22.8%	12946	32.9%	8968	5.4%	14074	8.5%	11636	50.7%	17602	76.7%
Comorbidity and health conditions												
Diabetes mellitus	4722	12.0%	7356	18.7%	3385	2.0%	4102	2.5%	52	0.2%	48	0.2%
Diabetes complications	277	0.7%	419	1.1%	108	0.1%	164	0.1%				
Hypertension	13413	34.1%	23340	59.3%	10063	6.1%	14115	8.5%	78	0.3%	112	0.5%
Ischaemic heart disease	2556	6.5%	3468	8.8%	1344	0.8%	1505	0.9%	34	0.1%	47	0.2%
Heart valve disease	492	1.3%	810	2.1%	298	0.2%	322	0.2%	21	0.1%	22	0.1%
Cerebrovascular accident	908	2.3%	1286	3.3%	390	0.2%	397	0.2%	9	0.0%	10	0.0%
Transient ischaemic attack	562	1.4%	1019	2.6%	127	0.1%	165	0.1%	<5	0.0%	<5	0.0%
Heart failure	1334	3.4%	1932	4.9%	260	0.2%	307	0.2%	5	0.0%	9	0.0%
Peripheral arterial disease	511	1.3%	937	2.4%	251	0.2%	304	0.2%	6	0.0%	8	0.0%
Venous thromboembolism	378	1.0%	585	1.5%	378	0.2%	373	0.2%	<5	0.0%	<5	0.0%
COPD	1503	3.8%	2379	6.0%	739	0.4%	914	0.6%	13	0.1%	12	0.1%
Asthma	1538	3.9%	2435	6.2%	11303	6.8%	14391	8.7%	2668	11.6%	3579	15.6%
Acute respiratory distress syndrome	100	0.3%	147	0.4%	392	0.2%	538	0.3%	68	0.3%	113	0.5%
Myo or Peri-carditis	14	0.0%	33	0.1%	114	0.1%	140	0.1%	<5	0.0%	<5	0.0%
Chronic renal disease	1580	4.0%	3236	8.2%	268	0.2%	365	0.2%	<5	0.0%	<5	0.0%
Acute renal failure	3906	9.9%	8844	22.5%	9777	5.9%	13057	7.9%	784	3.4%	976	4.3%
Renal failure (non-specified whether acuter or chronic)	46	0.1%	92	0.2%	158	0.1%	173	0.1%	197	0.9%	217	0.9%

Renal replacement therapy	13	0.0%	15	0.0%	17	0.0%	22	0.0%	0		0	
Chronic liver disease	128	0.3%	132	0.3%	191	0.1%	167	0.1%	<5	0.0%	<5	0.0%
Liver injury	<5	0.0%	<5	0.0%	<5	0.0%	<5	0.0%	0		0	
HIV	55	0.1%	72	0.2%	530	0.3%	407	0.2%	10	0.0%	12	0.1%
Other immunodeficiency (non-HIV)	21	0.1%	14	0.0%	49	0.0%	71	0.0%	22	0.1%	34	0.1%
Cancer	2005	5.1%	3186	8.1%	1235	0.7%	1560	0.9%	59	0.3%	65	0.3%
Autoimmune disease	3251	8.3%	5088	12.9%	9596	5.8%	11827	7.1%	420	1.8%	577	2.5%
Inflammatory Bowel Disease	108	0.3%	224	0.6%	587	0.4%	793	0.5%	11	0.0%	13	0.1%
Transplant (those indicated with immunosuppression therapy)	50	0.1%	76	0.2%	47	0.0%	79	0.0%	<5	0.0%	<5	0.0%
Down syndrome	<5	0.0%	0		39	0.0%	52	0.0%	21	0.1%	13	0.1%
Fragility	480	1.2%	568	1.4%	<5	0.0%	0		0		0	
Parkinson	214	0.5%	278	0.7%	23	0.0%	31	0.0%	0		<5	0.0%
Dementia	599	1.5%	654	1.7%	43	0.0%	29	0.0%	<5	0.0%	5	0.0%
Sepsis	<5	0.0%	7	0.0%	0		<5	0.0%	0		0	
Coagulation disorders	1087	2.8%	1769	4.5%	953	0.6%	1120	0.7%	70	0.3%	119	0.5%
Cystic fibrosis	<5	0.0%	5	0.0%	18	0.0%	32	0.0%	6	0.0%	6	0.0%
Lactating women	0		0		96	0.1%	324	0.2%	0		<5	0.0%
Pregnancy	0		0		275	0.2%	223	0.1%	0		<5	0.0%
Comedication												
Antibiotics	1866	4.7%	4836	12.3%	9021	5.4%	15668	9.5%	947	4.1%	1567	6.8%
Antiviral medications	95	0.2%	234	0.6%	210	0.1%	416	0.3%	10	0.0%	25	0.1%

