European Union

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003673

I-MOVE-COVID-19 network

Multidisciplinary European network for research, prevention and control of the COVID-19 pandemic

European study of COVID-19 vaccine effectiveness against hospitalised SARI patients laboratory-confirmed with SARS-CoV-2

Draft generic protocol

Version:	07
Date:	08 Feb 2021

Version February 2021

I-MOVE Influenza and COVID-19 Networks
WP4 coordinated by Epiconcept

Based on: current literature, I-MOVE generic influenza protocol for hospitalised older adults 2019–2020, I-MOVE-COVID-19 WP4 protocol for investigation into risk factors against severe COVID-19 among hospitalised patients, WHO Europe Guidance document: COVID-19 VE against SARI hospitalisations associated with lab-confirmed SARS-CoV-1 version 5 (20 January 2021), WHO/Euro SARI COVID VE questionnaire v7 (03 February 2021).

DOI: 10.5281/zenodo.4555315

Table of contents

Version history

Version	Date finalised	Created/modified by	Comments
1.0	06 oct 2020	Epiconcept	Initial draft for internal review
2.0	11 oct 2020	Epiconcept	Preliminary draft sent to I-MOVE-COVID-19 hospital sites for comment
2.1	23 oct 2020	Epiconcept	Updated draft after comments received from sites and discussion in site meeting on 15 oct 2020
3.0	09 dec 2020	Epiconcept	Version incorporating changes from further internal reviews
4.0	14 dec 2020	Epiconcept	Final draft sent to sites for 2020
5.0	20 jan 2021	Epiconcept	Draft v4 updated to be in line with WHO/Euro generic guidance document for SARI VE (version 5 from 20 jan 2021)
6.0	21 jan 2021	Epiconcept	Draft v5 updated after further discussion with Epiconcept and WHO/Euro colleagues
7.0	08 feb 2021	Epiconcept	Draft v6 updated after updates to WHO's SARI VE questionnaire v7

Abbreviations

COVID-19	Coronavirus disease 2019
Ct	Cycle threshold
CVE	COVID-19 vaccine effectiveness
EEA	European Economic Area
ECDC	European Centre for Disease Prevention and Control
EU	European Union
GP	General practitioner
ICD	International classification of diseases
ILI	Influenza-like illness
I-MOVE	Influenza – Monitoring Vaccine Effectiveness in Europe
MS	Member States
OR	Odds ratio
RF	Risk factor
RT- PCR	Reverse-transcription polymerase chain reaction
SARI	Severe acute respiratory infection
SARS-CoV-2	Severe acute respiratory syndrome – coronavirus 2
SES	Socioeconomic status
VE	Vaccine effectiveness

- *Arrow marks with italicised text indicate the points that study sites should adapt and provide details for in their study annexes*

- *Changes in this protocol relative to the previous version are indicated in yellow highlight*

1 Background

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). By the end of March 2020, there were over 400,000 cases of COVID-19 reported globally by over 150 countries, with an increasing proportion from countries in the European Union/European Economic Area (EU/EEA) and the United Kingdom (UK). As of 09 December, there were over 14 million cases and over 350,000 deaths reported in the EU/EEA and the UK.(1) The European Centre for Disease Prevention and Control (ECDC), in the latest rapid risk assessment (RRA; published 04 December 2020(2)), reports a recent declining trend in 14-day COVID-19 notifications after more than 2 months of increase and a drop in weekly test positivity, though the latter remains high (473 per 100,000 population; data to end November) for countries in the EU/EEA and the UK.

Data reported to ECDC show that clinical presentations of COVID-19 range from no symptoms (asymptomatic) to severe pneumonia, and that severe disease can lead to death. In the September 2020 RRA, ECDC report a median age of 33 years in data received in the 4 weeks prior to publication. In addition, since the middle of August 2020, incidence of newly diagnosed COVID-19 in those aged between 15 and 49 years (who comprise 65% of all cases) was continuously higher than for any other age-group. Despite this, deaths continue to dominate in the oldest age-group (49% of deaths being in those ≥80 years). In the 4 weeks prior to the September RRA, 4% of diagnosed COVID-19 cases in the EU/EEA with available data were hospitalised, while 0.2% had severe illness (most of these being aged 15–49 years).(3)

ECDC recommends that all hospitalised severe acute respiratory illness (SARI) patients be tested for SARS-CoV-2 virus, so as to detect community transmission and nosocomial outbreaks, as well as for monitoring the intensity and impact of the pandemic.(4) There are also several questions to which urgent answers are needed to improve our understanding of SARS-CoV-2, to inform us of the best interventions to prevent or delay the spread of COVID-19 and newly recommended treatment strategies.

One of the main questions is about the effectiveness of future pandemic vaccines. During the COVID-19 pandemic, having a pre-existing, well-established European platform already in place to rapidly provide severe acute respiratory illness (SARI) surveillance has provided ongoing case identification and will allow for the rapid evaluation of any pandemic vaccine and adaptation of preventive and control strategies.

I-MOVE (Influenza – Monitoring Vaccine Effectiveness in Europe), first established in 2007,(5) was the first network to monitor influenza vaccine effectiveness (VE) within and across the seasons in the European Union (EU) and the European Economic Area (EEA).

In February 2020, many partners, already involved in studies within the I-MOVE network, came together as the I-MOVE-COVID-19 consortium, and were successful in a bid for the European Commission H2020 call on "Advancing knowledge for the clinical and public health response to the novel coronavirus epidemic".

The I-MOVE-COVID-19 consortium aims to obtain epidemiological and clinical information on patients with COVID-19 as well as virological information on SARS-CoV-2, through different work packages (WPs): (a) provision of a flexible surveillance platform, adaptable to the epidemiological situation, through WP2 (primary care surveillance) and WP3 (hospital surveillance), (b) research studies, through WP4 and (c) evaluation of public health interventions (e.g. vaccination, antivirals) in WP2–4, in order to contribute to the knowledge base, guide patient management, and inform the public health response. The surveillance objectives (WP2 for primary care and WP3 for hospital level) have already been achieved through adaptation and expansion of the existing I-MOVE (influenza) network to include COVID-19. The network in 2020 includes primary care networks, hospitals, and national laboratory reference centres in 10 countries across the WHO European Region.[1]

The WP3 hospital surveillance for COVID-19 is coordinated by Public Health Scotland (PHS) with Epiconcept support. The hospital network comprises 11 study sites involving 13 hospitals in five EU Member States[2] (MS) and Albania, the intensive care/high dependency unit (ICU/HDU) network from all hospitals in England and hospital networks in England and Scotland (where coverage is not yet at 100%, but the aim is to include all hospitals). The laboratory component of the network includes regional and national reference centres from the participating countries. Hospital-based studies will utilise data collected from this surveillance network.

The WP4 (research studies) for COVID-19 is coordinated by Epiconcept. While each of the study sites can analyse their data separately, pooling the data for an overall analysis will provide a sample size big enough to answer study questions with reasonable precision.

While control and mitigation strategies such as testing, contact tracing and quarantine procedures help keep COVID-19 in check, having a critical level of immunised people in a population is a further and important method to minimise transmission. Having an effective and safe vaccine against SARS-CoV-2 will help reach this goal while minimising morbidity and mortality among the population. Many vaccines are under development and, as of 09 December 2020, one vaccine is already authorised and being used in the UK,(6) and there are several in different stages of clinical trials.(7) Post-marketing COVID-19 VE (CVE) studies with good precision will be key to determine if vaccines are effective or not among the target groups for vaccination. A high sample size is important to ensure a good precision around the point estimate, and for

[1]Albania, France, Ireland, Lithuania, the Netherlands, Portugal, Romania, Spain, Sweden, and the UK (England and Scotland).

[2]France, Lithuania, Portugal, Romania, and Spain.

the possibility to rapidly assess the VE. Pooling studies in several European countries may achieve these objectives.

This document presents the core European protocol for the 2020 hospital-based study of VE against COVID-19 in hospitalised SARI patients of all ages, outlining the agreed methods for collecting data on COVID-19 and SARS-CoV-2 in each of the individual studies, and including a plan for the pooled analysis. The specificities of each study can be detailed in the individual study annexes. This generic protocol will be updated according to the final vaccination strategy (vaccine products available in each of the participating sites, number of doses required, vaccination data sources, timeline for vaccine roll out, target/priority groups included) and the extent of circulation of each of the viruses at the time when the COVID-19 vaccine will be available for each country.

Importantly, this study is part of the I-MOVE network; as for I-MOVE influenza VE studies, the SARI definition will be used to recruit patients, even though not all patients with COVID-19 may exhibit SARI symptoms.

The study sites will carry out case–control studies, based on the test negative design (TND).(5,8,9)

2 Objectives

2.1 Primary objective

The primary objective will be to measure, for each European participating site country and, for pooled analyses, across all participating sites, the direct effect (effectiveness) of overall and product-specific COVID-19 vaccines against laboratory-confirmed SARS-CoV-2 in hospitalised SARI patients of all ages, in order to provide up-to-date information for the public health response at hospital level by informing prevention measures and highlighting target groups at risk.

