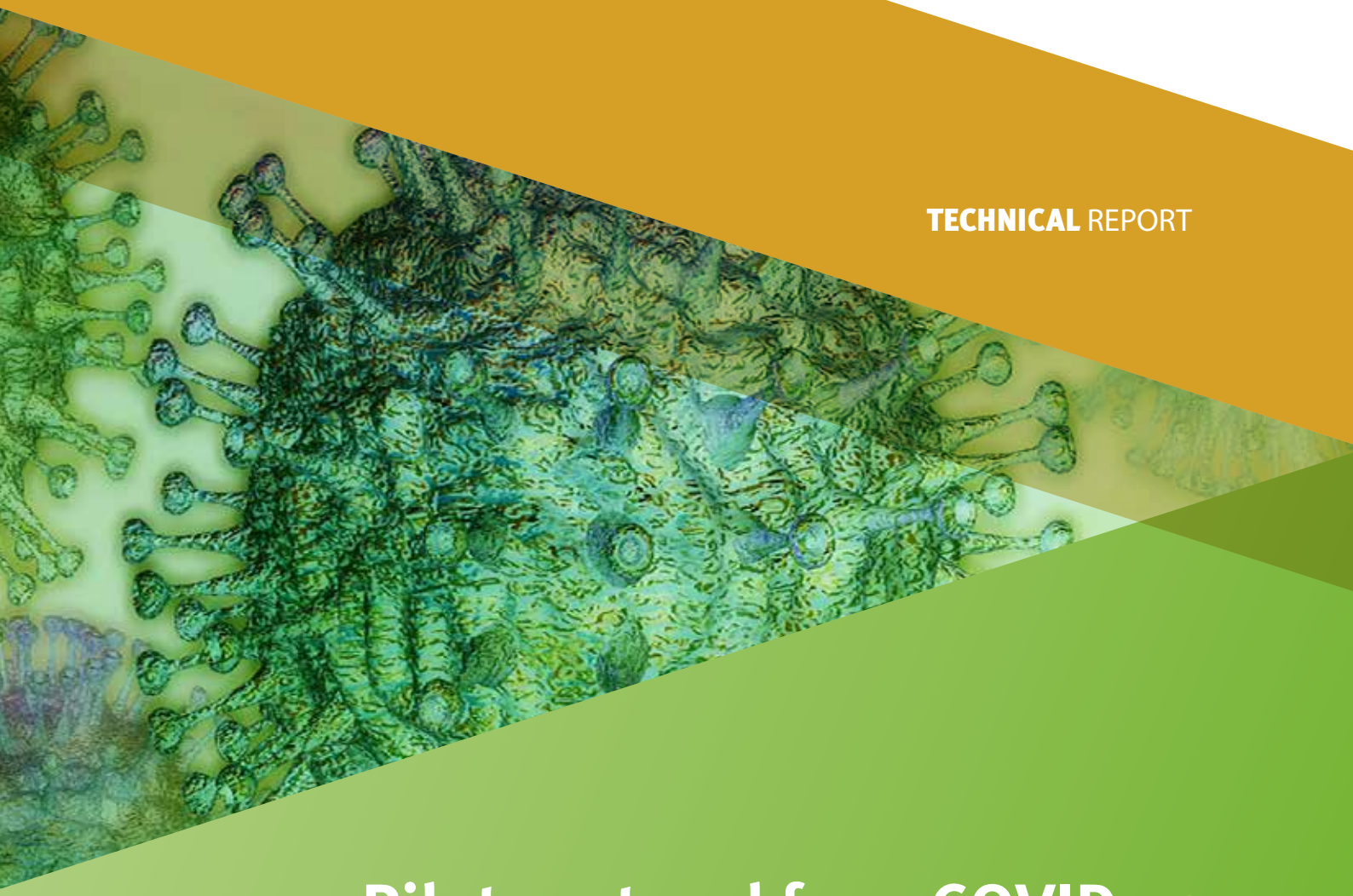
ecdc
EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL



TECHNICAL REPORT

Pilot protocol for a COVID-19 vaccine effectiveness study using health data registries

Version 1.0

www.ecdc.europa.eu

ECDC TECHNICAL REPORT

Pilot protocol for a COVID-19 vaccine effectiveness study using health data registries

Version 1.0

This pilot protocol was commissioned by the European Centre for Disease Prevention and Control (ECDC), as part of the activities within Framework Contract (FWC) ECDC/2021/018 'Vaccine Effectiveness, Burden and Impact Studies (VEBIS) of COVID-19 and Influenza Lot 4'. Production of the protocol was coordinated by Nathalie Nicolay.

Authors (alphabetical order)

Susana Monge Corella (Instituto de Salud Carlos III, Spain), Irina Kislaya (Instituto Nacional de Saúde, Portugal), Baltazar Nunes (Instituto Nacional de Saúde, Portugal), Anthony Nardone (Epiconcept, France), Alexis Sentís (Epiconcept, France), Marta Valenciano (Epiconcept, France)

Acknowledgements

The following experts participated in the first ECDC expert meeting on COVID-19 vaccine effectiveness pilot studies on 18 February 2022 and/or received the protocol for consultation and/or coordinated the implementation of the pilot protocol in their respective countries (or region): Amparo Larrauri (Instituto de Salud Carlos III, Spain), Cristina Buguri, Itziar Casado, Jesús Castilla, Iván Martínez-Baz (Instituto de Salud Pública y Laboral de Navarra), Hanne-Dorthe Emborg, Katrin Finderup (Statens Serum Institut, Denmark), Jostein Starrfelt (Norwegian Institute of Public Health).

The following ECDC staff provided substantial input during the production/revision of the document (alphabetical order): Sabrina Bacci, Nathalie Nicolay.

For questions or request of support of additional documents and tables presented in the protocol, please email vpd.vpd@ecdc.europa.eu

Suggested citation: European Centre for Disease Prevention and Control. Pilot protocol for a COVID-19 vaccine effectiveness study using health data registries. Stockholm: ECDC; 2022.

Stockholm, September 2022.

ISBN 978-92-9498-581-1
doi: 10.2900/421836
Catalogue number TQ-07-22-987-EN-N

Contents

Tables

Abbreviations

COVID-19	Coronavirus disease 2019
EEA	European Economic Area
EMA	European Medicines Agency
GP	General practitioner
HR	Hazard Ratio
RR	Risk Ratio
RT- PCR	Reverse-transcription polymerase chain reaction
SARI	Severe acute respiratory infection
SARS-CoV-2	Severe acute respiratory syndrome – coronavirus 2
VE	Vaccine effectiveness

Executive summary

The end of 2019 saw the emergence of a novel severe acute respiratory syndrome: coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19). Since 31 December 2019 and as of week 28 2022, 155 430 083 cases of COVID-19 (in accordance with the applied case definitions and testing strategies in the affected countries) have been reported in the European Union, including 1 124 471 deaths [1]. As of 22 July 2022, six COVID-19 vaccines – five of which are spike protein-based – had been given conditional marketing authorisation within the European Union/European Economic Area (EU/EEA) by the European Commission, based on the scientific opinion of the European Medicines Agency (EMA): Comirnaty (BNT162b2), Spikevax (mRNA-1273), Vaxzevria (AZD1222), Jcovden (Ad26.COV 2.5), and Nuvaxovid (NVX-CoV2373), and the non-spike protein-based COVID-19 vaccine (inactivated, adjuvanted) Valneva (VLA2001) [2,3].

In 2020, the European Commission emphasised the importance of continuously monitoring the safety and effectiveness of vaccines in the EU/EEA, and called on the European Centre for Disease Prevention and Control (ECDC) and EMA to develop a structured post-authorisation monitoring platform for vaccines, prioritising COVID-19 vaccines [4]. In November 2020, the European Commission proposed to the European Parliament and the Council of the EU a change to the mandates of EMA and ECDC in the context of its COVID-19 'lessons learned' package and the creation of a European Health Union, empowering the two agencies to jointly coordinate independent vaccine monitoring studies [5].