Corticosteroids	650	1.7%	1670	4.2%	1814	1.1%	3032	1.8%	131	0.6%	234	1.0%
Anti-inflammatory drugs	2576	6.5%	6975	17.7%	12694	7.7%	23799	14.4%	1576	6.9%	2293	10.0%
Opioids	1726	4.4%	4843	12.3%	3025	1.8%	5199	3.1%	43	0.2%	83	0.4%
Other Analgesic drugs	4198	10.7%	11723	29.8%	7721	4.7%	12810	7.7%	860	3.7%	1224	5.3%
Anti-Parkinson drugs	214	0.5%	583	1.5%	148	0.1%	238	0.1%	<5	0.0%	<5	0.0%
Anti-psychotic drugs	831	2.1%	2089	5.3%	2326	1.4%	3411	2.1%	121	0.5%	136	0.6%
Psychoanaleptics and psycholeptics	4136	10.5%	10830	27.5%	8599	5.2%	13267	8.0%	92	0.4%	116	0.5%
Anti-dementia drugs	221	0.6%	677	1.7%	7	0.0%	9	0.0%	0		0	
Statins	4861	12.4%	14512	36.9%	3303	2.0%	7279	4.4%	5	0.0%	10	0.0%
Immunosuppressant medication	139	0.4%	438	1.1%	339	0.2%	626	0.4%	13	0.1%	18	0.1%
Antineoplastic drugs	44	0.1%	106	0.3%	25	0.0%	49	0.0%	<5	0.0%	<5	0.0%
Anticoagulants drugs	3944	10.0%	10393	26.4%	1953	1.2%	3144	1.9%	40	0.2%	54	0.2%
Obesity	3076	7.8%	6145	15.6%	8098	4.9%	12251	7.4%	997	4.3%	1362	5.9%
Primary care visits in 1y prior to full vaccination												
None	21377	54.3%	2971	7.6%	70610	42.6%	34536	20.9%	7349	32.0%	2816	12.3%
≤5 visits	11302	28.7%	16224	41.2%	65216	39.4%	77675	46.9%	10330	45.0%	12187	53.1%
6-12 visits	4471	11.4%	13010	33.1%	21837	13.2%	38680	23.4%	4220	18.4%	6385	27.8%
>12 visits	2196	5.6%	7141	18.1%	7962	4.8%	14734	8.9%	1049	4.6%	1560	6.8%

Table 5. Baseline characteristics of people fully vaccinated with ModeRNA, AstraZeneca or Janssen 's vaccines and matched unvaccinated controls.

	Controls		Vaccinated ModeRNA		Controls		Vaccinated AstraZeneca		Controls		Vaccinated Janssen	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Overall	28817		28817		7684		7684		33148		33148	
Women	12739	44.2%	12739	44.2%	3500	45.5%	3500	45.5%	13854	41.8%	13854	41.8%
Birth year												
1900-1941	1280	4.4%	1218	4.2%	<5	0.0%	0		263	0.8%	158	0.5%
1942-1951	1571	5.5%	1696	5.9%	19	0.2%	21	0.3%	1465	4.4%	1785	5.4%
1952-1955	797	2.8%	825	2.9%	1808	23.5%	1838	23.9%	2164	6.5%	2297	6.9%
1956-1961	999	3.5%	979	3.4%	5844	76.1%	5821	75.8%	3068	9.3%	3106	9.4%
1962-1971	2908	10.1%	2923	10.1%	<5	0.0%	<5	0.0%	9657	29.1%	9637	29.1%
1972-1981	4669	16.2%	4630	16.1%	<5	0.0%	<5	0.0%	14886	44.9%	14745	44.5%
1982-1991	8291	28.8%	8244	28.6%	7	0.1%	0		1381	4.2%	1156	3.5%
1992-2001	6451	22.4%	6446	22.4%	-	-	-	-	252	0.8%	252	0.8%
2002-2019	1851	6.4%	1856	6.4%	-	-	-	-	12	0.0%	12	0.0%
Region												
2	1356	4.7%	1356	4.7%	-	-	-	-	2519	7.6%	2519	7.6%
3	4961	17.2%	4961	17.2%	1557	20.3%	1557	20.3%	8007	24.2%	8007	24.2%
7	10102	35.1%	10102	35.1%	4255	55.4%	4255	55.4%	14115	42.6%	14115	42.6%
14	12398	43.0%	12398	43.0%	1872	24.4%	1872	24.4%	8507	25.7%	8507	25.7%
Time0 (month)												
March	712	2.5%	712	2.5%	-	-	-	-	956	2.9%	956	2.9%
April	1426	4.9%	1426	4.9%	1386	18.0%	1386	18.0%	3283	9.9%	3283	9.9%
May	2876	10.0%	2876	10.0%	5744	74.8%	5744	74.8%	13804	41.6%	13804	41.6%
June	2450	8.5%	2450	8.5%	388	5.0%	388	5.0%	11738	35.4%	11738	35.4%
July	12645	43.9%	12645	43.9%	146	1.9%	146	1.9%	2708	8.2%	2708	8.2%