- *Each study site to specify the primary objective(s) of their study*

2.2 Secondary objectives

The secondary objectives are:

1. To measure overall and product-specific CVE against laboratory-confirmed SARS-CoV-2 in hospitalised SARI patients by:

- participating study site
- risk group (e.g. specific chronic conditions)
- sex
- age group (0–14 years, 15–49 years, 50–64 years, 65–79 years, 80+ years)
- COVID-19 vaccination target group
- time since vaccination and regularly over calendar time
- vaccine dose (one vs two), if applicable

1. To measure overall and product-specific CVE among SARI patients requiring hospitalisation against
   - specific genetic variant(s) of laboratory-confirmed SARS-CoV-2
   - more severe outcomes (ICU admission, invasive ventilation, in-hospital mortality)

2. To identify potential factors that may modify CVE: chronic conditions, the role of influenza vaccination, the role of settings such as long-term care facilities (LTCFs), the role of statins or other long-term medications (depending on availability of these data in the participating country)

In order to understand duration of protection of vaccine and to identify any differences in CVE among each of these strata, potential target groups for vaccination, and key SARS-CoV-2 virus phenotypic or genotypic evolutions that could affect vaccine performance.

➢ *Each study site to specify the secondary objectives of their study*

3 Methods

3.1 Study design

- At study site level: hospital-based, test-negative, case–control study in each participating hospital
- At EU/EEA level: multicentre hospital-based, test-negative, case–control study, using pooled data from several countries/regions

3.2 Study population

This study is intended to be conducted in countries with pre-existing SARI surveillance systems, or who are already part of the I-MOVE influenza hospital study, to recruit patients for hospital-based CVE studies. The study population for the CVE study will therefore consist of individuals of all ages hospitalised with SARI symptoms in one of the participating hospitals/services, with no contra-indication for COVID-19 vaccination

- *Study sites to describe the setting (number of hospitals included, number of beds, number and type of wards/specialties/services included)*
- *Study sites to describe the existing SARI surveillance system in place, if any, or to indicate if they are already participating in I-MOVE influenza studies*
- *Study sites to describe the study population*
- *Study sites to describe target group(s) for vaccination and order/timeline of vaccination by group (when known)*
- *Study sites to describe the epidemiological situation (incidence, number of COVID-19 hospitalisations, mortality)*

3.3 Study period

The study period starts when the COVID-19 vaccine is available in each of the participating countries and when SARS-CoV-2 is circulating, and when there is sufficient vaccine supply available. The study period is defined initially for each priority vaccination group, and begins for each vaccination group when the vaccination campaign in this group begins. For the general population, the study period will begin when the vaccination campaign is extended beyond target groups to the general population.

Participating hospitals carry out the study throughout the year. However, the study will end if SARS-CoV-2 is no longer circulating in the community (likely to be a 12-month period initially).

- *Each study to define the beginning of the pandemic CVE study period (day/month/year)*
- *Each study site to specify the date of the start of their vaccination campaign, by target group and for the general population (when known).*

3.4 Outcome

The outcome of interest for the primary analysis will be SARS-CoV-2 in patients of all ages hospitalised with SARI symptoms, laboratory-confirmed by PCR documented either on admission to hospital or within 14 days before admission.

Secondary outcomes of interest, in the same patient group, will be laboratory-confirmed genetic variants of SARS-CoV-2 and confirmed SARS-CoV-2 patients with severe outcomes (ICU admission, ventilation, death).

3.5 Definitions

3.5.1 Hospitalised patient

A hospitalised patient will be defined as a SARI patient who has been admitted to one of the participating hospitals during the study period, and has not been discharged home or home-equivalent within 24h.

3.5.2 SARI patient (possible COVID-19 case): WHO SARI case definition

A SARI patient will be defined using the WHO SARI case definition[3] as a hospitalised person with acute respiratory infection, with

a history of fever or measured fever of ≥ 38 C°

and cough

with onset within the last 10 days.

3.5.3 SARI patient (possible COVID-19 case): ECDC possible COVID-19 case definition

If the WHO SARI case definition is not possible, sites may use the ECDC case definition for a possible COVID-19 case.[4] Here, for the I-MOVE-COVID-19 hospital study, a possible COVID-19 case will be defined as

a hospitalised person with **at least one** of the following symptoms:

cough,

[3]WHO SARI case definition: https://apps.who.int/iris/bitstream/handle/10665/333912/WHO-2019-nCoV-Surveillance_Case_Definition-2020.1-eng.pdf?sequence=1&isAllowed=y

[4]ECDC possible COVID-19 case definition: https://www.ecdc.europa.eu/en/covid-19/surveillance/case-definition

fever,

shortness of breath, **or**

sudden onset of anosmia, ageusia or dysgeusia.

SARI patients with onset of symptoms within 14 days prior to hospital admission will be included in the study. Note that hospitals already participating in SARI surveillance systems should not modify the SARI inclusion criteria for surveillance. However, for the CVE analysis, we will only include those patients with onset of symptoms 14 days prior to hospital admission.

3.5.4 SARI patients confirmed as COVID-19 (confirmed cases)

A confirmed COVID-19 case will be defined as a patient hospitalised with SARI symptoms, with a respiratory sample positive for SARS-CoV-2 by PCR,(10) either on admission to hospital or documented within 14 days prior to hospital admission.

3.5.5 SARI who are negative for SARS-CoV-2 (controls)

A control will be defined as a patient hospitalised with SARI symptoms, with a respiratory sample **negative** for SARS-CoV-2 by PCR only, within 14 days prior to hospital admission.[5]

It would be beneficial if countries test for SARS-CoV-2 and influenza (during influenza season), as well as for all other respiratory viruses (as appropriate depending on time of year), if possible. If this is not feasible, then at least all samples negative for SARS-CoV-2 should be also tested for influenza during influenza season, if not already tested at primary care level.

Controls who are negative by PCR but have CT results suggestive of COVID, and those with prior SARS-CoV-2 infection in the 3 months prior to admission, may be excluded as controls in sensitivity analyses (see section 3.5.6 Exclusion criteria below).

- *Each study site to indicate which SARI case definition they will use (WHO or ECDC)*
- *Each study site to indicate which testing strategy they will use (testing all samples for both SARS-CoV-2 and influenza, or only testing for influenza in those negative for SARS-CoV-2)*
- *Each study site to indicate whether they can test for other respiratory viruses, or only SARS-CoV-2 and influenza.*

[5]Note: controls **must** have a negative PCR at admission.

3.5.6 Exclusion criteria

The patient will not be enrolled in the study if she or he:

- is unwilling to participate or unable to communicate and give consent (the consent may also be given by her/his legal representative, or by specific consent procedures, acceptable according to the local ethical review process)
- has a contraindication for the COVID-19 vaccine
- cannot be swabbed due to severe septum deviation, obstruction or other conditions that contra-indicate swabbing
- has a history of hospitalisation within the 14 days immediately prior to this admission (including transfers from another hospital)

Information will be collected on these and other potential exclusion factors and patients will be excluded from primary analyses according to available evidence (not all available at time of writing) on these factors.

In sensitivity analyses, the CVE will be estimated

- with different cut-offs of numbers of days between onset and swabbing, onset and hospitalisation, and between vaccination and onset of symptoms
- excluding those positive to a seasonal coronavirus (e.g. HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1)
- excluding those who are a current control (SARS-CoV-2 negative) but were positive by PCR or serology in the previous year before the current hospitalisation or reported clinically confirmed COVID-19, so as to determine the best cut-off period for having had a previous positive test during the previous year vs "any previous positive test" regardless of date
- excluding those who have received antivirals ≤14 days prior to swabbing (to avoid false negatives; the exact cut-off and types of antivirals will be determined as more research becomes available)

Please see section 8.3.5 on further analyses, in Annex 3.

- *Study sites to define how they obtain informed consent from those who are too unwell at time of recruitment (e.g. oral consent with witness for those in isolation until written consent possible, and/or consent of next-of-kin by telephone, etc.)*

3.6 Initial restriction to priority groups for vaccination

Initially, SARI patients will only be included in the analysis if they are part of a target group for COVID-19 vaccination, for which vaccination rollout has begun. This way all SARI patients included in the study will have had the chance to be vaccinated. SARI patients swabbed prior to rollout of the COVID-19 vaccination campaign in their particular country's target groups will not be included, as they will not be not eligible for vaccination. Later on in the pandemic, when vaccination is rolled out to the general population, the protocol will be updated to reflect this change and all SARI patients will be eligible for inclusion.

3.7 SARI patient identification – algorithm for patient inclusion

The SARI patients will be identified among patients hospitalised for at least 24 hours in one of the participating hospitals. SARI patients should be enrolled and swabbed within 48 hours of hospital admission, and should belong to a vaccination target group for which vaccination has already begun.

3.7.1 Recruitment strategies

For hospitals with electronic patient records and/or diagnosis codes commonly displayed, SARI-related ICD codes (or other codes used for SARI surveillance) will be sought. Patients admitted with any of the ICD codes listed in Table 1 will be approached; those meeting the SARI case definition and the inclusion criteria will be invited to be part of the study and sign informed consent (Figure 1).