As a result, at the end of 2020, utilising the lessons learned from other vaccine effectiveness (VE) studies (e.g influenza vaccine [6]), ECDC started building infrastructure to perform COVID-19 VE studies. The monitoring of COVID-19 VE over time aims to detect any reduction in vaccine performance that requires further investigation and public health action. The aim of the infrastructure is to build a system to regularly monitor VE and perform studies in different settings and populations and using different routinely available or ad hoc data sources. Different outcomes will be looked at such as severe disease, moderate disease, or transmission and in various and other specific population groups (e.g. healthcare workers).

This pilot protocol for ECDC studies of VE of COVID-19 Vaccines through routinely collected vaccination status and COVID-19 outcome data using health registries, version 1.0, presents the methodology that will be used to produce VE estimates using established health data registries across several study sites/countries (countries) in the EU/EEA. The study design in the current protocol is a retrospective cohort study using data collected routinely in electronic health records databases, selecting data on the resident community-dwelling population (i.e. excluding those living in closed institutions such as nursing homes and prisons) without any documented previous COVID-19 infection who belong to an age group for whom vaccination has been universally recommended at the time of the study. Outcomes include laboratory-confirmed SARS-CoV-2 infection (either symptomatic or asymptomatic) or COVID-19-related hospital admission (either all wards or intensive care units) or death. Other data to be collected include socio-demographic (age, sex), clinical (comorbidities, previous history of SARS-CoV-2 infection) and COVID-19 vaccination variables (brand, number, and dates of doses). The protocol outlines the agreed methods for analysing available data related to COVID-19 and SARS-CoV-2 outcomes both at country level and at European level, including a plan for the pooled analysis.

This pilot protocol is primarily intended to guide the implementation of a first pilot ECDC-funded study using routinely collected data, and includes a concise overview of the systems of the countries included in the pilot study. ECDC encourages the conduct of VE studies, using this protocol as a basis, in countries not currently planning to participate in ECDC-funded studies. The use of consistent protocols will facilitate the comparability of results across study sites/studies.

Background

ECDC COVID-19 vaccine effectiveness studies

In 2018, the Council Recommendation on strengthened cooperation against vaccine-preventable diseases (2018/C 466/01) called on the European Commission to work with the Member States and with the support of the European Medicines Agency (EMA) and in cooperation with ECDC to 'continuously monitor the benefits and risks of vaccines and vaccinations, at EU level, including through post-marketing surveillance studies' [7].

In 2020, the European Commission emphasised the importance of continuously monitoring the safety and effectiveness of vaccines in the EU/EEA and called on ECDC and EMA to develop a structured post-authorisation monitoring platform for vaccines, prioritising COVID-19 vaccines. In November 2020, the European Commission proposed to the European Parliament and the Council of the EU a change to EMA's and ECDC's mandates in the context of its COVID-19 lessons learned package and the creation of a European Health Union, empowering the two agencies to jointly coordinate independent vaccine monitoring studies.

As a result, at the end of 2020, utilising the lessons learned from other VE studies, ECDC started building infrastructure to perform COVID-19 VE studies. The objective of the infrastructure is to build a system to regularly monitor vaccine effectiveness and perform studies in different settings, and depending on the setting, to provide information on different outcomes (severe disease, moderate disease, transmission, etc). The multi-country approach of the effectiveness studies is also one of the key features characterising the studies, with a foreseen approach of progressive inclusion of more countries over time.

Within the infrastructure, one specific project has the aim to assess VE and the impact of COVID-19 vaccines through routinely collected vaccination status and COVID-19 outcome data using health registries. The current protocol describes the pilot methods to implement such studies. The studies will rely on routinely collected vaccination status information and COVID-19 outcome data using established health data registries across several study sites/countries in the EU/EEA. The main objective is to set up a robust network of study sites/countries with the capacity to use health registries under common criteria to monitor COVID-19 vaccines' effectiveness and their impact on a continuous basis. The overall aim will be to provide a real-life monitoring system to detect any changes in VE over time and population subgroups. In future, once the system is established, it may be used to address further specific questions.

The pilot study

The pilot phase aims to assess the feasibility of using health registries under common methods criteria to monitor COVID-19 vaccines' effectiveness and explore the validity of pooled estimates. Four study sites/countries (Denmark, Navarre (Spain), Norway and Portugal) contribute to the pilot phase and provide COVID-19 vaccines estimates over a six-month study period. The project aims to expand to additional countries after the pilot phase.

The main focus of the pilot phase is to document VE against hospitalisation due to COVID-19 in people aged 65 years and above (with results stratified by age group (65-79 and 80+ years). For each site, VE estimates are based on events reported between January to June 2022. As the pilot phase is launched in March 2022, events are collected retrospectively for the months of January to March 2022 and prospectively for April to June 2022.

During the pilot phase, it is expected that estimates produced in June 2022 are given for all additional outcomes of interest as listed in the principal objectives and in all age groups of interest.

In addition to measure COVID-19 VE against hospitalisation in the study population, other specific objectives of the pilot phase are:

- To evaluate the proposed cohort study methods to monitor and detect changes in VE as a signal detection mechanism, since the pilot covers the period immediately before and after the expansion of the Omicron variant in Europe;
- To explore reasons for any differences in VE estimates across countries (e.g. different study populations, different definition of variables, different epidemiological situation);
- To assess the feasibility of the proposed methods regarding database linkage, data quality, VE estimation, data management and analysis;
- To describe the challenges encountered (legal, technical, methodological); and
- To inform the revision of the study protocol and study procedures.

Overall aim

The overall aim of the project based on this protocol is to monitor the real-time performance of COVID-19 vaccines in the community-dwelling resident population in EU/EEA countries in order to detect any signal in terms of reduced VE, so that public health vaccine recommendations may be adjusted accordingly. To achieve this aim, hazard ratio or rate ratio of outcomes of interest will be estimated using information routinely collected in various registries, including vaccination, population and health databases, and using data linkage.

Objectives

Principal objectives

To measure VE of COVID-19 vaccines over time, in community-dwelling resident populations in EU/EEA countries, among people without evidence of a prior SARS-CoV-2 infection against Hospital admission due to COVID-19 with laboratory-confirmed SARS-CoV-2 infection.

VE will be measured for people who have a complete vaccination status with primary series and a complete vaccination status with booster dose compared to those non-vaccinated (see definitions below).

Secondary objectives

To measure vaccine effectiveness against:

- Laboratory-confirmed SARS-CoV-2 infection (including both symptomatic and asymptomatic infections);
- ICU admission due to COVID-19 with laboratory-confirmed SARS-CoV-2 infection; and
- Death due to COVID-19 with laboratory-confirmed SARS-CoV-2 infection.

To measure relative VE by comparing outcomes occurrence in people vaccinated with a complete primary vaccination course and with first booster COVID-19 vaccine dose versus people vaccinated with only a complete primary vaccination course of COVID-19 vaccines (see definition below).

To measure COVID-19 vaccine effectiveness:

- By age group: 5-17, 18-49, 50-64, 65-79, 80+ years;
- Among people with evidence of prior SARS-CoV-2 infection;
- By time since last vaccine dose: number of weeks elapsed between current time and the date of the last vaccine dose received, to evaluate the presence of waning immunity; and
- By vaccine products (primary series with homologous regimens).

The above estimates will be measured once during the piloting in June 2022.

Alternatives

- Different age groups may be used, but it should be always feasible to build the above categories.
- Study sites can contribute to all or just a subset of the established objectives (for example, only some outcomes within the main objectives), or only to some of the subgroup analyses.

→ Study sites/countries to specify the study objectives.

Methodology

Study design

A retrospective cohort study using data collected routinely in electronic health records databases. Comparison of the risk or rate of outcome occurrence between people with different vaccination status.

Study setting

A brief table outlining the characteristics of the study sites/countries contributing to this pilot study is available in Annex 3.

→ Study sites/countries to include more details on the areas covered by the study, population and representativeness. Include brief information on the vaccination rollout, such as prioritised groups and dates in which different populations were incorporated into the vaccination programme.

Study period

VE estimates are produced on a monthly basis between January to June 2022 (Month of data extraction and analysis, Table 1) using events reported between 1 October 2021 and 25 April 2022 (e.g. Study periods, Table 1).