August	7341	25.5%	7341	25.5%	20	0.3%	20	0.3%	578	1.7%	578	1.7%
September	1367	4.7%	1367	4.7%	-	-	-	-	81	0.2%	81	0.2%
Month of full vaccination												
April	404	1.4%	404	1.4%	-	-	-	-	<5	0.0%	<5	0.0%
May	854	3.0%	854	3.0%	-	-	-	-	1811	5.5%	1811	5.5%
June	2525	8.8%	2525	8.8%	-	-	-	-	11164	33.7%	11164	33.7%
July	2237	7.8%	2237	7.8%	4191	54.5%	4191	54.5%	13561	40.9%	13561	40.9%
August	7961	27.6%	7961	27.6%	2951	38.4%	2951	38.4%	4972	15.0%	4972	15.0%
September	12312	42.7%	12312	42.7%	383	5.0%	383	5.0%	1370	4.1%	1370	4.1%
October	2524	8.8%	2524	8.8%	159	2.1%	159	2.1%	268	0.8%	268	0.8%
Flu vaccination in 5y prior												
No vaccination	27617	95.8%	25512	88.5%	7219	93.9%	6493	84.5%	31715	95.7%	30250	91.3%
<5 vaccinations	1082	3.8%	2388	8.3%	415	5.4%	943	12.3%	1276	3.8%	2414	7.3%
≥5 vaccinations	118	0.4%	917	3.2%	50	0.7%	248	3.2%	157	0.5%	484	1.5%
Other vaccinations in 5y prior												
No other vaccinations	26070	90.5%	24610	85.4%	7307	95.1%	6752	87.9%	30644	92.4%	29400	88.7%
≥1 vaccination	2747	9.5%	4207	14.6%	377	4.9%	932	12.1%	2504	7.6%	3748	11.3%
Comorbidity and health conditions												
Diabetes mellitus	905	3.1%	1248	4.3%	878	11.4%	1070	13.9%	1757	5.3%	2163	6.5%
Diabetes complications	24	0.1%	56	0.2%	33	0.4%	65	0.8%	90	0.3%	105	0.3%
Hypertension	2640	9.2%	4025	14.0%	2277	29.6%	2853	37.1%	4909	14.8%	6503	19.6%
Ischaemic heart disease	409	1.4%	565	2.0%	340	4.4%	356	4.6%	733	2.2%	835	2.5%

Heart valve disease	91	0.3%	98	0.3%	46	0.6%	69	0.9%	141	0.4%	146	0.4%
Cerebrovascular accident	123	0.4%	160	0.6%	105	1.4%	113	1.5%	231	0.7%	253	0.8%
Transient ischaemic attack	81	0.3%	140	0.5%	37	0.5%	44	0.6%	87	0.3%	111	0.3%
Heart failure	186	0.6%	213	0.7%	79	1.0%	81	1.1%	211	0.6%	212	0.6%
Peripheral arterial disease	83	0.3%	149	0.5%	87	1.1%	114	1.5%	159	0.5%	217	0.7%
Venous thromboembolism	90	0.3%	124	0.4%	67	0.9%	46	0.6%	141	0.4%	159	0.5%
COPD	270	0.9%	395	1.4%	211	2.7%	292	3.8%	470	1.4%	600	1.8%
Asthma	2042	7.1%	2762	9.6%	340	4.4%	410	5.3%	1872	5.6%	2082	6.3%
Acute respiratory distress syndrome	40	0.1%	78	0.3%	31	0.4%	29	0.4%	72	0.2%	160	0.5%
Myo or Peri-carditis	18	0.1%	44	0.2%	7	0.1%	8	0.1%	25	0.1%	35	0.1%
Chronic renal disease	184	0.6%	384	1.3%	68	0.9%	98	1.3%	198	0.6%	247	0.7%
Acute renal failure	2021	7.0%	3122	10.8%	601	7.8%	831	10.8%	2558	7.7%	3215	9.7%
Renal failure (non-specified whether acuter or chronic)	31	0.1%	54	0.2%	8	0.1%	5	0.1%	20	0.1%	10	0.0%
Renal replacement therapy	<5	0.0%	7	0.0%	<5	0.0%	<5	0.0%	5	0.0%	<5	0.0%
Chronic liver disease	39	0.1%	57	0.2%	35	0.5%	34	0.4%	100	0.3%	81	0.2%
Liver injury									<5	0.0%	<5	0.0%
HIV	65	0.2%	78	0.3%	29	0.4%	20	0.3%	141	0.4%	118	0.4%
Other immunodeficiency (non-HIV)	17	0.1%	19	0.1%	<5	0.0%	5	0.1%	9	0.0%	7	0.0%
Cancer	415	1.4%	725	2.5%	335	4.4%	355	4.6%	688	2.1%	660	2.0%
Autoimmune disease	1743	6.0%	2212	7.7%	783	10.2%	920	12.0%	2307	7.0%	2947	8.9%