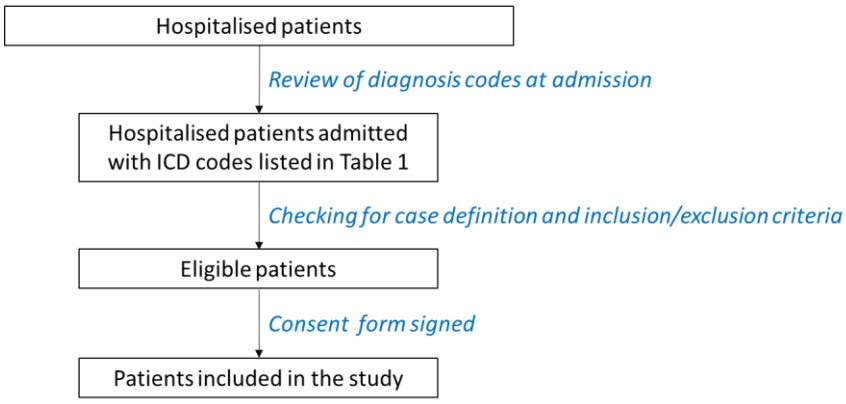


Figure 1: proposed inclusion algorithm for hospitals/services relying on common use of ICD codes, I-MOVE-COVID-19 hospital-based COVID-19 vaccine effectiveness study.

For hospitals where ICD codes at admission are not systematically collected or accessible, systematic screening of all patients admitted will be organised. This should be done by sensitisation of the medical staff at the beginning of the study (Figure 2), followed by regular study coordinator review.

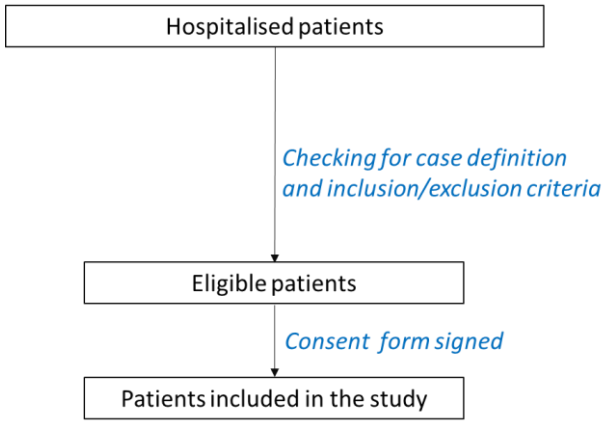


Figure 2: proposed inclusion algorithm for hospitals/services systematic screening of all admitted patients, I-MOVE-COVID-19 hospital-based COVID-19 vaccine effectiveness study.

Retrospective recruitment (or "catching-up" already diagnosed patients) is not recommended for the CVE study, as not all COVID-19 patients exhibit SARI symptoms and it will be difficult to determine retrospectively the reasons for testing.

- *Each study site to describe procedures used to identify study participants*

In case of test scarcity, extreme workloads, or budget limiting inclusion to a threshold of patients, the study sites may need to switch from exhaustive to systematic sampling (e.g. inclusion of patients every second day, or only on certain days in the week). Systematic sampling procedures should be planned ahead by the study sites. During the period of systematic selection, the study sites will make sure to document the sampling fraction.

- *Study sites not testing all SARI cases to describe the systematic sampling procedure. If systematic sampling is not done, explain criteria for testing*

Table 1. List of diagnosis codes for which patients could be screened for onset of SARI symptoms, I-MOVE-COVID-19 hospital-based VE study.

Category	Morbidity	ICD-9	ICD-10
Influenza-like illness	Cough	786.2	R05
	Difficulty breathing	786.05	R06
	Sore throat	784.1	R07.0
	Dysphagia	787.20	R13
	Fever	780.6	R50.9
	Headache	784.0	R51
	Myalgia	729.1	M79.1
	Fatigue/malaise	780.79	R53.1, R53.81, R53.83
Cardiovascular diagnosis	Acute myocardial infarction or acute coronary syndrome	410-411, 413-414	I20-23, I24-25
	Heart failure	428 to 429.0	I50, I51
Respiratory diagnosis	Emphysema	492	J43.9
	Chronic obstructive pulmonary disease	496	J44.9
	Asthma	493	J45
	Myalgia	729.1	M79.1
	Dyspnoea/respiratory abnormality	786.0	R06.0
	Respiratory abnormality	786.00	R06.9
	Shortness of breath	786.05	R06.02
	Tachypnoea	786.06	R06.82
	Other respiratory abnormalities	786.09	R06.00, R06.09, R06.3, R06.89
Infections	Pneumonia and influenza	480-488.1	J09-J18
	Other acute lower respiratory infections	466, 519.8	J20-J22
	Viral infection, unspecified	790.8	B34.9
	Bacterial infection, unspecified	041.9	A49.9
	Myocarditis	429.0	I40.9
	Bronchitis	490, 491	J40, 41
Inflammation	SIRS* non-infectious without acute organ dysfunction	995.93	R65.10
	SIRS* non-infectious with acute organ dysfunction	995.94	R65.11

**SIRS: Systemic inflammatory response syndrome*

Category	Morbidity	ICD-9	ICD-10
Abdominal symptoms	Vomiting	787.0	R11
	Diarrhoea	009.3, 787.91	A07.9, K52.9
	Abdominal pain	789.0	R10
Diagnoses related to deterioration of general condition or functional status	General physical deterioration, lethargy, tiredness	780.79	R53.1, R53.81, R53.83
	Anorexia	783.0	R63.0
	Feeding difficulties	783.3	R63.3
	Abnormal weight loss	783.21	R63.4
	Other symptoms and signs concerning food and fluid intake	783.9	R63.8
	Disorientation/altered mental status	780.97	R41.0
	Dizziness and giddiness	780.4	R42
	Infective delirium	293.0, 293.1	F05
	Coma	780.01	R40.2
	Transient alteration of awareness	780.02	R40.4
	Other alteration of consciousness (somnolence, stupor)	780.09	R40.0, R40.1
	Febrile convulsions (simple), unspecified	780.31	R56.00
	Complex febrile convulsions	780.32	R56.01
Other	Anosmia, ageusia, myalgia	781.1, 729.1	R43.0, R43.2, M79.1

3.8 Laboratory methods

Study nurses or physicians will collect respiratory specimens (see Section 4.4) from all eligible patients, respecting safety standards for COVID-19 and following WHO biosafety guidelines.[6]

- *Each study site to describe the type (nasopharyngeal/oropharyngeal or both) and number of swabs taken for each patient*

[6]Any non-propagative diagnostics (e.g. sequencing, RT-PCR) should be conducted at a facility using procedures equivalent to biosafety level 2 (BSL-2), while propagative work (e.g. virus culture, isolation or neutralisation assays) should be conducted at a containment laboratory with inward directional airflow (BSL-3). Patient specimens from suspected or confirmed cases should be transported as UN3373, 'biological substance category B'. Viral cultures or isolates should be transported as category A, UN2814, 'infectious substance, affecting humans'.(11)

Quality control tests should systematically be run using PCR to ensure presence of cells in the respiratory specimens. In the absence of cells, a negative result should be considered inconclusive and a second swabbing should take place if possible.

The ECDC-recommended SARS-CoV-2 laboratory confirmation is by viral RNA detection with nucleic acid amplification tests, such as RT-PCR.(4,10) Isolates will undergo molecular analysis for currently circulating SARS-CoV-2 virus. During the influenza season, it is recommended that influenza virus tests should also be performed, as long as there is circulation of influenza viruses.(4) This is especially important here in order to identify any potential association between influenza and SARS-CoV-2.

Following the procedures outlined by each study, a systematic sample of isolates (or all isolates) will undergo gene sequencing. The sampling procedure can include sequencing all isolates, or a systematic sample thereof. The systematic sample should be representative of cases and be large enough to provide reasonable precision when calculating proportions of virus change over time.

Gene sequences, if sequencing is performed, should also be uploaded to GISAID's open access EpiCoV platform. Gene sequence information can be provided directly to the I-MOVE-COVID-19 central hub, or the GISAID EpiCoV accession number can be provided alongside the I-MOVE-COVID-19 unique identifier to link these data (see Annex 5). Processed genetic information, e.g. name of genetic clade, can also be included within the epidemiological database.

- *Each study site to describe the laboratory procedures (samples taken, storage, transport)*
- *Each study site to describe the tests and the kits used (and their sensitivity, specificity, PPV) for COVID-19 and, if needed, other respiratory virus detection*
- *Each study site to describe if the laboratory participates in QA/QC (Quality Assurance/Quality Control) schemes*
- *Each study site to describe the selection of specimens and the procedures for genetic and antigenic characterisation, where appropriate (see Annex 4 for an example of presentation of these results)*
- *Each study site to describe genetic and antigenic analyses and specify sequencing methods*
- *Study sites to describe whether specimens are tested for other respiratory viruses (e.g. whether influenza continues to be tested systematically during the season and stops once the influenza season is over, or is only tested when the COVID-19 result is negative, etc.)*

3.9 Exposure (vaccination)

3.9.1 Definition of vaccination status

Current pandemic COVID-19 vaccine

An individual will be considered as vaccinated against COVID-19 with a product-specific vaccine (see section "COVID-19 vaccination status ascertainment") during the current pandemic under the following categories:

- ***Fully vaccinated*** (two-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received both doses** at least 14 days* before onset
- ***Fully vaccinated*** (single-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received one dose** at least 14 days* before onset
- ***Partially vaccinated*** (two-dose vaccine only): to be defined according to vaccine product recommendations, but most likely a patient will be considered partially vaccinated if they have **received one of two doses** at least 14 days* before onset
- A SARI patient will be considered as ***unvaccinated*** if s/he **did not receive COVID-19 vaccine or** if s/he **was vaccinated after onset** of symptoms.