Table 1. Follow-up period corresponding to each monthly estimate, pilot study, 2022

Month of data extraction and analysis	Study period (Period of reported events)
January 2022	1 October to 25 November 2021
February 2022	1 November to 26 December 2021
March 2022	1 December 2021 to 25 January 2022
April 2022	1 January to 25 February 2022
May 2022	1 February to 28 March 2022
June 2022	1 March to 25 April 2022

→ Study sites/countries to specify the study period

Study population

The study population includes people in the National Vaccination Plan and/or the reference population registries fulfilling the following criteria during the different study periods:

- Residents in the EU/EEA country performing the study;
- Belonging to the population with vaccine recommendation in place at time of study period;
- Excluding those living in nursing homes or institutions such as prisons.

Alternative

- If information on specific populations (such as people in nursing homes, prisons, or other closed institutions) is not available, all individuals will be included in the eligible populations. This limitation and its implication should be described in the limitations section.

→ Study sites/countries to specify and describe the study population.

Definition

Vaccination status

The vaccination status will be based on vaccine doses administered up to the date in which vaccination status is assessed (as a time-changing variable), and individuals will be classified as follows:

- Non-vaccinated: has not received any vaccine dose.
- Complete vaccination with primary series of COVID-19 vaccines: people who received the primary series of COVID-19 vaccine doses defined as one dose of Ad26.COV2.S (Janssen) vaccine or two doses of ChAdOx1-S (Oxford/Astra Zeneca), BNT162b2 (Pfizer-BioNTech) or mRNA-1273 (Moderna), or combination of any of the

three vaccines . The two doses should be administered no less than 19 days apart for BNT162b2 or 21 days apart for ChAdOx1-S or mRNA-1273. Complete vaccination status with primary series is achieved 14 days after the date of administration of the final dose required for complete primary series vaccination.

- Complete vaccination with first booster COVID-19 vaccine dose: people who received an additional dose of BNT162b2, mRNA-1273 or ChAdOx1-S at least three months after completion of the primary series. Complete vaccination with first booster status is achieved 14 days after the date of administration of the booster dose.

The 14-day period between the administration of the final dose of primary series or the booster dose will be considered as a separate vaccination status category.

Any person who received at least one dose of vaccine but do not fulfil the definition of complete vaccination with primary series of COVID-19 vaccines will be considered incompletely vaccinated. This category will be analysed, but will not be reported as a category for this study.

People who received a subsequent vaccine dose (any vaccine dose after complete vaccination) that does not fulfil the definition criteria of a booster (delay between dose, unauthorised vaccine product) will be dropped from the risk set on the date they receive the subsequent vaccine dose.

Outcomes

Outcomes of interest are defined as:

- Laboratory-confirmed COVID-19 infection: any SARS-CoV-2 positive test (PCR or rapid antigenic test) regardless of the indication for the test, the presence of symptoms or the severity of infection.
- COVID-19 related hospital admission: laboratory-confirmed infection with admission to hospital 24 hours before or up to 3 weeks after the positive test or symptoms onset, in which admission criteria is compatible with SARI (based on similar criteria as in SARI surveillance) or in which COVID-19 is the main diagnosis in the discharge record (for example, based on ICD coding).
- COVID-19 related ICU admission: hospital admission due to COVID-19 with admission to ICU during hospital stay.
- COVID-19 related death: death for which COVID-19 is recorded as the cause of death OR, if cause of death not available, laboratory-confirmed infection with death in the 30 days after the positive test or symptoms onset.

For each outcome, the censoring date of outcome occurrence will be defined based on date of symptoms onset (if symptomatic), date of laboratory diagnosis, date of hospital admission (if hospitalised), date of ICU admission (if admitted to ICU) and date of death (if deceased).

Alternatives

- Study sites/countries may contribute to some outcome estimates.
- Sites/countries not being able to identify hospitalisations due to COVID-19 (using the proposed or other similar definition) may provide hospitalisations with COVID-19 instead, provided this is well documented in the site-specific protocol and the reporting methods (Annex III).

→ *Study sites/countries to define the outcomes used and the definitions.*

Other variables for subgroup analyses

Previous infection

- No previous infection: No previous positive SARS-CoV-2 test recorded at the date of follow-up period (first day of follow-up for each individual).
- Previous infection: positive SARS-CoV-2 test recorded ≥60 days before the date of beginning of follow-up period.

People within the first 60 days after a positive SARS-CoV-2 test are defined as a COVID-19 episode and should not be included in the risk pool.

The main analysis will be carried out among people without previous evidence of SARS-CoV-2 infection.

→ *Study sites/countries to specify the definitions used for stratification variables.*

Age group

Age will be calculated at the beginning of each study period using the date of birth, and categorised into the following age-groups 5-17, 18-49, 50-64, 65-79, 80+. Alternative age groups may be discussed upon needs.

Time since vaccination

Time since last dose administration (second dose or booster dose for complete vaccination status and complete vaccination status with booster, respectively) will be computed at each point in time by constructing a time-dependent variable.

Assessment of time since vaccination will start at the end of the induction period. This means that, for time since complete vaccination, time 0 will be day 14 after the date of administration of the dose of complete vaccination. For time since the booster, time 0 will be day 14 after the date of administration of the booster dose.

For completely vaccinated people, the time since the administration of the latest dose will be divided into three periods:

- From time 0 (as defined above) up to ≤84 days after time 0 (i.e. <12 weeks);
- Days 85 – 168, both included, after time 0 (i.e. ≥12 weeks & <24 weeks); and
- ≥169 days, after time 0 (i.e. ≥24 weeks).

For completely vaccinated + people who have received booster doses, the time will be divided into three similar periods, although for this pilot study the number of people in the last category will be scarce:

- From time 0 (as defined above) up to ≤84 days after time 0 (i.e. <12 weeks);
- days 85 – 168, both included, after time 0 (i.e. ≥12 weeks & <24 weeks); and
- ≥169 days, after time 0 (i.e. ≥24 weeks).

Vaccine product

Homologous primary vaccination series is defined as:

- one dose of Ad26.COV2.S (Janssen);
- two doses of ChAdOx1-S (Oxford/Astra Zeneca), administrated with a minimum of 21 days between the first and second dose;
- two doses of an mRNA vaccine (either mRNA-1273 (Moderna) or BNT162b2 (Pfizer-BioNTech)), administrated with minimum of respectively 19 days or 21 days between first and second dose.

→ *Study sites/countries to specify the definitions used for stratification variables.*

Effect modifiers or confounding variables

Sociodemographic

- Sex;
- Age group;
- Individual level socioeconomic status (normally not available): Educational level, occupation, unemployment, income, as available in registries;
- Area level socioeconomic condition (postal code, municipality or other): income per capita, GDI per capita, inequality or deprivation index [8], unemployment rate, as available in other complementary data sources;
- Others.

Comorbidities and health-seeking behaviours

Different variables can be used to account for comorbidities:

- Number of chronic comorbidities (0, 1, ≥2), see Annex 1.
- Comorbidities index (e.g. Charlson index)
- Use of medication for chronic medical condition (yes, no), see Annex 1.
- Number of consultations in primary care over the last 12 months, or other relevant time-frame (0, 1, 2, ≥3)
- Number of SARS-CoV-2 tests performed in the last year, or other relevant time-frame.
- Hospitalisation in the previous year, or other relevant time-frame (yes, no)
- Vaccination with other 'respiratory' vaccines: seasonal influenza or pneumococcus (yes, no)

→ *Study sites/countries to specify the definitions used for effect modifiers/confounding variables.*

Data sources

- The study will be using routinely collected data from various population health registries available at national or subnational level. Each database should contain a unique identifier for every individual to allow data linkage between databases.

Sources of information on the reference population

- The reference population database (for example, census database, health coverage database, etc.) with individual records of the target population. It should contain variables that allow the identification of non-residents, or temporary residents.

- In any of databases, it is desirable that it is possible to identify specific populations, such as those living in nursing homes or other institutions.

Sources of information on the vaccination status

Source of information on the vaccination status: vaccination registry or vaccination record databases with records of people and dates of COVID-19 vaccination, including vaccine product.

Alternative

- For relative effectiveness (i.e. direct comparison between people in different vaccination groups), the individualised data source for the reference population may not be needed, because comparison can be established between vaccinated individuals alone.

→ Study sites/countries to specify and describe data sources and database.