Inflammatory Bowel Disease	90	0.3%	122	0.4%	39	0.5%	41	0.5%	165	0.5%	170	0.5%
Transplant (those indicated with immunosuppression therapy)	11	0.0%	60	0.2%	15	0.2%	0		17	0.1%	5	0.0%
Down syndrome	7	0.0%	<5	0.0%	0		<5	0.0%	5	0.0%	<5	0.0%
Fragility	36	0.1%	40	0.1%	5	0.1%	<5	0.0%	10	0.0%	18	0.1%
Parkinson	28	0.1%	28	0.1%	11	0.1%	17	0.2%	27	0.1%	36	0.1%
Dementia	63	0.2%	71	0.2%	9	0.1%	8	0.1%	45	0.1%	47	0.1%
Sepsis	<5	0.0%	<5	0.0%	<5	0.0%	0		<5	0.0%	0	
Coagulation disorders	254	0.9%	351	1.2%	107	1.4%	116	1.5%	319	1.0%	358	1.1%
Cystic fibrosis	5	0.0%	9	0.0%	<5	0.0%	<5	0.0%	9	0.0%	<5	0.0%
Current lactating women	18	0.1%	50	0.2%	<5	0.0%	0		9	0.0%	9	0.0%
Pregnant in prior 9 months	52	0.2%	43	0.1%					61	0.2%	5	0.0%
Comedication												
Antibiotics	1523	5.3%	2798	9.7%	409	5.3%	732	9.5%	2055	6.2%	3217	9.7%
Antiviral medications	35	0.1%	105	0.4%	21	0.3%	31	0.4%	81	0.2%	110	0.3%
Corticosteroids	330	1.1%	647	2.2%	129	1.7%	208	2.7%	570	1.7%	812	2.4%
Anti-inflammatory drugs	1982	6.9%	4008	13.9%	780	10.2%	1478	19.2%	3466	10.5%	5686	17.2%
Opioids	556	1.9%	1201	4.2%	326	4.2%	577	7.5%	1190	3.6%	1855	5.6%
Other Analgesic drugs	1564	5.4%	3067	10.6%	707	9.2%	1241	16.2%	2506	7.6%	3926	11.8%
Anti-Parkinson drugs	42	0.1%	91	0.3%	19	0.2%	39	0.5%	77	0.2%	122	0.4%
Anti-psychotic drugs	399	1.4%	637	2.2%	190	2.5%	274	3.6%	790	2.4%	1096	3.3%

Psychoanaleptics and psycholeptics	1566	5.4%	2940	10.2%	877	11.4%	1435	18.7%	3127	9.4%	4487	13.5%
Anti-dementia drugs	19	0.1%	72	0.2%	<5	0.0%	12	0.2%	15	0.0%	34	0.1%
Statins	856	3.0%	2236	7.8%	944	12.3%	2065	26.9%	2015	6.1%	3745	11.3%
Immunosuppressant medication	60	0.2%	163	0.6%	35	0.5%	55	0.7%	110	0.3%	172	0.5%
Antineoplastic drugs	7	0.0%	26	0.1%	7	0.1%	10	0.1%	20	0.1%	17	0.1%
Anticoagulants drugs	656	2.3%	1557	5.4%	570	7.4%	895	11.6%	1259	3.8%	1798	5.4%
Obesity	1488	5.2%	2366	8.2%	761	9.9%	1111	14.5%	2344	7.1%	3376	10.2%
Primary care visits in 1y prior to full vaccination												
None	13205	45.8%	5668	19.7%	2694	35.1%	915	11.9%	13261	40.0%	6660	20.1%
≤5 visits	10552	36.6%	13049	45.3%	3279	42.7%	3545	46.1%	12555	37.9%	14653	44.2%
6-12 visits	3640	12.6%	7146	24.8%	1247	16.2%	2258	29.4%	4983	15.0%	8062	24.3%
>12 visits	1420	4.9%	2954	10.3%	464	6.0%	966	12.6%	2349	7.1%	3773	11.4%

Table 6. Effectiveness of complete vaccination against SARS-CoV-2 infection by vaccine brands, time after vaccination and age categories.

	Controls		Vaccinated		HR adj.*	LCI	UCI	VE adj.* (%)	LCI (%)	UCI (%)
	Person-days	Cases	Person-days	Cases						
PF										
Time after 2nd dose:										
≥7d	15300000	606	15400000	599	0.78	0.68	0.88	22	12	32
≥60d	5308014	156	5344028	150	0.68	0.53	0.87	32	13	47
≥90d	2270810	58	2287654	54	0.67	0.45	1.01	33	-1	55
≥120d	750580	14	755853	18	0.78	0.30	1.99	22	-99	70
Age:										
12-17 years	789253	39	792481	6	0.12	0.05	0.29	88	71	95
18-59 years	11000000	458	11100000	464	0.85	0.74	0.98	15	2	26
≥60 years	3467961	108	3509850	129	0.76	0.57	1.02	24	-2	43
MD										
≥14 post 2nd dose	1670979	54	1688332	34	0.43	0.28	0.66	57	34	72
AZ										
≥14 post 2nd dose	605920	26	612280	26	0.82	0.47	1.42	18	-42	53
JA										
≥14 post 1st dose	2655668	122	2658467	327	2.21	1.77	2.76	-121	-176	-77

*The variables remaining in the adjusted model were: For PF: any other vaccination during the prior 5 years (for people aged 12-17 years), hypertension and visits to primary care the year before the complete vaccination (for aged 18-59 years) or antibiotics prescriptions and visits to primary care the year before the complete vaccination (for aged ≥60 years). For MD: opioids prescriptions and visits to primary care the year before the complete vaccination. For AZ: visits to primary care the year before the complete vaccination. For JA: adjusted by obesity, antibiotic prescriptions or prescriptions of anxiolytic, hypnotic, sedative or antidepressant drugs and visits to primary care the year before starting the vaccination.

Table 7. Effectiveness of complete vaccination against hospitalization with COVID-19 by vaccine brands, time after vaccination and age categories.