**The exact number of days will depend on the vaccine; this number may change and the protocol will be updated when more information is available.*

It is crucial that the vaccination status and date of vaccination variables are collected with the utmost care to ensure data completeness and quality.

3.9.2 COVID-19 vaccination status ascertainment

The main exposure of interest in this study will be vaccination history with any COVID-19 pandemic vaccine. The vaccination history includes date of administration, type of vaccine and brand name, and number of doses. Documenting the batch codes (where this is feasible) will allow identification of the vaccine brand, the vaccine content and the dose.

The sources of information for the vaccination status may include:

- vaccination registry (preferred option)
- consultation of the patient's vaccination card or patient's hospital notes
- interview with the patient's GP

- interview with the patient's pharmacist
- data from the patient's insurance company showing evidence of pharmacy delivery or re-imbursement for COVID-19 vaccine during the 2020 winter season
- interview with the patient and/or his/her relatives.

  - *Each study site to describe how vaccination status ascertainment will be performed and validated*

  - *Each study site to document*
    - *vaccine products used*
    - *places of vaccination (GPs, specific vaccination centres, etc.)*
    - *precise mode of vaccine ascertainment (self-report, card, registry, etc.)*
    - *if no precise dates of vaccination collected, the variable allowing a patient to be defined as vaccinated or unvaccinated*
    - *vaccine status ascertainment validation*

3.10 Confounding factors and effect modifiers

3.10.1 Pre-existing chronic conditions

Patients (in particular, those who are ultimately included as controls) with underlying conditions may be included due to an exacerbation of these conditions, unrelated to SARI. These patients may be more likely to be infected with COVID-19, or to develop more severe disease than the source population. Furthermore, these patients may be more likely to be vaccinated against COVID-19 than the source population. We may therefore overestimate the CVE. We will document the presence of underlying conditions among all recruited SARI patients.

Underlying conditions which could be potential confounding factors/effect modifiers are shown in Table 2 (for their ICD codes, see Annex 1).

3.10.2 Severity of underlying condition(s)/healthcare utilisation

The severity of an underlying condition could be an effect modifier or a confounding factor, i.e. not just presence of underlying condition. In order to document and control for healthcare-seeking behaviour in control groups and the severity of the underlying conditions, information on the number of hospital admissions due to underlying conditions in the 12 months prior to inclusion in the study will be collected. We will also collect information on the number GP consultations (face-to-face or telephone/video) in the previous 12 months, where available.

Table 2. List of mandatory and optional pre-existing conditions as potential effect modifiers or confounding factors

Mandatory	Optional
• Asthma	• anaemia/chronic haematologic disease
• immunodeficiency (including HIV) and organ transplant	• asplenia

• cancer (solid organ and haematological)
• diabetes mellitus
• heart disease (excluding hypertension)
• hypertension
• lung disease
• obesity ***or***
  ○ height and weight, ***or***
  ○ BMI[7] (sites to include whichever is feasible/available)
	• chronic liver disease/cirrhosis
	• dementia
	• neuromuscular disorders
	• renal disease (exclude acute renal failure)
	• rheumatologic diseases
	• stroke
	• tuberculosis

- *Each study site to update the mandatory list to include pre-existing conditions defining target groups for vaccination in your country*
- *Each study site to define the list of chronic conditions to be included and state whether they are used to define target groups for vaccination, as well as any pre-existing medications being taken, and describe what the sources of information for these will be*

3.10.3 Ethnicity (optional)

Some studies have shown that certain ethnic groups may be at higher risk, either for becoming infected with, or for developing severe COVID-19. Uptake of, or access to vaccination may also be linked to ethnicity. Not all sites collect, or are able to collect, this information. Even if all sites were able to collect ethnic group, each group may be defined differently in different countries. The definitions for each ethnic group will need to be standardised across the I-MOVE-COVID-19 sites before we can robustly collect this information and use it to investigate VE by ethnic group for pooled data. However, any site(s) who can or do already collect this information may find it useful for examining national VE by ethnic group.

- *Each study site to indicate whether they can or do collect ethnic group, and define their ethnic groups and (proposed) method of collecting the data (e.g. by self-reported ethnicity, from patient notes, or any other method)*

[7]Note: obesity defined as BMI>29.

3.10.4 Medication status for chronic condition(s) (optional)

The use of specific types of chronic medications prior to vaccination or illness may modify or confound the effect of the vaccine.

Definition of medication status for pre-existing chronic condition(s):

- An individual will be considered as "on medication" if s/he has received more than one dose of the medication during the 6 months before
    - the first dose of pandemic COVID-19 vaccination (if date of medication use available: for the analysis measuring the effect of chronic medications on CVE) or
    - onset of SARI symptoms or
    - hospitalisation
- An individual will be considered as "not on chronic medication" if s/he did not receive medication
    - before the periods specified above in the protocol to document such medication use

3.10.5 Chronic medication use status ascertainment (optional)

The medication history includes the date the patient started on the medication(s) where known; else just the year, if the patient was known to have been on medication before vaccination or symptom onset, or if the precise date is unknown. If both of these are unknown, then a simple yes/no response to whether the patient was on the chronic medication before hospitalisation will be used. In addition, medication history will include medication brand name, and number/ frequency of doses.

The sources of information for chronic medication status may include:

- consultation of the patient's hospital record
- interview with the patient's GP
- interview with the patient's pharmacist
- data from the patient's insurance company showing evidence of pharmacy delivery or re-imbursement for chronic medication during the winter 2020 season
- interview with the patient and/or his/her relatives

➢ *Each site will define chronic medication use based on data collected*

➢ *Each site to describe how chronic medication status ascertainment will be performed*

3.10.6 Pregnancy status

Pregnancy status will be collected and coded for women aged 15–55 years as follows: pregnant (yes/no/do not know), and if yes: trimester (1/2/3/do not know).

3.10.7 Smoking history

Smoking history will be collected and coded as follows: never-smoker, former smoker (stopped smoking at least one year before inclusion in the study), current smoker (or stopped in the past year).

3.10.8 Healthcare worker

The definition of a healthcare worker for the purposes of this study is anyone working (paid or on a regular voluntary basis) in healthcare who has contact with any type of patient) during his/her work. This includes: doctors, nurses, emergency medical personnel, medical and nursing students having contact with patients, as well as porters and cleaners. It also includes anybody working with resident contact in a nursing/residential home for the elderly. Study sites should collect information on healthcare worker status where possible.

- *Each study site to indicate whether they can collect information on healthcare worker status*

3.10.9 Other occupations (optional)

As some occupations predispose to greater exposure, and may be a proxy for attitudes towards vaccination, where possible countries may collect additional information on occupation for stratification in analysis by type of occupation (optional).

- *Each study site to indicate whether they intend to/can also collect information on occupation in general (optional)*

3.10.10 Other vaccinations

We will collect information on vaccinations received for influenza (including date of vaccination) and for pneumococcal disease (pneumococcal polysaccharide vaccine, PPV, and/or pneumococcal conjugate vaccine, PCV), where available.

Previous influenza and pneumococcal vaccinations

Vaccination against influenza in the current influenza seasons (where this information is available) and year of vaccination against pneumococcal diseases will be collected.

The sources of information for these vaccinations may include:

- vaccination registry
- consultation of the patient's vaccination card
- interview with the patient's GP
- interview with the patient's pharmacist
- data from the patient's insurance company showing evidence of pharmacy delivery or re-imbursement of influenza vaccine during the 2020 winter season
- interview of the patient and/or his/her relatives

➢ *Each study site to indicate whether vaccination information (including date of vaccination) will be available for influenza and/or pneumococcal disease*

➢ *Each study site to describe how previous influenza and pneumococcal vaccination status is documented*

3.10.11 Antiviral administration

The use of antivirals prior to swabbing may lead to misclassification biases. We will run sensitivity analyses excluding patients who were administered antivirals prior to swabbing. We will document whether the patients received any antiviral treatment in the 2 weeks preceding symptom onset and the type (curative or preventive) of antivirals received.

➢ *Each study site to list any antivirals administered*

3.10.12 Functional impairment/frailty

Frailty may be associated with both vaccination and the risk of developing severe symptoms in case of COVID-19 infection. There are different ways in which countries may capture the presence of functional impairment related to the ability of patients to do a range of daily activities without assistance. Where possible, the Barthel Index(12) should be used. If this is not possible, countries may use simple questions related to the ability of patients to do a range of daily activities (e.g. walking, bathing) without assistance. Finally, in the absence of these, a proxy for frailty may be used, such as inclusion of a question on residence, with response options for long-term care facility and residence at home/not at home but "with support" or "without support".