Sources of information on the outcomes

Data will be extracted from different electronic health record databases:

- COVID-19 laboratory-confirmed infection;
- Epidemiological surveillance databases (for notifiable diseases);
- Primary healthcare consultation;
- Hospital admission/discharge; and
- Death or mortality registries with record of cause of death.

→ Study sites/countries to specify and describe sources of information for each outcome.

Sources of information on effect modifiers or confounders

- Electronic databases informing on comorbidities: possibly primary healthcare records, data on medication prescriptions, or any other population-based data source that can provide information on comorbidities for all cohort people (and not only for cases).
- Electronic databases informing on healthcare-seeking behaviour: possibly healthcare administrative database (i.e. number of consultations), laboratory records (i.e. number of test performed) or non-COVID-19 vaccination records (i.e. other vaccines administered).

Alternative

- Stratifying for effect modifiers or adjusting for the confounders will add validity to the estimations, but can be included or not as available.

→ Study sites/countries to specify and describe sources of information for each effect modifier or confounding factor.

Construction of the cohort

Vaccination group and characteristics at baseline

The reference population database will be linked with the electronic databases on vaccination, comorbidity, and/or health-seeking behaviours registries using the unique identifier and a deterministic data linkage procedure (no random component in the linkage procedure).

People will enter the study in their corresponding group of vaccination status based on the data available in the vaccination registry.

→ Study sites/countries to specify the data linkage method used.

Time-changing characteristics

The vaccination status will be assessed, and people classified into the same or updated vaccination status on a daily (preferable) or weekly basis, generating a new record in the dataset for each new assessment. The variable with the time since last vaccine dose will be updated accordingly (Table 2). Effect modifiers and confounding variables can be included as time-changing if possible, and classification status reassessed on a daily or a weekly basis as appropriate.

Table 2. Example of update of the time dependant vaccination variable

Individual ID	Study Week	Reference date for the week	Vaccination status	Time since last dose	Other variables classified at baseline (e.g., age, sex, or comorbidities)
123456	1	1 Dec 2021	Complete vaccination	≥25 weeks	Constant
123456	2	8 Dec 2021	Complete vaccination	≥25 weeks	Constant
123456	3	15 Dec 2021	-	<12 weeks	Constant
123456	4	22 Dec 2021	Booster dose	<12 weeks	Constant

Alternative

- Study sites may choose to use previous infection as a time-changing variable and reassess on a daily or weekly basis, as appropriate, resulting in a more complex data set. Of note, because the 60 days following a COVID-19 event are considered to be part of the same episode, no person experiencing an outcome during the study period can re-enter the study as previously infected in the two-month timeframe of the study. However, previously infected individuals may finish the 60 days episode during the eight weeks of the study timeframe. In this last case, individuals would not be in the data set since time 0, but have a delayed entry into the previous infection group at time t, at the calendar date when they fulfil 60 days since the previous COVID-19 episode.

→ Study sites/countries to list time-changing variables.

Identification of outcomes during follow-up

Information for identification of outcomes and their occurrence dates will be obtained by individual deterministic linkage between the cohort built previously and the databases containing the information on the outcomes.

Outcome classification for each person will be assessed on a daily or weekly basis from the study start date (t0) and up to the administrative censoring date (Table 3). The date of outcome occurrence will be defined as the day before the date indicated in the database.

Table 3. Example of dataset with information on the outcome

Individual ID	Study Week	Reference date	Vaccination status	Time since last dose	Other variables classified at baseline	Outcome: infection	Outcome: hospitalisation
123456	1	1 Dec 2021	Complete vaccination	≥25 weeks	Constant	0	0
123456	2	8 Dec 2021	Complete vaccination	≥25 weeks	Constant	0	0
123456	3	15 Dec 2021	-	<12 weeks	Constant	0	0
123456	4	22 Dec 2021	Booster dose	<12 weeks	Constant	1	0

Analysis plan

Description of the sample selection

The total number of people in each of databases fulfilling the inclusion criteria (see xx). Proportion of them (%) that have a successfully cross-matched between the different databases. Total numbers (%) of people excluded. More details in the dummy table below (Table 4).

Table 4. Dummy table: description of sample selection by health registry data source (one per country/site)

	Data source 1 excluded	Data source 1 included	Data source 2 excluded	Data source 2 included	Data source 3 excluded	Data source 3 included	Data source 4 excluded	Data source 4 included
Name of the source								
Percentage of positive cross-matched/reference population database								
Date of extraction								
Last date of available data								
Number of observations								
Number of people								

**% of positive cross-matched according to the reference data source (i.e. the administrative population).*

Description of the study population

For each group (i.e. in each category of the vaccination status variable and within each stratified analysis), study/sites and overall population study, we will describe: the number of people, total person-time, and number of events. Using percentages or medians and interquartile ranges as appropriate, cohort groups will be described by age, sex, and the rest of stratification or confounding variables. The proportion of SARS-Cov2-positive people with viruses sequenced will be described to study the feasibility of stratifying VE estimates by variants of concern (pilot sites/countries have already confirmed that a low proportion of viruses were sequenced).

The study-site background information can include the distribution of vaccination coverage or status with time and the cases, hospitalisations, and deaths with time. More details can be found in Table 5 below.

Table 5. Socio-demographic, microbiological, and clinical history of study population included by vaccination status (one per country/site)

Characteristics	Total number of people	Total number of people included		Not vaccinated among included		Complete vaccination with primary series of COVID-19 vaccines among included		Complete vaccination with first COVID-19 vaccine booster dose among included	
	n	n	%	n	%	n	%	n	%
Number of people									
Total people in study population									
Linked									
Not linked									
Sequenced people									
Delta									
Omicron									
Others									
Unknown									
Sex									
Male									
Female									
Missing									
Age group									
5 to 17 years									
18 to 49 years									

Characteristics	Total number of people	Total number of people included		Not vaccinated among included		Complete vaccination with primary series of COVID-19 vaccines among included		Complete vaccination with first COVID-19 vaccine booster dose among included	
50 to 64 years									
65 to 79 years									
80+ years									
Missing									
Special populations									
Healthcare workers									
Prisons									
Long-term care facilities									
Missing									
Comorbiditites									
None									
1 or 2									
2+									
Missing									
Vaccine brand									
Comirnaty									
Spikevax									
Jcovden									
Vaxzevria									
Nuvaxovid									
Missing									
Unknown									
Outcome									
Symptomatic									
Missing									
Hospitalisation									
Missing									
Death									
Missing									

** The proportion of the missing data will be used to determine if each specific variable can be included in the model and how (e.g., missing could eventually be included in the model as a category). It is not planned to improve quality by filling the missing data, although it will be encouraged when possible.*
Information on the number of people, total person-time of follow up and number of events by vaccination status, age, previous infection and vaccine brand, for the different events under study will be collected in the format for periodic reporting (Annex 2).

Estimation of the vaccine effectiveness at site/country level

Crude vaccine effectiveness

VE will be estimated using hazard ratio (HR) or rate ratio (RR) of defined outcome(s) in people for the different vaccination statuses as defined above (complete vaccination with primary series of COVID-19 vaccines or complete vaccination with first booster COVID-19 vaccine dose) versus unvaccinated in the study population [9-11].

VE will be estimated using hazard ratio (HR) or rate ratio (RR) of defined outcome(s) in people completely vaccinated with primary series of COVID-19 vaccines d versus unvaccinated in the study population.

Crude VE = (1-HR) x 100

Using as the time scale the calendar date or week (as appropriate), a stratified Cox regression using as stratum each study week will be modelled. This will perform an independent Cox regression in each calendar week, thus making it more probable that the proportionality of hazards assumption is fulfilled. Weekly estimates are then combined into a single estimate for the full time period.

Each person will enter the study in the different vaccination status groups on the date they are first classified into that group. This will be date of the beginning of the study, except for people that change vaccination status groups throughout follow-up, which will be censored without event in the group that they leave and are recorded as a delayed entry in the group where they are newly classified. In the example of Table 3, the individual with ID 123456 enters the completely vaccinated group at t0 (1 December 2021) is censored without event at t14 (14 December 2021) and enters the risk group of booster dose at t15 (15 December 2021) (and thus will not contribute to estimates of effectiveness in the booster group between days 0 and 14). The time of start in each group will be provided in the survival command.