	Controls		Vaccinated		HR adj.	LCI	UCI	VE adj. (%)	LCI (%)	UCI (%)
	Person-days	Cases	Person-days	Cases						
PF										
Time after 2nd dose:										
≥7d	15400000	109	15500000	41	0.21	0.13	0.32	79	68	87
≥60d	5348418	38	5385463	23	0.38	0.22	0.68	62	32	78
≥90d	2291437	17	2308747	10	0.42	0.13	1.34	58	-34	87
≥120d	755873	5	761389	6	0.40	0.08	2.15	60	-115	92
Age:										
18-59 years	11100000	67	11200000	14	0.08	0.03	0.20	92	80	97
≥60 years	3473731	42	3517447	27	0.36	0.21	0.60	64	40	79
MD										
≥14 post 2nd dose	1679652	9	1696874	<5	0.26	0.08	0.84	74	16	92
AZ										
≥14 post 2nd dose	609604	10	616452	<5	0.08	0.01	0.57	92	43	99
JA										
≥14 post 1st dose	2670642	26	2687229	33	0.96	0.56	1.63	4	-63	44

The variables remaining in the adjusted model were: For PF: COPD, prescription of analgesic drugs before initiating the vaccination, prescription of anticoagulants, and visits to primary care the year before the complete vaccination among people aged ≥60 years, and hypertension, prescription of antibiotics and visits to primary care the year before the complete vaccination among aged 18-59 years. For MD: adjusted by prescriptions of anticoagulants and visits the year before the full vaccination. For AZ: visits. For JA: prescriptions of opioids, anticoagulants, obesity and visits the year before the vaccination.

Table 9. Effectiveness of complete vaccination with Pfizer against death with COVID-19 among patients aged ≥60 years.

	Control		Vaccinated		Main analysis					
	Person-days	Cases	Person-days	Cases	HR adj.*	LCI	UCI	VE adj.* (%)	LCI (%)	UCI (%)
PF										
≥7 days post D2 (aged ≥60 years)	3477690	5	3520588	<5	0.13	0.04	0.43	87	57	96

*HR was adjusted by influenza vaccinations in the 5 years before the time zero.

Table 8. Summary of different vaccine effectiveness estimations (and HR) against hospitalization with COVID-19 calculated in final model (Table 7) and post-hoc analysis.

		Post-hoc analysis	
	Final model (the outcome date was the date of hospital admission)	Using the date of SARS-COV-2 positive test as outcome date (instead of the date of hospital admission) and removing patients with prior positive test	Excluding cases hospitalised by other primary reason (as explained in Annex. Section 2)
Pfizer ≥7days after full vaccination	79% (68-87%)	78% (65-85%)	77% (66-85%)
18-59 y-o	92% (80-97%)	-	91% (81-96%)
≥60 y-o	64% (40-79%)	-	66% (41-81%)
Pfizer ≥60days after	62% (32-78%)	-	64% (33-81%)
Pfizer ≥ 90days after	HR: 0.42 (0.13-1.34) VE: 58% (-34% to 87%)	-	HR: 0.25 (0.06-1.08)
Pfizer ≥ 120days after	HR: 0.40 (0.08-2.15) VE: 60%	-	HR: 0.38 (0.07-2.13)
ModeRNA	74% (16-92%)	74% (16-92%)	93% (47-99%)
AstraZeneca	92% (43-99%)	92% (37-99%)	-
Janssen	HR: 0.96 (0.56-1.63) VE: 4% (-63% to 44%)	HR: 0.90 (0.53-1.55) VE: 10% (-55 to 37%)	HR: 0.69 (0.37-1.28) VE: 31% (-28% to 63%)

Table 3. Hazard ratio of SARS-CoV-2 infection and hospitalisation with COVID-19 immediate after 1st dose COVID-19 vaccination in comparison with controls by vaccine brands and age categories.

	Controls		Vaccinated				
SARS-CoV-2 infection	**Person-days**	**Cases**	**Person-days**	**Cases**	**HR adj.***	**LCI**	**UCI**
0-13 days post PF D1	8923572	1020	8942342.1	1650	1.31	1.21	1.42
Age:							
12-17 years	834469	276	837489	135	0.43	0.35	0.53
18-59 years	6133466.1	683	6141507.1	1366	1.60	1.46	1.76
≥60 years	1828788	59	1836311	133	1.50	1.10	2.05
0-13 days post MD D1	1395769	223	1400014	291	1.21	1.01	1.44
0-13 days post AZ D1	1256284	131	1259474	83	0.53	0.41	0.70
Hospitalisation with COVID-19	**Person-days**	**Cases**	**Person-days**	**Cases**	**HR adj.***	**LCI**	**UCI**
0-13 days post PF D1	8931320	60	8949751	113	1.26	0.91	1.73
0-13 days post MD D1	1397659	15	1401271	23	1.26	0.65	2.45
0-13 days post AZ D1	1257322	15	1259824	13	0.62	0.28	1.35
0-13 days post JA D1	793853.0	<5	794968	16	3.38	1.12	10.26