➢ *Each study site to describe the measure(s) used to capture functional impairment/frailty*

3.10.13 Presence of influenza and other respiratory viruses

The COVID-19 vaccine may potentially have an impact on influenza and other respiratory viruses. This will mean that controls will not represent the source population giving rise to the cases (i.e. they will have lower vaccination coverage than the rest of the target population). We will try to document the presence of other respiratory infections (e.g. influenza) among patients testing negative for SARS-CoV-2, as well as in those who are positive.

- *Each study site to list the other respiratory infection viruses tested for (including influenza)*

3.10.14 Setting (LTCF vs community)

Older and vulnerable populations, already at greater risk of severe disease, are often situated in localised settings such as those for long-term care, where they are more at risk for localised outbreaks than residents in the general community (as was observed in the early phase of the pandemic, when many hospitalisations arose in long-term care populations). It is also possible, however, that an entire LTCF (residents and staff) is vaccinated, providing them with less chance of exposure than the general population. Stratifying by setting (LTCF vs community) will help to adjust for differences between SARI patients who are LTCF residents and those who are not.

- *Each study site to ensure setting is captured, particularly for SARI patients aged over 64 years*

3.10.15 Socioeconomic status or deprivation (optional)

Individuals with lower socioeconomic status (SES), who may be living in crowded conditions and have less access to good nutrition and potentially more co-morbidities, will be at greater risk of infection and severe disease, and may also be less able to access vaccination services. Stratifying by SES, if collected, will allow comparison of CVE between those of lower and higher SES.

- *Each study site to decide on the best way to represent SES for their population and to describe the measure(s) used to capture SES or level of deprivation*

3.10.16 Previous SARS-CoV-2 infection

Individuals who have been previously infected may have a greater response to the vaccine or be less likely to be reinfected even if unvaccinated. Study participants should be asked about prior test results and prior symptoms to elucidate possibility of prior infection. We will exclude those with prior infection in sensitivity analyses.

- *Each study site to collect and describe measure(s) used to capture prior SARS-CoV-2 infection*

3.10.17 Practice of non-pharmaceutical interventions (optional)

Individuals not practising routine use of NPIs will be at greater risk of infection, and may also be less likely to access vaccination services. Stratifying by NPI use, if collected, will allow comparison of CVE between those regularly using and not using these key NPIs. Key NPIs proposed are: use of a mask in public, frequently washed hands with soap and water for at least 20 seconds, used sanitiser when soap and water were unavailable, ensured physical distancing in public (remaining at least 1 m away from others), with the following options: always, usually, sometimes, never, not applicable.

- *Each study site to decide on the best way to collect NPI use in their population and to describe the measure(s) used to capture use of NPIs. In particular it is important that these questions are asked without judgement*
- *Sites to adjust distance etc. to match national recommendations*

3.11 Sample size

Providing VE estimates for each separate study is one of the objectives of this project. Therefore, the minimum sample size should be estimated for each study in order to obtain precise VE estimates. The pooled analyses should not prevent study teams from including a big enough sample size to obtain exact estimates for each separate study.

- *Each study site to specify the minimum sample size calculation*

In VE estimation, sample size estimation is different from sample size estimation in hypothesis testing. Rather than being concerned about a VE estimation to cross 0% or not, we are more concerned with the precision around the estimate. For example, if we have a VE of 70%, a lower boundary confidence interval of 1% does not provide us with a very informative VE estimate, even if the confidence interval does not include 0%. We are more concerned to have a VE estimate that is precise around the point estimate of 70% (e.g. with a lower boundary of 50%). The precision around the estimate is more informative than whether the confidence intervals include 0% or not. Indeed, if we have a low VE estimate, which can be the case in particular stratified analyses, we would need a huge sample size to provide a VE estimate that does not include 0%. For example, if the true VE is 5–10%, then a study providing a lower boundary not including 0% would be unreasonably large.

The following sample size estimates focus on precision of the VE estimate (Table 3). As the lower confidence interval boundary is always larger than the upper confidence interval boundary, we focus on a precision of the lower confidence interval, ranging between 10 and 30%. We also assume a case to control ratio of 1:1. We include varying vaccine coverage among the source population between 30% and 50%, varying vaccine effectiveness (with the OR between 0.1 and 0.7).

Table 3. Sample size calculations

Precision to lower CI boundary	Controls /case	Detectable OR	Vaccine coverage in source population/ controls	Number of cases	Number of controls	CVE	CI
0.3	1	0.1	0.3	60	60	90	60–98
0.3	1	0.2	0.3	85	85	80	51–92
0.3	1	0.3	0.3	118	118	70	40–85
0.3	1	0.4	0.3	157	157	60	30–77
0.3	1	0.5	0.3	203	203	50	20–69
0.3	1	0.6	0.3	255	255	40	10–60
0.3	1	0.7	0.3	314	314	30	0–51
0.2	1	0.1	0.3	96	96	90	70–97
0.2	1	0.2	0.3	148	148	80	60–90
0.2	1	0.3	0.3	216	216	70	50–82
0.2	1	0.4	0.3	299	299	60	40–73
0.2	1	0.5	0.3	395	395	50	30–64
0.2	1	0.6	0.3	507	507	40	20–55
0.2	1	0.7	0.3	633	633	30	10–46
0.1	1	0.1	0.3	241	241	90	80–95
0.1	1	0.2	0.3	433	433	80	70–87
0.1	1	0.3	0.3	681	681	70	60–77
0.1	1	0.4	0.3	985	985	60	50–68
0.1	1	0.5	0.3	1346	1346	50	40–58
0.1	1	0.6	0.3	1764	1764	40	30–49
0.1	1	0.7	0.3	2240	2240	30	20–39
0.3	1	0.1	0.4	42	42	90	60–98
0.3	1	0.2	0.4	63	63	80	49–92
0.3	1	0.3	0.4	91	91	70	40–85
0.3	1	0.4	0.4	125	125	60	30–77
0.3	1	0.5	0.4	165	165	50	20–69

0.3	1	0.6	0.4	212	212	40	10–60
0.3	1	0.7	0.4	265	265	30	0–51
0.2	1	0.1	0.4	68	68	90	70–97
0.2	1	0.2	0.4	111	111	80	60–90
0.2	1	0.3	0.4	168	168	70	50–82
0.2	1	0.4	0.4	238	238	60	40–73
0.2	1	0.5	0.4	323	323	50	30–64
0.2	1	0.6	0.4	421	421	40	20–55
0.2	1	0.7	0.4	534	534	30	10–46
0.1	1	0.1	0.4	170	170	90	80–95
0.1	1	0.2	0.4	323	323	80	70–87
0.1	1	0.3	0.4	528	528	70	60–77
0.1	1	0.4	0.4	786	786	60	50–68
0.1	1	0.5	0.4	1098	1098	50	40–58
0.1	1	0.6	0.4	1466	1466	40	30–49
0.1	1	0.7	0.4	1891	1891	30	20–39
0.3	1	0.1	0.5	32	32	90	60–98
0.3	1	0.2	0.5	51	51	80	51–92
0.3	1	0.3	0.5	77	77	70	40–85
0.3	1	0.4	0.5	109	109	60	30–77
0.3	1	0.5	0.5	148	148	50	20–69
0.3	1	0.6	0.5	193	193	40	10–60
0.3	1	0.7	0.5	246	246	30	0–51
0.2	1	0.1	0.5	51	51	90	70–97
0.2	1	0.2	0.5	90	90	80	60–90
0.2	1	0.3	0.5	142	142	70	50–82
0.2	1	0.4	0.5	208	208	60	40–73
0.2	1	0.5	0.5	289	289	50	30–64
0.2	1	0.6	0.5	384	384	40	20–55
0.2	1	0.7	0.5	495	495	30	10–46
0.1	1	0.1	0.5	129	129	90	80–95
0.1	1	0.2	0.5	262	262	80	70–87
0.1	1	0.3	0.5	447	447	70	60–78
0.1	1	0.4	0.5	687	687	60	50–68
0.1	1	0.5	0.5	983	983	50	40–58
0.1	1	0.6	0.5	1337	1337	40	30–49
0.1	1	0.7	0.5	1751	1751	30	20–39

The sample size estimates above are for the crude analysis and an adjusted analysis would require a higher sample size.

The sample size should also be respected for each population subgroup for which a sub (stratified) analysis (e.g. effect modification) is planned.

See also the Analysis section on sample size requirements for analyses.