Date of end of follow-up will be the earliest among date of death, date of outcome, date of dropout in the study for any other reason or administrative censoring.

Status by the end of follow-up will be event (1) or free of event (0), which will be the dependent variable for the Cox regression.

Alternative

- Study sites may choose to perform the analysis using incidence rates (and Poisson Regression) instead of survival analysis (and Cox regression). Results should be very similar and that will be a valid option as long as the time-changing nature of vaccination status and other variables is accounted for.

→ Study sites/countries to define the analytical approach.

Adjusted vaccine effectiveness

The regression to estimate HR or RR will be adjusted by fixed or time changing confounders, as appropriate, and as previously defined. Firstly, partially adjusted HR or RR will be estimated, adjusting by age group, sex, and region in the country if appropriate. Secondly, a fully adjusted HR or RR estimate will be produced adjusting by variables related to socioeconomic condition, comorbidities, health-seeking behaviour, and previous infection (when not used as a stratification variable).

Examples of R code for performing the stratified Cox regression in this type of dataset will be provided in a detailed accompanying Statistical Analysis Plan.

Relative vaccine effectiveness of booster dose

Relative VE will be estimated using hazard ratio (HR) or rate ratio (RR) of defined outcome(s) in people completely vaccinated with COVID-19 vaccine booster dose versus those completely vaccinated with primary series of COVID-19 vaccine doses within the population study.

Relative crude and adjusted relative VE will be calculated.

Stratified vaccine effectiveness

Stratified crude and adjusted will be calculated for each stratification variable.

Methods for pooling estimates (multi sites/countries analysis)

Country-specific HRs/RRs and standard errors for the effect of COVID-19 vaccination obtained from the study studies will be combined in a model that incorporates random effects of the studies, to account for unmeasured country-specific factors [12].

The country-specific vaccination status-disease effects (HRs/RRs) will be weighted by the inverse of their marginal variances (generic inverse variance method). The marginal variance is the sum of the individual study-specific

variances and the variance of the random study effects (τ^2). This will give the pooled HR/RRs and a standard error. The confidence interval around the pooled effect (the range of values that contain the true average HR/RR with 95% certainty) will then be calculated. τ^2 and I^2 will be used to describe between-studies heterogeneity, along with the p-value of the heterogeneity test [13]. Potential factors or specific pilot sites/countries characteristics that could be the source of qualitative heterogeneity will be described. Prediction intervals (the limits that will contain estimates of 95% of studies with similar intra and between studies heterogeneity) will also be computed.

The country-specific HR/RRs and their confidence intervals, along with the pooled HR/RRs, will be presented graphically in a forest plot. This model will also be compared against a twostage analysis with fixed study effects to assess the effects of model assumptions.

The crude effect, the basic adjusted effect (age, sex, region), and the fully adjusted effect (adding the rest of available covariates) will be compared to assess the degree of confounding by different factors and guide the hypothesis around the explanation of differences across sites/countries.

Pooled estimates will be obtained overall for each of the study outcomes and vaccination status categories, and for all the subgroup analyses outlined in this protocol. For each pooled estimate, only sites/countries contributing to that estimate will be used.

Sensitivity analyses will describe pooled estimates obtained while excluding some study participants for whom variables were collected differently, who have differences in the study setting that could affect the estimates (e.g. different SARS-CoV-2 genetic variant with predominant circulation), or for whom estimates significantly differ across sites/countries (i.e. site confidence intervals do not overlap with the pooled estimate confidence interval), particularly if the I2 estimate is >50%.

Data-checking and validation

The following data-checking and data validation will be undertaken before analysis:

- Identification of inconsistencies (e.g unusual values and outliers);
- Inclusion/exclusion criteria adherence;
- Missing values, missing clinical details, missing laboratory results;
- Duplicate cases and multiple admissions;
- Too long a delay from the date of onset of symptoms to laboratory specimen collection date;
- Consistency of and among dates (onset, admission, discharge, swabbing); and
- Missing data for essential variables that can lead to exclusion of the records from the analyses.

→ *Study sites/countries to list data checking and validation items*

Ethical requirements

Approval by an ethics committee is a requirement.

All sites must conform with national and EU ethical and data protection requirements. Aggregated VE estimates from the different sites used in the study are handled according to the Helsinki Declaration revised by the World Medical Organization in Fortaleza in 2013.

→ *Study sites/countries to provide information on ethical approval.*

Potential biases and limitations

- The availability and degree of detail in the study variables will be limited by what is recorded in registries whose final aim was different than the performing of this study. Moreover, a high degree of variability between countries is expected, making the pooled results more difficult to interpret as a homogeneous estimation.
- Control for confounding will also be limited, since many relevant variables are not monitored or are not available in the data. Previous infections as a relevant cofounder may not be properly recorded, and therefore may not be controlled for due to limitations in the availability of data and different testing policies among participating countries.
- Given the staggered rollout of COVID-19 vaccination in most European countries, there is collinearity between age, moment of infection, and type of vaccine used. Therefore, the interpretation of effects by since time from vaccination and by age needs to be carried out with caution from a causal perspective.
- The non-vaccinated people will be the reference group in the analysis. However, they are increasingly different from the vaccinated population. For example, they might take fewer preventive precautions and be more at risk and have an underlying higher risk of COVID-19, overestimating VE. On the other hand, they

may be more likely to be people with previous infection that we failed to identify in our registries (something likely after the widespread use of self-tests), and so we would underestimate VE.
- Many countries have stopped exhaustively registering COVID-19 diagnoses in the Omicron wave, and many have ended the systematic testing of all mild COVID-19 cases. This can make estimates against infection and/or symptomatic infection unreliable.

→ Study sites/countries to describe the main limitations of their study and how those can affect the results.

Data-sharing for the purpose of pooling results at a European level

The specific information to be shared as part of the ECDC pilot study should be included in the Excel files accompanying this protocol. An example table is available as Annex 2, and includes:

- Point estimation of every VE estimated in the study (for each outcome, in every subgroup), i.e. exponential coefficient from the Cox or Poisson regression.
- Lower and upper limits of the confidence interval of every VE estimated in the study (for each outcome, in every subgroup), i.e. exponential lower and upper limits of the confidence intervals estimated from the Cox or Poisson regression.
- Number of people, of person-days of follow-up and of events of every VE estimated in the study (for each outcome, in every subgroup).

For the longitudinal monitoring proof of concept, sites/countries will carry out the study monthly in April, May, and June 2022, providing data to the European pool no later than the 20th of each month.

References

1. European Centre for Disease Prevention and Control. COVID-19 Situation update week 38, updated 24 September 2021. ECDC: Stockholm; 2021. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/Communicable-disease-threats-report-24-september-2021.pdf
2. European Medicines Agency. COVID-19 vaccines: authorised. EMA: Amsterdam; 2021. Available at: https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorised
3. European Centre for Disease Prevention and Control (ECDC). Overview of the implementation of COVID-19 vaccination strategies and deployment plans in the EU/EEA. Stockholm: ECDC; 21 Apr 2022. Available at: https://www.ecdc.europa.eu/en/publications-data/overview-implementation-covid-19-vaccination-strategies-and-deployment-plans
4. European Commission. Proposal for a regulation of the European parliament and of the Council amending Regulation (EC) No 851/2004 establishing a European Centre for disease prevention and control. November 2020. Available at: https://ec.europa.eu/info/sites/default/files/proposal-mandate-european-centre-disease-prevention-control_en.pdf
5. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Building a European Health Union: Reinforcing the EU's resilience for cross-border health threats. Brussels. 11 Nov 2020. Available at: https://ec.europa.eu/info/sites/default/files/communication-european-health-union-resilience_en.pdf
6. Valenciano M, Ciancio BC, Moren A, the influenza vaccine effectiveness working group. First steps in the design of a system to monitor vaccine effectiveness during seasonal and pandemic influenza in EU/EEA Member States. Euro Surveill. 2008;13(43):pii=19015. Available at: https://www.eurosurveillance.org/images/dynamic/ee/v13n43/art19015.pdf
7. Council Recommendation on strengthened cooperation against vaccine-preventable diseases. Brussels. 20 November 2018. Available at: https://data.consilium.europa.eu/doc/document/ST-14152-2018-REV-1/en/pdf
8. Ribeiro AI, Launay L, Guillaume E, Launoy G, Barros H. The Portuguese version of the European Deprivation Index: Development and association with all-cause mortality; 2018. PLoS ONE 13(12): e0208320. Available at: https://doi.org/10.1371/journal.pone.0208320
9. Clinical Trials (Second Edition). Study Design, Endpoints and Biomarkers, Drug Safety, and FDA and ICH Guidelines 2016, Chapter 9 – Biostatistics – Part I, Pages 203-226. Available at: https://doi.org/10.1016/B978-0-12-804217-5.00009-6
10. Nunes B, Rodrigues AP, Kislaya I, Cruz C, Peralta-Santos A, Lima J, et al. mRNA vaccine effectiveness against COVID-19-related hospitalisations and deaths in older adults: a cohort study based on data linkage of national health registries in Portugal, February to August 2021. Euro Surveill. 2021 Sep;26(38):2100833. Available at: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.38.2100833
11. Monge S, Rojas-Benedicto A, Olmedo C, Mazagatos C, José Sierra M, Limia A, et al. Effectiveness of mRNA vaccine boosters against infection with the SARS-CoV-2 omicron (B.1.1.529) variant in Spain: a nationwide cohort study. Lancet Infect Dis. 2022 Sep;22(9):1313-20.
12. Chapter 10: Analysing data and undertaking meta-analyses. Cochrane Training. Available at: https://training.cochrane.org/handbook/archive/v6.2/chapter-10#section-10-3
13. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006 Jun;11(2):193–206.