*HR of the SARS-CoV-2 infection were adjusted by: For PF: prescription of antibiotics, corticosteroids, statins, venous thromboembolism, asthma, cancer and visits to primary care the year before the time zero among people aged ≥60 years, prescription of antibiotics, corticosteroids, prescription of anxiolytic, hypnotic, sedative or antidepressant drugs, asthma, acute respiratory distress syndrome, cancer, inflammatory bowel disease, influenza vaccinations in the 5 years before the time zero and visits to primary care the year before the time zero among aged 18-59 years, and prescription of antibiotics, analgesics, other vaccinations in the 5 years before the time zero and visits to primary care the year before the time zero among aged 12-17 years. For MD: prescriptions of anticoagulants, influenza vaccinations in the 5 years before the time zero and visits to primary care the year before the time zero. For AZ: prescriptions of anti-inflammatories, prescription of anxiolytic, hypnotic, sedative or antidepressant drugs and visits to primary care the year before the time zero. HR of the hospitalisation with COVID-19 were adjusted by: For PF: prescription of antibiotics, corticosteroids, anticoagulants, hypertension, heart failure, COPD, cancer, myo/peri-carditis, influenza vaccinations in the 5 years before the time zero and visits to primary care the year before the time zero. For MD: prescription of anticoagulants, cancer, obesity, influenza vaccinations in the 5 years before the time zero and visits to primary care the year before the time zero. For AZ: prescription of anti-inflammatories, anti-Parkinson, immunosuppressive, prescription of anxiolytic, hypnotic, sedative or antidepressant drugs, heart failure, asthma and obesity. For JA: prescriptions of analgesics, obesity, chronic kidney disease, cerebrovascular accident.

Annex. Additional information

Section 1. Quantification of false negative COVID-19 infections without positive lab test in BIFAP

Aim

The aim of the review was to find evidence to confirm or reject the COVID-19 diagnosis through PC physicians free text comments and explore the reasons for the lack of positive tests.

Methods

A manual review of a random sample of 100 anonymised clinical histories of individuals with COVID-19 diagnosis recorded in primary care but no linked positive COVID-19 test was performed blinded to vaccination status.

The recorded episodes of COVID-19 diagnosis were identified through SNOMED (Systematized Nomenclature of Medicine) codes, as reported in Annex-Table S1. COVID□19 diagnosis codes were introduced in 2020 into the coding schemes used in BIFAP (i.e. the International Classification of Primary Care ICPC-2, the International Classification of Diseases ICD-9 and mapped to Snomed-CT). More detail of the process to identify COVID-19 diagnosis recorded in primary care, and comparison with positive lab-test results (as gold-standard) is reported elsewhere [88].

Annex-Table S1. SNOMED description of COVID-19 diagnosis mapped to available ICPC/ICD-9 codes in primary care clinical histories and frequency of true positives found against test results.

SNOMED description	**SNOMED codes**
Coronavirus infection (disorder)	186747009
Disease caused by severe acute respiratory syndrome coronavirus 2 (disorder)	840539006
Diagnosis of COVID-19 infection confirmed by laboratory testing (disorder)	63681000122103
Pneumonia caused by Human coronavirus (disorder)	713084008
Pneumonia caused by severe acute respiratory syndrome coronavirus 2 (disorder)	882784691000119100130840008
Disease caused by Coronaviridae (disorder)	27619001
Polymerase chain reaction positive for severe acute respiratory syndrome coronavirus 2 (finding)	62531000122108
Asymptomatic severe acute respiratory syndrome coronavirus 2 infection (finding)	189486241000119100
Procedure for action related to case of disease due to SARS-CoV-2 (procedure)	64121000122109

Testing positive for IgG against SARS-CoV-2 (finding)	64671000122103
Outcome: case of COVID-19 still under follow-up (finding)	63511000122107
Positive result of rapid test for detection of IgM and IgG antibodies against SARS-CoV-2 in blood (finding)	63621000122102
Detection of severe acute respiratory syndrome coronavirus 2 (observable entity)	871562009
SARS-CoV-2 antigen testing positive (finding)	64731000122108
Secondary triage for severity level in patient with disease due to SARS-CoV-2 (procedure)	64031000122106
Diagnosis of COVID-19 infection confirmed by laboratory testing (disorder)	63681000122103
Detection of severe acute respiratory syndrome coronavirus 2 antigen (observable entity)	871553007
Positive serologic study for COVID-19 (finding)	62951000122108

Results

During the manual review of the clinical histories of 100 false diagnosis (N= 59 vaccinated and N= 41 unvaccinated), evidence was found to claim the following:

- 56% diagnosis were confirmed (49% among unvaccinated and 61% among vaccinated individuals) through a mention, in free text, to a positive antigen test or PCR (performed in other regions, in nurse home, workplace, private setting or without specification), confirming the diagnosis.
- 25% diagnosis were rejected (27% among unvaccinated and 24% among vaccinated individuals) through a mention, in free text, to a final negative test (N=13), prevalent COVID-19 (N=2), final alternative flu diagnosis (N=2) and referred to a COVID-19 vaccination consultation (N=1).
- 19% diagnosis were not linked with any extra information (being 24% among unvaccinated and 15% among vaccinated individuals).

Based on the confirmed rate and the amount of diagnosis among vaccinated and unvaccinated individuals, we estimated that 1,251 (2,051*61%) and 796 (1,624*49%) COVID-19 cases, respectively, were confirmed diagnosis without a linked positive test. That represented 7% (796/12,059) of estimated total cases among unvaccinated and 11% (1,251/11,690) among vaccinated.