3.12 Data

3.12.1. Sources of information

Data will be collected using a standardised questionnaire/data collection form. The source(s) of data may include:

- hospital medical records
- interview with patient or his/her family
- interview with patient's GP
- interview with patent's pharmacist
- vaccination register
- laboratory

➢ *Each study site to define the sources of information used for each variable collected and the potential limitations*

3.12.2. Collected information

Collected information falls under the following main categories:

- study identification
  - country, hospital
  - vaccination target groups
  - first ward of referral
  - ICU/other ward of admission
- patient characteristics *(ethnic group optional)*
- SARI signs, symptoms
  - current
  - previous clinical symptoms (if no prior tests done)
- other symptoms
- dates
  - vaccination (COVID-19, influenza, pneumococcal disease)
  - onset of symptoms
  - admission, discharge
  - swabbing
- laboratory
  - type of swab (nasopharyngeal, sputum, etc.) *(optional)*

    - type of test
    - results (including information antigenic and genetic analysis, where available)
    - previous positive PCR or antigen test for SARS-CoV-2, if feasible (for sensitivity analyses)
- underlying chronic conditions, including obesity (see sections 3.9.1–3.9.4)
    - use of medications for chronic conditions *(optional)*
    - number of hospitalisations for chronic conditions in the previous 12 months *(optional)*
    - number of GP consultations in the previous 12 months *(optional)*
- presence of influenza and other respiratory viruses (see section 3.9.10)
- vaccination and antivirals (see section 3.9.8–3.9.9)
    - pandemic vaccination including number of doses, date, product
    - current influenza vaccination
    - pneumococcal vaccination status, type of vaccine and either date or year of vaccination *(optional)*
    - antiviral administration *(optional)*
- non-pharmaceutical interventions *(optional)* (section 3.10.17)
- functional status or proxy by residence type (see section 3.9.9)
- setting (e.g. LTCF)
- SES/deprivation *(optional)*

Pandemic vaccine data collected will be revised as more information on the vaccine(s) and target groups becomes available.

These are described in more detail below (see also Annex 1 for a complete variable list including coding).

3.12.3. Study identifiers

We will document the following study characteristics.

- Country, site, priority vaccination target group(s)
- Hospital (note: actual name of hospital will not be collected, but a unique number will be assigned by each participating site to each hospital, to allow adjustment by hospital in analyses)
- Patient unique ID (note: this is not a patient identifiable ID such as date-of-birth or national ID number, but a unique identifier for the pooled database)

3.12.4. Hospital/ward information

We will document the following dates and other hospital information:

- Date of onset, admission, discharge, death
- First ward of referral

- Any hospital stay (for pre-existing chronic condition) in previous 12 months *(optional)*
- Date of swab/sample

3.12.5. Patient characteristics

We will document following patient characteristics to describe the study population.

- Age
- Sex
- Smoking history (see section 3.9.5)
- Pregnancy
- Healthcare worker
- Occupation *(optional)*
- Clinical frailty score at admission (where possible; see section 3.9.10) *(optional)*
- Ethnic group *(optional)*
- SES/deprivation (optional)

  - *Each country to describe type of clinical frailty score in use, where available*
  - *Each country to describe community measures in place to limit exposures*

3.12.6. Clinical characteristics (symptoms and markers of severity)

We will document following clinical characteristics and markers of disease severity.

The four following key symptoms indicate respiratory illness:

- fever or feverishness
- cough
- shortness of breath
- sore throat

In addition, as many study sites will also use this protocol to measure influenza VE, the following four symptoms should also be collected:

- headache
- myalgia
- malaise
- deterioration of general condition (asthenia, weight loss, anorexia)

The following three symptoms have been associated with COVID-19 illness and are part of the ECDC COVID-19 case definition. They should also be collected where possible:

- anosmia
- ageusia
- dysgeusia

It would also be helpful to collect whether these symptoms appeared with **sudden onset.**

The following 14 symptoms are **optional** for the hospital-based COVID-19 CVE study:

- coryza, rhinitis
- chest pain
- chills
- fatigue
- nausea
- vomiting
- abdominal pain
- diarrhoea
- conjunctivitis
- confusion
- dizziness
- tachypnoea or other signs of low oxygen saturation (restlessness)
- rash or other dermatological manifestation
- palpitations/rapid heartbeat

We also will collect the following information, which can be used to indicate severity.

- Oxygen use
- ICU admission
- Invasive ventilation
- Death

We will collect the **date of first key symptom onset**, as well as information on COVID-19 test(s) and laboratory results, including information on antigenic and genetic analysis, when available. It is vital that this information is collected, as well as **date of vaccination**, **date of swab** (to allow estimation of and

stratification by delay from swab to onset), **date of admission** (to allow estimation of and stratification by time from onset to hospitalisation, and to measure length of hospital stay), and **date of discharge/death** (to allow measurement of length of hospital stay).

3.12.7. Case definition

Collection of good quality symptom information is crucial for the CVE study in order to be able to validate the case definition used. As a minimum, we need to collect data on the symptoms required for the WHO or ECDC case definitions. The following variables are imperative for application of the WHO SARI case definition:

- fever or feverishness
    - if fever: measured fever (with temperature), or feverishness
- cough
- onset date
- admission date

The following additional variables will also be imperative if the study site is using the ECDC case definition:

- shortness of breath
- sudden onset
- anosmia
- ageusia
- dysgeusia

3.12.8. Data entry validation

For hospitals using electronic medical records, if paper questionnaires are used, a sample of them will be checked against the medical records and against the study database. The agreement between patient records/reports by study participants will be measured when/if records are available.

- ➢ *Each study site to specify how data are validated*

3.12.9. Data management

Data entry and transfer

Web-based data collection methods or paper-based methods can be used. Data entry will include checks to minimise data entry errors. Double data entry is recommended if paper forms are used for data collection.

Laboratory information will be reported to the study site coordinator using the reporting procedures existing in each study site for the severe COVID-19 risk factor study.

For the multi-centre pooled analysis, study sites may send an anonymised database to Epiconcept through the secure data transfer system EpiFiles. Each individual study database will be sent to the Epiconcept team database using a secure protocol (see Annex 4: Dataflow for pooled database). All personal identifier information such as names, addresses, and medical registration codes will be deleted before data transmission to Epiconcept, where all individual data will be pooled. Study databases can be sent to Epiconcept in any format (e.g. Stata, CSV, EpiInfo, etc.).

Epiconcept provides the option of web-based data collection methods, if so desired by the countries. These methods can also be combined with paper-based methods.

If the Epiconcept web-based data collection methods are not used, data can be coded as outlined in Annex 2, but it is not required.

There are two methods for data transfer for the COVID-19 VE study:

(1) As for I-MOVE influenza for most sites, data collection through your usual method with transfer of your electronic database to Epiconcept through the secure data transfer platform, EpiFiles

(2) As for I-MOVE influenza for some sites, data entry directly into the Epiconcept software Voozanoo, which Epiconcept will adapt to include COVID-19 variables

- *Study sites to specify procedures of data management and procedures to comply with the GDPR requirements*
- *Study sites to indicate which of the two data collection options they will use*
- *Study sites to provide a codebook that includes the variable names, variable descriptions, and the coding of variable values (if not using Voozanoo).*

Data cleaning

Summary and frequency tables as well as visual representations of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of discharge from hospital before date of onset of symptoms). Ideally, these checks will be included as warnings in the electronic questionnaire in order to avoid inconsistencies in the data entry. These values will be checked against the questionnaires or queried with the hospitals. Any changes to the data will be documented and stored separately from the crude database. Any recoding of data (e.g. age) will be documented. A guide and/or an example Stata do-file for data cleaning will be provided if so desired.

- *Study sites to specify the data checking and cleaning process*

Data storage and data management for pooled analysis

The minimum dataset will be transmitted to Epiconcept where individual data will be pooled. Data will be stored in the EU data repository, as required by the European Commission (EC), and all data management procedures must comply with the General Data Protection Regulations (GDPR).

The Epiconcept team will conduct the pooled analysis. Data validation, cleaning and verification will be carried out at study level. A country (or study) identifier will be included in each record (e.g. ES for Spain, UK for the United Kingdom), a hospital code will be included (e.g. a unique number), and each record will be given a unique number. This number will also be included in the study team's database and will be used by the Epiconcept team and the study sites during pooling, so that records can be traced back whilst maintaining anonymity, if there should be any further queries. Tracing back will be performed by the study sites, not by the coordinating team. Study databases can be sent in any format.

Summary and frequency tables and graphic displays of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of respiratory specimen collection before date of onset of symptoms). Any improbable, illegal or missing values will be reported to the study site in question.

Any subsequent changes to the data will be fully documented and stored separately from the crude database, to ensure reproducibility and transparency of data management.

A study site-specific flowchart of exclusions and restrictions will be shared with each of the study sites. Variables will be recoded and new variables generated. The recoded data will be stored separately from the crude data and recoding will be documented.

Missing data

Any missing data will be described. If many data are missing and there is no evidence of bias in the missing data, and variables that are considered good predictors of the missing data are available, multiple imputation methods at study level will be used to replace missing values. A sensitivity analysis will be carried out comparing results from the complete case analysis (where records with missing data will be dropped) and the full set analysis (with imputed data).

3.13 Data analysis

3.13.1. Individual level analysis

The analysis will be carried out first for each individual study site and shared with the site study team for validation. Epiconcept can provide example scripts if desired or carry out the site-specific data analysis at the site's request.

Briefly, cases and controls will first be described by baseline characteristics. Patients will be described according to:

- sex
- age group
- healthcare worker status
- time: month of symptom onset
- COVID-19 vaccination status
- symptoms
- absence, presence of at least one, presence of more than one high-risk condition
- specific chronic conditions (e.g. respiratory, cardiovascular diseases)
- pregnancy, smoking status
- influenza and pneumococcal vaccination status
- respiratory co-infections (where available)
- severity (ICU, on ventilation, death)

An example layout of this descriptive analysis is provided in Table 4 on the next page.