Annex 1. Data dictionary (example)

Individual characteristics

	Variable	Type	Coding	Definition
Operational	Extract date	Date	dd/mm/yyyy	Database extraction date
Patient characteristics	Sex	Numeric	0 = female	Sex of patient
			1 = male	
			3 = other	
			8 = do not know	
	dob	Date	dd/mm/yyyy	Date of birth (only if no age; once age calculated from dob this will be dropped)
	Age groups	Numeric (categorical)	1=5-17	
			2=18-49	
			3=50-64	
			4=65-79	
			5=80+	
	postcode	Numeric		Postcode of residence (four digits)
	residence	Numeric	0 = at home, not dependent on home support/care	Patient residence at time of event onset. Whether patient was living at home or was institutionalised, or had pre-hospital dependence on home support/care
			1 = at home, but dependent on home support/care	
			2 = institutionalised	
			3 = Do not know	
	hcw	Numeric (categorical)	0 = No	Whether the patient is a healthcare worker
			1 = Yes	
			8 = do not know	

Outcome

	Variable	Type	Coding	Definition
COVID case	swabdate	Date	dd/mm/yyyy	Respiratory specimen collection date
	lab_covtest	Numeric (categorical)	0 = No	Tested for SARS-CoV-2
			1 = Yes	
			8 = Do not know	
	lab_covtesttype	Numeric (categorical)	1 = RT-PCR	Type of laboratory test used
			2 = Serology	
			3 = Rapid test	
			4 = Other	
			8 = Do not know	
	lab_covtesttype_sp	Text		Specify other type of laboratory test
	lab_covid	Numeric (categorical)	0 = Negative	Laboratory result: virus type SARS-CoV-2
			1 = Positive	
			8 = Do not know	
			1 = Positive	
			8 = Do not know	
Hospital/ward information	prevhosp	Numeric (categorical)	0 = No	Prior admission to hospital (at least once in previous 12 months)
			1 = Yes	
			8 = Do not know	
	admitdate	Date	dd/mm/yyyy	Date of hospital admission
	hospitalward	Text		Ward
	dischargedate	Date	dd/mm/yyyy	Date of hospital discharge
COVID_icu	icu		0 = No	Admission to intensive care unit (ICU)
			1 = Yes	
			8 = Do not know	
	icuadmitdate	Date	dd/mm/yyyy	Date first admitted to ICU
	icudisdate	Date	dd/mm/yyyy	Date last discharged from ICU
Death	death	Numeric (categorical)	0 = No	Person is deceased
			1 = Yes	

	Variable	Type	Coding	Definition
			8 = Do not know	
	deathdate	Date	dd/mm/yyyy	Date of death
	deathcause	Numeric (categorical)	1 = died from COVID-19 2 = died other cause 8 = died unknown cause	Cause of death

Vaccination status

	Variable	Type	Coding	Definition
COVID-19 vaccination	panvaccany	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received at least one dose of COVID-19 vaccine
	Panvaccdate_i	Date	dd/mm/yyyy	Vaccination date (for each dose, i)
	Panvacctype_i	Text		Type of vaccine (for each dose, i)
	panvaccdose	Numeric	0, 1, 2	Number of doses received

Signs and symptoms

	Variable	Type	Coding	Definition
Signs/symptoms at admission)	Feverishness	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sub-febrility (37–38°C)
	Fever	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	History of fever (to construct SARI case definition)
	Malaise	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Malaise (to construct SARI case definition)
	Headache	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Headache (to construct SARI case definition)
	Myalgia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Myalgia (to construct SARI case definition)
	Sob	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Shortness of breath (to construct SARI case definition)
	general_deter	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Deterioration of general condition (asthenia or loss of weight or anorexia or confusion or dizziness) (to construct SARI case definition)
	Vomit	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Vomiting
	Diarr	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diarrhoea
	abdopain	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Abdominal pain
	Ageusia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Loss of sense of taste
	anosmia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Loss of sense of smell
	onsetdate	Date	dd/mm/yyyy	Date of onset of symptoms

Confounding variables 1

	Variable	Type	Coding	Definition
Underlying chronic conditions (ICD and ICPC-2 codes)	Anaemia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Anaemia/chronic haematologic disease
	Asplenia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Asplenia (absence of/damage to spleen)
	Asthma	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Asthma
	Cancer	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cancer (any)
	Hypert	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Hypertension
	Demente	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Dementia
	Diabetes	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diabetes
	Heartdis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Heart/cardiac disease (excluding hypertension)
	Immuno	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	HIV (including other immunodeficiency, organ transplantation)
	Liverdis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chronic liver disease (excluding cancer)
	Lungdis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Lung disease (excluding asthma)
	Neuromusc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Neuromuscular disorder
	Height	Numeric (integer)		Height of patient in metres
	Weight	Numeric (integer)		Weight of patient in kg
	Bmi	Numeric (1 d.p)		BMI of patient (calculated afterwards)
	Obese	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Obesity (only if height, weight, and BMI not collected; can be calculated)
	Rendis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Renal disease (excluding cancer and acute renal failure)
	Rheumat	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rheumatologic disease
	Stroke	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Stroke
	Tuberc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Tuberculosis

Confounding variables 2

	Variable	Type	Coding	Definition
Patient characteristics	Frailty	Numeric (categorical)	To be updated with coding depending on score used	Clinical frailty score at admission (where possible) or Barthel Index
Pre-symptomatic treatment/intervention: medication	statin_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on statins two weeks before symptoms onset
	chemo_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chemotherapy (within six months or currently) for cancer
	gliclaz_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Gliclazides (for diabetes or heart failure) two weeks before symptoms onset
	psychotrop_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Psychotropic drugs (including benzodiazepine, etc.) two weeks before symptoms onset
	ECA_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on angiotensin-converting enzyme inhibitors two weeks before symptoms onset
	ARB_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on angiotensin II receptor blockers two weeks before symptoms onset
	Metformin_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on metformin two weeks before symptoms onset
	APD_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on antiplatelet drugs two weeks before symptoms onset
	AINES_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient on non-steroidal anti-inflammatory drugs two weeks before symptoms onset
	Antiviral	Numeric (categorical	0 = No 1 = Yes 8 = Do not know	Patient on antiviral drugs two weeks before symptoms onset
Pre-symptomatic treatment/intervention: vaccination	flu_vacc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received current seasonal influenza vaccination
	flu_vaccdate	Date	dd/mm/yyyy	Date of last influenza vaccination
	ppv_vacc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received PPV23 vaccination
	ppv_vaccdate	Date	dd/mm/yyyy	Date of last PPV23 vaccination
	pcv_vacc	Numeric (categorical)	0 = No 1 = Yes	Received PCV7/10 or 13 vaccination