Discussion

For those diagnosis without a linked positive test, the review on the physicians' free text comments suggested that a proportion of true cases (11% among vaccinated and 7%

among unvaccinated patients) were missed in our gold-standard source. Those missed episodes would only be recovered through manual reviewed of all false positive or alternative automatic algorithms based on machine learning and still the precise date of the missed infections may not be clear. Those options may not be affordable in all research projects, so in our point of view, that differential error in the measure of the covid outcome by vaccination status, must be taken into account in the estimation of the covid vaccines effectiveness.

Many of the sampled diagnosis without linked positive test found in the current study (56 out of 100) were based on information from true positive test performed in health centres in regions that did not provide data to BIFAP or were antigen tests at home. That suggested that complete data for people may be stored in different unlinked sources which are crucial to share and merge for reasons of public interest. This is important particularly in a country such as Spain with multiple regional healthcare systems. A scheme supporting the sharing of databases should be considered essential in those situations.

Also, missing positive test may be due to the different clinical protocols developed during the pandemic. In some periods, COVID-19 might have been diagnosed based on symptoms or contact with a positive person, rather than laboratory testing.

Furthermore, some factors, such as nasal steroids use, epistaxis during sampling or the phase of the infection itself, are associated with false-negative results on a COVID-19 test that could also explain the lack of positive tests linked to BIFAP.

In summary, vaccinated had more COVID-19 diagnosis without linked positive test (19.8% versus 15.5% among unvaccinated) and more than half of them (61% versus 49% of unvaccinated) were later confirmed in free text comments. That resulted in higher proportion of missing cases among them in studies using only positive test (11% vs 7% among unvaccinated). We do not know to what extent that is consequence of higher missing free text information among unvaccinated (24%) than vaccinated (15%) individuals, or other reasons.

Conclusion

Thanks to the physicians' free text comments available at event level, we estimated the outcomes missed (without a linked positive test) and the reasons behind. That provided crucial information for those studies assessing the effectiveness of covid vaccines using as

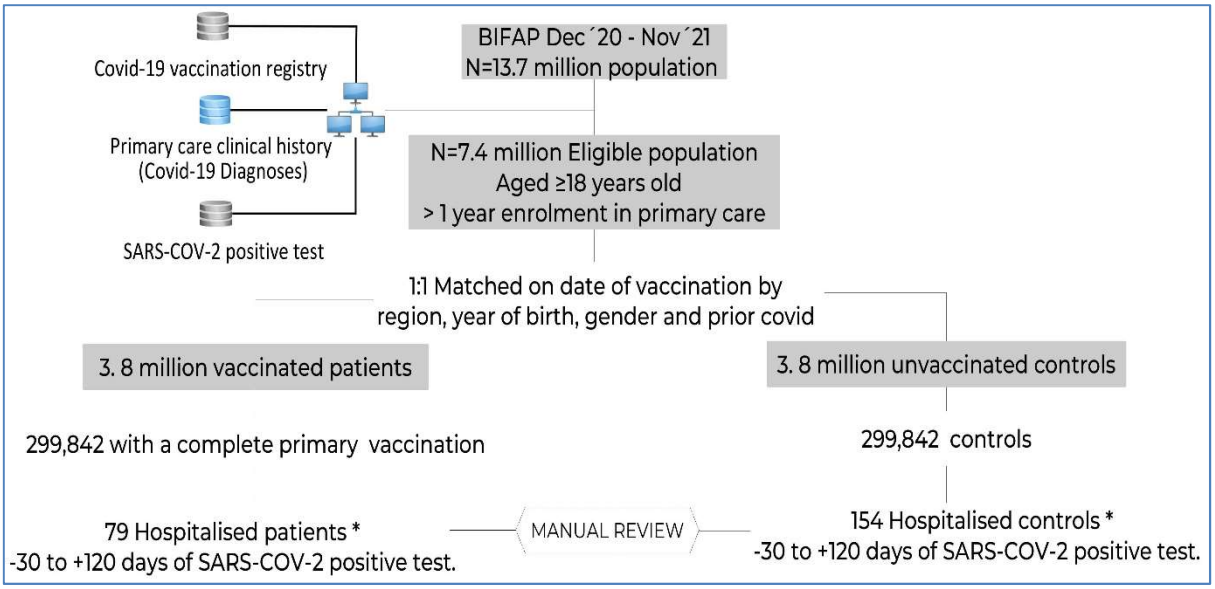
outcome only positive test. In those studies, the differential measurement error of COVID-19 positive test (non-severe COVID-19) among vaccinated and unvaccinated must be taken into account for estimates' correction and their proper interpretation.

Section 2. Validation of Severe COVID-19 information recorded in BIFAP

In this study, the main reason of hospitalization for cases of severe COVID-19 included in main effectiveness analysis and the potential misclassification of the severity of the SARS-CoV-2 infection (i.e. hospitalised for other reasons) by vaccination status and vaccine brand (Janssen-JA, ModeRNA-MD, Pfizer-PZ, AstraZeneca-AZ) was assessed post-hoc. For that we performed the following methods:

- Cases of severe COVID-19 were automatically identified in a cohort of 299,842 pairs of patients completely vaccinated and unvaccinated controls (matched 1:1 on date of vaccination, year of birth, gender and region) aged ≥18 and free of prior SARS-CoV-2 infection, between December 2020- November 2021. **Figure S2.**

Figure S2. COVID-19 severe cases ascertainment in BIFAP



*A recorded diagnosis of SARS-CoV-2 infection, COVID-19 or pneumonia in the hospital registry regardless the actual primary cause of admission.