This study is a case control study (test-negative design). The measure of association is an odds ratio (OR). This can be measured by logistic regression. An OR = 1 indicates no association between an exposure and the outcome. An OR>1 indicates a potential risk factor, an OR < 1 indicates a potential protective factor, noting that the confidence interval around the OR helps with its interpretation.

For vaccination as preventive factor, the CVE can be computed as CVE = (1 – OR)*100. A 95 % confidence interval is computed around the point estimate.

Univariable analysis will be carried out to measure the CVE against being a laboratory-confirmed COVID-19 case. Stratified analyses (by sex and age group, for example) can follow to better understand potential effect modifiers and confounders.

Table 4: Example of descriptive table for cases and controls; I-MOVE-COVID-19 hospital-based vaccine effectiveness study, Europe, 2021.

Variables	Number of laboratory-confirmed COVID-19 cases /total n (%)	Number of test-negative controls /total n (%)
Median age (IQR)	X	X
Missing	X	X
Age groups		
0–14	x/x (x)	x/x (x)
15–44	x/x (x)	x/x (x)
45–64	x/x (x)	x/x (x)
65–79		

≥ 80	x/x (x)	x/x (x)
Missing	X	X
Sex		
Female	x/x (x)	x/x (x)
Missing	X	X
Healthcare worker	x/x (x)	x/x (x)
Missing	X	X
Days between onset of symptoms and swabbing		
0	x/x (x)	x/x (x)
1	x/x (x)	x/x (x)
2	x/x (x)	x/x (x)
3	x/x (x)	x/x (x)
4–7	x/x (x)	x/x (x)
COVID-19 vaccination	x/x (x)	x/x (x)
Missing	x	x
Etc.		

Prior to multivariable analysis, a model development strategy will be determined. In the final step, multivariable analysis will be carried out to take confounding factors and potential effect modifiers into account. This will provide adjusted ORs from which the CVE can be estimated using the formula above.

3.13.2. Output tables presenting CVE estimates

In order to present the results in the most transparent manner and to enable the reader to best understand the data, tables similar to the one illustrated by Table 5 can be used (variables presented just as example of the output format). Useful information includes numbers of cases and controls (overall and vaccinated) and presentation of results for different models.

Table 5: Example of table showing vaccine effectiveness against COVID-19 adjusted for various covariables by sex and age group, hospital-based I-MOVE-COVID-19 vaccine effectiveness study, WHO European Region, 2021.

Type/subtype	Population included	Analysis scenarios/adjustments made	CVE (%)	(95%CI)
COVID-19	All ages	N (cases/ vaccinated; controls/ vaccinated) Crude Adjusted for onset week (cubic spline)		

	Adjusted for sex
	Adjusted for chronic condition
	Adjusted for age (cubic spline)
	Adjusted for onset week, age (cubic spline)
	Adjusted for onset week, chronic condition
	Adjusted for onset week, age (cubic spline), chronic conditions, sex
0–49 years	N (cases/ vaccinated; controls/ vaccinated)
	Crude
	Adjusted for onset month, age (cubic spline)
50 years and over	N (cases/ vaccinated; controls/ vaccinated)
	Crude
	Adjusted for onset week, age (cubic spline), chronic condition, sex

3.13.3. Pooled analysis

Epiconcept conducts the pooled analysis. The higher sample size for this analysis will provide more power (and precision). Data can be coded as outlined in Annex 2, or a codebook can be provided by the study teams to Epiconcept that includes the variable names, descriptions and coding. Epiconcept performs all necessary data cleaning. Epiconcept documents and shares any further data cleaning and analysis with all study coordinators to ensure it can be reproduced.

See Annex 3 for detailed guidelines to the pooled analysis. For the pooled data, interim analyses will be conducted in different periods if appropriate and according to the available sample size.

The timing to conduct each interim analysis will depend on the time needed to reach the appropriate sample size. This will depend mainly on the incidence of hospitalisation, COVID-19 incidence, vaccination coverage, the recruitment strategy within hospitals and the number of participating hospitals/services per hospital.

The pooled analysis will be carried out in a similar way to the site-specific analysis. Country or study site will be included potentially as a fixed effect or as a random effect in a multilevel model. Statistical heterogeneity between study sites will be determined, using Q-test and the I^2 index(13).

3.14 Consent

Each study will comply with national ethics committee requirements. Informed consent will be required from all participants or legal tutors. The national ethics committees will specify whether oral or written consent will be required. Specific consent procedures may be needed for unconscious patients and patients with deterioration of general condition or functional status, unable to sign the consent (e.g. oral witnessed consent, consent by the next of kin, etc). A copy of the ethical approvals should be sent to the coordinating centre, with a copy of the template used to obtain informed consent.

- *Each study site to describe the procedures to comply with the national ethics committee requirements and the type of informed consent needed, as well as whether consent can be obtained for a legal tutor*
- *Each study site to send a copy of the ethical approval to the coordinating centre*

3.15 Dissemination of results

The enrolment of COVID-19 cases will be regularly updated by each study coordinator on a website developed for the multicentre study. Initial CVE estimates will be disseminated as soon as possible, in the second quarter of 2021, with a later CVE in the third quarter of 2021. (Note that this may be revised depending on how the pandemic progresses.)

3.15.1. Publications, scientific communication

Results of the individual studies should only be published in open-source journals (this is a requirement of the European Commission's H2020 funding received for this surveillance project). Each study coordinator will decide which scientific conferences will be attended in order to present the results. An article presenting the results of the pooled analysis and CVE estimates for the EU/EEA will be submitted to an open-source, peer-reviewed journal.

The list of authors will respect the recommendations of authorship stated by the International Committee of Medical Journal Editors: http://www.icmje.org/ethical_1author.html. The actual authorship for the pooled article will be discussed and agreed with the study teams at the beginning of the study.

3.16 Training

Investigators and data collectors will be trained on the study protocol before the start of the study. They will receive the protocol, questionnaires and laboratory respiratory specimen collection procedures.

- *Each study site to describe the training to be organised*

4 Logistical aspects

4.1 Study leader

In each study site, a principal investigator will coordinate the study at the country level and act as focal point for the European study. Epiconcept is in charge of the pooled analysis.

4.2 Human resources

In each hospital/hospital network, an investigator will be in charge of monitoring data collection at the hospital level. Study investigators at the hospital will collect information from cases and controls. The specific human resources needed in each country are detailed in the study annexes. Epiconcept ensures the overall coordination of the various studies.

4.3 Supervision

If feasible in the pandemic context, site visits and joint workshops (remote if required) will be organised by the coordinating team/study sites in order to carry out an appraisal of the ongoing studies in the various countries involved. The appraisal team will be composed of two persons from the various project partners.

4.4 Respiratory specimen collection

By default, the respiratory specimen will be collected through nasal/nasopharyngeal swabbing or concurrent nasal and oral/oropharyngeal swabbing (or endotracheal aspirates in ICU). Personal protection must be used in accordance with guidelines.

- *Each study site to describe the specimen collection procedures.*

4.5 Laboratory tests

High specificity is needed for COVID-19 confirmation. COVID-19 laboratory confirmation will be done using RT-PCR or multiplex RT-PCR.

- *Each study site to describe the tests and the kits used for COVID-19 and influenza; and, if needed, other respiratory virus detection*
- *Each study site to specify sequencing methods.*

Quality control tests should systematically be run using PCR to test for presence of cells in the respiratory specimens.

- *Each study site to describe quality controls for specimens*
- *Each study site to describe genetic and antigenic analyses.*

4.6 Standard operating procedures

Standard operating procedures should be used by investigators during all the steps of the study for identification of study subjects, data collection, laboratory methods, data entry, monitoring, etc. Epiconcept has prepared a data entry SOP for use with Epiconcept's online Voozanoo 4 questionnaire, for sites using this platform.

- *Each study site to develop (or adapt pre-existing) study SOP to be used by the study team*

4.7 Reports

Each study site will write a report at the end of the study and submit it to the study coordination team. Epiconcept will write a final report presenting the results of the pooled analysis.

5 Limitations

With any multi-centre study, there is always the potential for heterogeneity among sites. In addition, during a pandemic with such high caseloads for hospitals, there may be difficulties in collecting all data, and not all included cases will have laboratory confirmation. There is also the possibility that very severely ill patients (e.g. those who are extremely frail and/or in nursing homes) may not be admitted to hospital at all, and would be missed by the study. Potential limitations to the VE estimates for COVID-19 are discussed below.

5.1 Potential biases

5.1.1 Bias from pooled estimates

With data from a number of different hospitals from different countries being pooled, any bias in the individual studies will influence the pooled estimate. The power of the test for the presence of heterogeneity between individual studies is low if there are few sites/countries. In this case, the test may not be able to detect heterogeneity between them, despite it being present. It is important that heterogeneity is also assessed using qualitative knowledge about differences between studies. Depending on the nature of the bias, the inclusion of biased studies in the pooled estimate could lead to over- or underestimation of the true association between COVID-19 vaccination and the outcome.