	Variable	Type	Coding	Definition
			8 = Do not know	
	pcv_vaccdate	Date	dd/mm/yyyy	Date of last PCV7/10 or 13 vaccination
	bcg_vacc	Numeric (categorical)	0 = No	Received BCG vaccination
			1 = Yes	
			8 = Do not know	
	bcg_vaccyear	Numeric	yyyy	Year of BCG vaccination

Annex 2. Format for periodic reporting

(Example, the full periodic reporting tables are available upon request to ECDC)

Vaccination status categories	N*	person-days	Events	HR crude			HR adjusted1**			HR adjusted2***		
Non-vaccinated				REF	REF	REF	REF	REF	REF	REF	REF	REF
Complete vaccination												
Complete vaccination + booster												
Non-vaccinated				REF	REF	REF	REF	REF	REF	REF	REF	REF
Complete vaccination + booster <12 weeks												
Complete vaccination + booster ≥12 & <24 weeks												
Complete vaccination + booster ≥24 weeks												
Completely vaccinated				REF	REF	REF	REF	REF	REF	REF	REF	REF
Complete vaccination + booster<12 weeks												
Complete vaccination + booster ≥12 & <24 weeks												
Complete vaccination + booster ≥24 weeks												

** Number of people contributing to each group. Because vaccination status is time changing, the sum of N in all categories will be greater than the total study sample size.*

*** HR adjusted1: Adjusted by age (preferably by age groups considered in the protocol), sex and region in the country.*

**** HR adjusted2: Additionally adjusted by the rest of confounding variables.*

Annex 3. Study sites/countries methods overview, pilot 2022

1. A site methods overview for Portugal

Overview of methods

Outcomes to be analysed	☒ Laboratory-confirmed SARS-CoV-2 infection ☒ Symptomatic SARS-CoV-2 infection ☒ Hospital admission due to COVID-19 ☒ ICU admission due to COVID-19 ☒ Death due to COVID-19
Definition used for symptomatic infection	Laboratory-confirmed infection with at least one symptom of COVID-19 (fever, cough, difficulty breathing, anosmia, ageusia) with onset within [-15; +25 days] from positivity date
Definition used for COVID-19 hospitalisation	Admission for at least 24h with COVID-19 as the primary diagnosis (ICD10 code U07.1) at discharge
Definition used for ICU admission due to COVID-19	Registry of ICU admission in patients with COVID-19 as the primary diagnosis (ICD10 code U07.1) at discharge
Definition used for COVID-19 death	All-cause death with positive RT-PCR test occurred within previous 30 days or intrahospital death of patients admitted with COVID-19 as the primary diagnosis (ICD10 code U07.1) or death coded as COVID-19 in mortality registry (ICD10 code U07.1)
Population of reference for VE estimation	☐ No restriction by previous SARS-CoV-2 infection ☒ Only people without previous infection ☐ People both with or without previous infection (main and secondary analyses)
Type of VE estimation	☒ Absolute (i.e. vs. unvaccinated) ☒ Relative (i.e. between vaccinated in different categories of vaccination)
Which vaccines have been administered in this population (within the National Vaccination Programme)?	☒ AstraZeneca ☒ Janssen ☒ Moderna ☐ Novavax ☒ Pfizer ☐ Other (specify): _____
Which identifier is being used for data-linkage?	National Health Service User number that is used as personal identifier in different electronic heath registries (EHR)
Approx. number of people fulfilling selection criteria by 1 Dec 2021	Approximately 7 500 000
Is information available on which people belong to special populations?	☐ Nursing homes ☐ Institutions such as prisons ☒ Healthcare workers ☐ None of the above is available
Method used to analyse data	☒ Incidence Rates and Poisson regression ☒ Survival analysis and Cox regression
Software used for the analysis	R software
Comments (challenges, limitations, or other)	Target population corresponds to people aged 12 years or more eligible for COVID-19 vaccination, National Health Service user and resident in Portugal mainland (excludes residents in Azores and Madeira Autonomous Regions) – covers approximately 96% of the PT population.

Information on variables and data sources

(Example table, add or remove rows as needed)

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
Outcome	Laboratory-confirmed SARS-CoV-2 infection	BI-SINAVE – contains information on all SARS-CoV-2 laboratory-confirmed infections notified in Portugal	
	Symptomatic SARS-CoV-2 infection	SINAVE+TRACE COVID-19 Symptoms from -15 days to +25 days from the date of positivity are considered. The following symptoms of both systems are crossed: feverishness, fever, cough, through pain, body pain, dyspnea, headache, chest pain, nasal congestion, diarrhea, nausea or vomiting, abdominal pain, widespread weakness or fatigue, arthralgias, anosmia and ageusia.	Quality of data on symptoms decline during peaks
	Hospital admission due to COVID-19	BI-SINAVE +BIMH Date of laboratory confirmation of SARS-CoV-2 is obtained from BI-SINAVE and date and cause of hospital admission are obtained from BIMH	In BIMH episodes are coded at discharge, there is some delay in coding, and only public hospitals are covered
	Death due to COVID-19	BI-SINAVE+BIMH+ SICO + National Health Service User (NHSU) database. The date of death is obtained from NHSU, the cause of death is obtained from SICO, date of laboratory confirmation of SARS-CoV-2 is obtained from BI-SINAVE Information on intra-hospital deaths among those admitted with COVID-19 as primary diagnosis are obtained from BIMH	Delay in coding for cause-specific mortality
	ICU admission	BI-SINAVE +BIMH	In BIMH episodes are coded at discharge, there is some delay in coding, and only public hospitals are covered
Vaccination status	*Only for 12-17 years old* **Completely vaccinated** (with Pfizer) vs. **unvaccinated** Individual will be considered unvaccinated if he/she had no registry of any vaccine uptake. Individual will be considered completely vaccinated 14 or more days following the uptake of the second dose of Pfizer vaccine.	VACINAS – nationwide EHR with data on COVID-19 and other vaccines administrated in Portugal, contains information on brand and date of uptake of each dose of COVID-19 vaccine and other vaccines	For 12- to 17-year-olds (this group is not yet eligible for the booster dose, and was vaccinated mainly with Pfizer) People with intervals between doses <19 days for primary scheme will be excluded from the analysis;
	Only for 12-17 years old **Time since complete vaccination** with Pfizer will be categorised in 28-days intervals (14-41 days, 42-69 days, 70-97 days, 98-123 days and 124+ days), VE will be estimated for	VACINAS	People with intervals between doses <19 days for primary scheme will be excluded from the analysis;

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
	each category of vaccination status considering unvaccinated as the reference group		
	For 18+ years of age **Completely vaccinated** (any vaccine brand) vs. **Unvaccinated** and **booster dose** (any vaccine brand) vs. **Unvaccinated** Individual will be considered unvaccinated if he/she had no registry of any vaccine uptake. Individual will be considered completely vaccinated if he/she has received a complete vaccination schedule according to country recommendations (one dose of Ad26.COV2.S (Janssen) vaccine or two doses of ChAdOx1-S (Oxford/Astra Zeneca), BNT162b2 (Pfizer-BioNTech) or mRNA-1273 (Moderna), or any combination of the three) **14 or more** days after last dose uptake until receiving a booster dose Individual will be considered vaccinated with the booster 14 days following the booster dose uptake	VACINAS	People with intervals between doses <19 days for primary scheme will be excluded from the analysis; People with intervals between doses <19 days for primary scheme will be excluded from the analysis; people with less than 90 days between primary vaccination and booster dose will be excluded from the analysis
	For 18+ years of age **Booster dose** (any vaccine brand) **vs. Completely vaccinated with 124+ days after vaccination.** Time since complete vaccination will be categorised in 28-days intervals (14-41 days, 42-69 days, 70-97 days, 98-123 days and 124+ days) Time since booster will be categorised in 28-days intervals (14-41 days, 42-69 days, 70-97 days, 98+days) VE will be computed for each category of vaccination status to the booster, considering Completely vaccinated with 124+days since vaccination as the reference group	VACINAS	People with intervals between doses <19 days for primary scheme will be excluded from the analysis; People with intervals between doses <19 days for primary scheme will be excluded from the analysis; people with less than 90 days between primary vaccination and booster dose will be excluded from the analysis
Stratification	By age group cohorts (12-17, 18-39, 40-64, 65+ years)		
Adjustment	Age	NHSU The NHSU is used as the central source system and all users included in the study come from this source. Data on sex, date of birth, address and date of death are obtained from NHSU.	
	Sex	NHSU	
	Health Region	NHSU	
	Number of COVID-19 tests in 2020-2022	BI-SINAVE	
	Number of chronic conditions	Primary Care Information System (SIM@SNS) database used to identify comorbidities including anemia, dementia, diabetes, cardiac disease, neuromuscular	HER cover public sector only