- **Severe COVID-19** was defined as admissions to hospital or Intensive Care Units (ICU) 'with' a diagnosis of SARS-CoV-2 infection, COVID-19 or pneumonia and a positive SARS-CoV-2 test result from 30 days before to 120 days after.

- Infections might be confirmed through positive PCR, antigens, or any other confirmatory criteria established by clinical protocols.

- A **manual review** of primary care physicians' free text comments **(gold-standard)** included in the clinical histories around the cases records and blinded to vaccination status, was performed to ratify **hospitalizations 'for' COVID-19 as primary reason**.

- Free text comments could include hospital discharge letters, reasons to hospital referrals as well as descriptions of the general practice consultations.

Results of the post-hoc validation

- Overall, 233 cases of severe COVID-19 were automatically identified and reviewed. Table S2.
- Admission 'for' COVID-19 was confirmed in 44% of vaccinated (by vaccine brand: 61% PF; 30% JA; 0% MD; 0% AZ) and 69% of unvaccinated cases (74% PF; 46% JA; 56% MD; 90% AZ).
- Admission for other reasons was proved in 29% of vaccinated (15% PF; 42% JA; 75% MD; 0% AZ) and 11% of unvaccinated (9% PF; 23% JA; 11% MD; 0% AZ). Other reasons were mainly acute cardiovascular episodes.
- Remaining episodes (27% of vaccinated and 19% of unvaccinated) did not have additional information related to admission or stay in hospital or the reason of hospitalization was not clarified.

Table S2. Hospitalisation for COVID-19 or other reasons according to the manual review, by vaccine brand

	Automatically detected cases N= 233			
Case status after manual review:	**Vaccinated N= 79**		**Unvaccinated N= 154**	
	N	**%**	**N**	**%**
Pfizer N=150	**41**		**109**	
Hospitalisation 'for' COVID-19	25	61%	81	74%
Hospitalisation for other reason	6	15%	10	9%
No information	10	24%	18	17%
Janssen N=59	**33**		**26**	
Hospitalisation 'for' COVID-19	10	30%	12	46%
Hospitalisation for other reason	14	42%	6	23%
No information	9	27%	8	31%
ModeRNA N=13	**4**		**9**	
Hospitalisation 'for' COVID-19	0	-	5	56%
Hospitalisation for other reason	3	75%	1	11%
No information	1	25%	3	33%
AstraZeneca N=11	**1**		**10**	
Hospitalisation 'for' COVID-19	0	-	9	90%
Hospitalisation for other reason	0	-	0	-
No information	1	100%	1	10%

We concluded that:

- Overall, the information recorded in the clinical histories suggested misclassification of the COVID-19 severity that was differential among vaccinated and controls.
- Missing information on the reason to hospitalization was, overall, more frequent among vaccinated than controls suggesting differential recording and/or medical assistance.
- Correcting the vaccine effectiveness estimates by excluding admissions for other reasons is recommended to avoid biased results, under the hypothesis of episodes with missing information were true severe COVID-19.
- Although the sample size was low and limited the precision of the predictive values, current validation parameters could be considered to adjust vaccine effectiveness estimates when manual validation cannot be performed.

Section 3. The Validation Of Covid-19 information In The Pharmacoepidemiological Research Database for Public Health System by vaccination status

Preprint publication, 23 Dec 2022
https://doi.org/10.22541/au.167179419.92721401/v1

Abstract

Purpose

To validate Covid-19 information records in The Pharmacoepidemiological Research Database for Public Health System (BIFAP), commonly used for pharmacoepidemiological research in Spain.

Methods

The recorded Covid-19 cases in primary care (PC) or positive test registries (gold-standard) were identified among vaccinated patients against SARS-CoV-2 infection of any age. They were matched with unvaccinated controls by birth year, vaccination date, region, and sex, between December 2020-October 2021. The sensitivity (SE), specificity (SP), positive (PPV), negative (NPV) predictive values, and date accurateness were estimated for PC by vaccination status and age brands.

Results

Among 21,702 patients with positive tests and 20,866 with recorded Covid-19 diagnoses, the SE, SP, PPV, and NPV were, respectively, 79.98%, 99.95%, 80.24% and 99.94% among vaccinated, and 78.67%, 99.96%, 84.51% and 99.94% among controls. For those aged ≥70 years old, SE (71.15-72.85%) was lower while PPV (84.68-88.04%) was higher compared to <70 years old participants. 94.12% of the total true positive cases (N=17,191) were recorded within ±5 days from the date of the test result.

Conclusions

PC Covid-19 diagnosis recorded in BIFAP showed high validation parameters. SE was similar and PPV was slightly lower among vaccinated than unvaccinated controls. According to case definition, correction of vaccines effectiveness estimates by such misclassification is recommended. Data shows the influence of age. Among the elderly, Covid-19 diagnosis was less recorded but when recorded it was more accurate than among younger patients. These findings permit the design of informed algorithms for performing Covid-19-related research.