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
		disease, rheumatologic disease, obesity, tuberculosis, stroke, pulmonary disease, except asthma, liver disease, and hypertension belong to the group without immunosuppression, and HIV, renal disease, and cancer belong to the group with immunosuppression Data on number of comorbidities is available	Due to data protection issues data on each individual comorbidity is not accessible
	Other vaccines uptake in the last four years (influenza, PCV7, PCV10, PCV13, PPV23)	VACINAS	
	European deprivation index for Portugal at municipality level (Ribeiro AI et al 2018)	Census 2011	Do be updated with Census 2021 as definitive results become available

2. A site methods overview for Denmark

Overview of methods

Outcomes to be analysed	☒ Laboratory-confirmed SARS-CoV-2 infection ☐ Symptomatic SARS-CoV-2 infection ☒ Hospital admission due to COVID-19 ☐ ICU admission due to COVID-19 ☒ Death due to COVID-19
Definition used for symptomatic infection	-
Definition used for COVID-19 hospitalisation	Positive test 14 days before admission, 48hrs after. We are working on a definition for admission 'with' or 'due to' covid.
Definition used for COVID-19 death	Positive test 30 days before death
Population of reference for VE estimation	☐ No restriction by previous SARS-CoV-2 infection ☐ Only people without previous infection ☒ People both with or without previous infection (main and secondary analyses)
Type of VE estimation	☒ Absolute (i.e. vs. unvaccinated) ☒ Relative (i.e. between vaccinated in different categories of vaccination)
Which vaccines have been administered in this population (within the National Vaccination Programme)?	☒ AstraZeneca ☐ Janssen ☒ Moderna ☐ Novavax ☒ Pfizer ☐ Other (specify): _____
Which identifier is being used for data-linkage?	Unique identifier (CPR number)
Approx. number of people fulfilling selection criteria by 1 Dec 2021	5 million
Is information available on which people belong to special populations?	☒ Nursing homes ☐ Institutions such as prisons ☒ Healthcare workers ☐ None of the above is available
Method used to analyse data	☐ Incidence Rates and Poisson regression ☒ Survival analysis and Cox regression
Software used for the analysis	SAS and R

1. Information on variables and data sources

(Example table, add or remove rows as needed)

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
Outcome	SARS-CoV-2	The Danish Microbiology Database (MiBA)	
	SARS-Cov-2-related admission	the Danish National Patient Register (LPR) and MiBA	
	SARS-Cov-2-related death	MiBa and the Danish Civil Registration System (CPR)	
Vaccination status	Vaccination	Danish Vaccination Registry (DDV)	
Stratification	Sex	CPR	
	Age group	CPR	
Adjustment	Sex	CPR	
	Age	CPR	
	Comorbidity	LPR	
	Number of SARS-CoV-2 tests performed	MiBa	Not necessarily health-seeking behaviour, but due to corona passport
	Previous vaccination (PPV23, influenza)	DDV	

3. A Site methods overview for Navarre

Overview of methods

Outcomes to be analysed	☒ Laboratory-confirmed SARS-CoV-2 infection ☐ Symptomatic SARS-CoV-2 infection ☒ Hospital admission due to COVID-19 ☒ ICU admission due to COVID-19 ☒ Death due to COVID-19
Definition used for symptomatic infection	-
Definition used for COVID-19 hospitalisation	Confirmed admission due to COVID-19 according to a Public Health medical doctor revision of all hospitalised patients with a COVID-19 positive result.
Definition used for COVID-19 death	Confirmed death from COVID-19 according to a Public Health medical doctor revision of all deaths in patients with a COVID-19 positive result. Other definitions are possible.
Population of reference for VE estimation	☐ No restriction by previous SARS-CoV-2 infection ☒ Only people without previous infection ☐ People both with or without previous infection (main and secondary analyses)
Type of VE estimation	☒ Absolute (i.e. vs. unvaccinated) ☒ Relative (i.e. between vaccinated in different categories of vaccination)
Which vaccines have been administered in this population (within the National Vaccination Programme)?	☒ AstraZeneca ☒ Janssen ☒ Moderna ☐ Novavax ☒ Pfizer ☐ Other (specify): _____
Which identifier is being used for data-linkage?	Unique identifier
Approx. number of people fulfilling selection criteria by 1 Dec 2021	650.000 inhabitants
Is information available on which people belong to special populations?	☒ Nursing homes ☐ Institutions such as prisons ☒ Healthcare workers ☐ None of the above is available
Method used to analyse data	☒ Incidence Rates and Poisson regression ☒ Survival analysis and Cox regression
Software used for the analysis	SPSS, Stata, SAS

1. Information on variables and data sources

(Example table, add or remove rows as needed)

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
Outcome	SARS-CoV-2	Test results database, that includes the microbiology Database	
	Hospitalisation due to COVID-19 ICU admission due to COVID	Hospital admissions database	
	Death from COVID-19	Clinical and administrative record. Mortality register	
Vaccination status	Vaccination	Vaccination Registry	
Stratification	Sex	Administrative record	
	Age group	Administrative record	
Adjustment			
	Age	Administrative record	
	Comorbidity	Primary healthcare records	
	Calendar month		

4. A Site methods overview for Norway

Overview of methods

Outcomes to be analysed	☒ Laboratory-confirmed SARS-CoV-2 infection ☐ Symptomatic SARS-CoV-2 infection ☒ Hospital admission due to COVID-19 ☒ ICU admission due to COVID-19 ☒ Death due to COVID-19
Definition used for symptomatic infection	Unavailable
Definition used for COVID-19 hospitalisation	Hospitalised with COVID-19 as main cause
Definition used for COVID-19 death	COVID-19 mentioned on death certificate, either as main or contributing cause
Population of reference for VE estimation	☐ No restriction by previous SARS-CoV-2 infection ☐ Only people without previous infection ☒ People both with or without previous infection (main and secondary analyses)
Type of VE estimation	☒ Absolute (i.e. vs. unvaccinated) ☐ Relative (i.e. between vaccinated in different categories of vaccination)
Which vaccines have been administered in this population (within the National Vaccination Programme)?	☒ AstraZeneca ☐ Janssen ☒ Moderna ☐ Novavax ☒ Pfizer ☐ Other (specify): _____
Which identifier is being used for data-linkage?	Hashed national ID number
Approx. number of people fulfilling selection criteria by 1 Dec 2021	4,3 million adults
Is information available on which people belong to special populations?	☒ Nursing homes ☐ Institutions such as prisons ☒ Healthcare workers ☐ None of the above is available
Method used to analyse data	☒ Incidence Rates and Poisson regression ☒ Survival analysis and Cox regression
Software used for the analysis	R

Information on variables and data sources

(Example table, add or remove rows as needed)

Group of variables	Variables (e.g. vaccination date, region, chronic disease)	Data source	Limitations, other comments
Outcome	Infection	MSIS – Reporting system for infectious diseases	
	Hosptialisation due to COVID-19	Norwegian Intensive Care and Pandemic Registry	
	Intensive care unit admission	Norwegian Intensive Care and Pandemic Registry	
	Death due to COVID-19	Norwegian Death Registry	
Vaccination status	Date of vaccinations, vaccination types	SYSVAK – National Immunisation Register	
Stratification/ Adjustment	Age, sex, county of residence, country of birth, date of death	The National Population Register	
	Comorbidities	Norwegian Patient Registry (NPR): individual level data from all public specialist health-care services in Norway.	
	Resident in nursing home	Institution register from Norwegian Labour and Welfare Administration	