There are many conditions which could lead to bias in a single site or hospital. With this new virus, there are new and evolving surveillance systems and strategies in each participating country. There are not only different tests being used, but a variation in the number of tests used to declare an individual negative, for example. Another example is that, when under high pressure (e.g. high volume of patients to be admitted during a peak in the epidemic for any site), it is possible that some hospitals may switch to admitting only suspected COVID-19 patients, while others focus on non-COVID-19 patients. In the event of the former type of hospital being an I-MOVE-COVID-19 VE participating hospital, this could affect the recruitment of controls and result in cases being predominantly recruited from one hospital over another. If a participating site only has one hospital providing data, this could mean they are only able to provide information on cases. Conversely, if the single participating hospital was designated a non-COVID-19 admitting hospital, this site would only be able to provide information on controls.

To allow for complete assessment of heterogeneity, sites need to document all changes in their COVID-19 surveillance system during the study period.

- *Each country to document any changes in COVID-19 surveillance during the study period, including allocation of participating hospitals to COVID-19 or non-COVID-19 admission status*

5.1.2 Negative confounding

Negative confounding refers to biases that reflect the fact that high-risk groups (people more likely to develop severe complications) will be more likely to be vaccinated and therefore reduce CVE. If negative confounding is present, the CVE will be underestimated. Adjustment for potential negative confounding factors documented in the study (e.g. presence of chronic diseases) will minimise negative confounding.

5.1.3 Positive confounding

Positive confounding refers to biases that reflect a 'healthy vaccine effect'. People with a healthy lifestyle will be more likely to accept vaccination, thus leading to an increase of measured CVE. Or, similarly, people being in a state of "extreme frailty" will not be offered vaccination and, because they are frail, may be more likely to have severe disease. Persons with risk-taking behaviours may also be averse to vaccination, which may also increase their exposure to disease. If positive confounding is present, CVE will be overestimated.

5.1.4 Unmeasured confounding

Positive and negative confounding will be minimised through stratification and multivariable analysis. It will not be possible to rule out the presence of characteristics in the study population for which no information is collected in the study questionnaire and that therefore could lead to positive or negative confounding. Therefore, some residual unmeasured confounding may remain.

- *Each study site to describe the potential limitations and representativeness of the subjects included*

5.1.5 Previous infection in cases or control; inclusion of asymptomatic controls

Individuals who have been previously infected may have a greater response to the vaccine or be less likely to be reinfected even if unvaccinated. It is possible that some of the controls (those testing negative for SARS-CoV-2) may have themselves been positive for SARS-CoV-2 some time before, but were asymptomatic. The proportion of these (potentially immune individuals) in each country's dataset would depend on the circulation of the virus in the community in the months before the hospitalisation of the control. Knowledge of their prior infection could affect their likelihood to be vaccinated. For example, if someone knew that they had had COVID-19, despite having no symptoms (e.g. if they had had a screening test), they may be subsequently less likely to be vaccinated. This would lower vaccination coverage among controls and increase CVE.

Ascertainment of which controls may have had previous SARS-CoV-2 infection can be attempted by asking about previous SARS-CoV-2 tests and results, as well as prior clinical symptoms. However, among the controls, there could potentially be several patients with prior SARS-CoV-2 infection. Results should be interpreted in light of this, and an estimate of a range of potential bias should be calculated around the CVE estimates. Sensitivity analyses should be conducted excluding any SARI patient with previous SARS-CoV-2 infection confirmed either by PCR or by ELISA.

As antibody tests become more widespread, then this may be included in the protocol.

5.1.6 Inclusion of influenza-positive controls

It is possible that SARI patients who are also influenza positive will be unsuitable controls. There is limited information on co-infection with influenza and COVID-19 from the first wave of the pandemic in Europe, partly due to the timing of the pandemic being towards or after the end of the 2019–20 influenza season in many countries. The low number of coinfections described in the literature(14,15) could be due to lack of opportunity (there being little influenza circulating at that time) or to a negative correlation between the two infections, with those positive for COVID-19 being unlikely to also be positive for influenza. In addition, those receiving COVID-19 vaccination are highly likely to have also received influenza vaccine. There is therefore the potential for there to be a relationship between being positive for influenza and receiving COVID-19 vaccination, which introduces bias. Sensitivity analyses will be conducted excluding controls who are positive for influenza.

5.1.7 Validation of exposure

The vaccination status is the exposure of interest and the validity of vaccination data should therefore be checked carefully. If the vaccination status is reported by the patient only without further proof, information bias may occur. We will validate the vaccination status of cases and controls using an independent source (i.e. vaccination register, GPs).

- *Each study site to describe how the source of exposure validation and its potential limitations*

5.1.8 Other potential biases

Controls could come from different source populations with varying risk for infection with SARS-CoV-2, varying probability for acquiring COVID-19 vaccination, etc. (e.g. depending on time of year). Time (onset

date) will be used to adjust for seasonal differences. Analyses will also be stratified by time (e.g. onset quarter).

- *Each country to describe timeline of vaccination for different target groups*

5.2 Representativeness of subjects included in the study

The study includes only cases that are hospitalised. Health-seeking behaviour may differ by country depending on the case management strategy (e.g. recommendation to stay at home with mild symptoms, and only seeing a GP if symptoms persist, and then hospitalisation if severe). In some cases, the management strategy will have an impact on the delay between onset of symptoms and hospitalisation. This, in turn, may have an impact on the time lag between onset and respiratory specimen collection, and may affect positivity rates between study sites. Beside the collection of dates of onset/admission/respiratory specimen collection, health-seeking behaviour and case-management strategies should be described for each study and it should be noted how these may affect the CVE estimates.

In some sites (those where the CVE study is not part of routine SARI surveillance), very severely ill patients will not be able to give informed consent. If this is the case, some very severe cases may not be included.

Importantly, the representativeness of the controls needs consideration. (For example, if controls were to be all influenza **and** COVID-19 negative, we need to consider whether they are representative of the source population in terms of vaccine coverage.)

- *Each study site to describe the potential limitations in terms of representativeness of the subjects included*
- *Each study site to describe case-management strategy in their country*

6 References

1. European Centre for Disease Prevention and Control. COVID-19 situation update for the EU/EEA and the UK, as of 9 December 2020 [Internet]. European Centre for Disease Prevention and Control. 2020 [cited 2020 Dec 9]. Available from: https://www.ecdc.europa.eu/en/cases-2019-ncov-eueea

2. European Centre for Disease Prevention and Control. Rapid risk assessment: Risk of COVID-19 transmission related to the end-of-year festive season [Internet]. European Centre for Disease Prevention and Control. 2020. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/Risk-assessment-COVID-19-transmission-related-the-end-of-year-festive-season.pdf

3. European Centre for Disease Prevention and Control. Rapid risk assessment: Increased transmission of COVID-19 in the EU/EEA and the UK – twelfth update [Internet]. European Centre for Disease Prevention and Control. 2020 [cited 2020 Oct 11]. Available from: https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-increased-transmission-twelfth-update

4. European Centre for Disease Prevention and Control. Rapid risk assessment: Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK – seventh update [Internet]. European Centre for Disease Prevention and Control. 2020 [cited 2020 Mar 27]. Available from: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic

5. Valenciano M, Ciancio B, I-MOVE study team. I-MOVE: a European network to measure the effectiveness of influenza vaccines. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2012 Sep 27;17(39).

6. UK Government. UK authorises Pfizer/BioNTech COVID-19 vaccine [Internet]. GOV.UK. [cited 2020 Dec 14]. Available from: https://www.gov.uk/government/news/uk-authorises-pfizer-biontech-covid-19-vaccine

7. World Health Organization. COVID-19 vaccines [Internet]. [cited 2020 Dec 14]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines

8. Committee for Medicinal Products for Human Use. Guideline on influenza vaccines. Non-clinical and clinical module [Internet]. London, UK: European Medicines Agency; 2014 Jul. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/07/WC500170300.pdf

9. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. Vaccine. 2013 Apr 19;31(17):2165–8.

10. World Health Organization. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases: Interim guidance [Internet]. 2020 [cited 2020 Apr 6]. Available from: https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117

11. World Health Organization. Laboratory biosafety guidance related to coronavirus disease 2019 (COVID-19): Interim guidance [Internet]. WHO; 2020. Available from: https://apps.who.int/iris/bitstream/handle/10665/331138/WHO-WPE-GIH-2020.1-eng.pdf

12. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud. 1988;10(2):61–3.

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.

14. Ozaras R, Cirpin R, Duran A, Duman H, Arslan O, Bakcan Y, et al. Influenza and COVID-19 coinfection: Report of six cases and review of the literature. J Med Virol. 2020 Jun 4;

15. Antony SJ, Almaghlouth NK, Heydemann EL. Are coinfections with COVID-19 and influenza low or underreported? An observational study examining current published literature including three new unpublished cases. J Med Virol. 2020 Jun 12;

7 Bibliography

1. Blettner, M, Sauerbrei, W, Schlehofer, B, Scheuchenpflug T, Friedenreich C. Traditional reviews, metaanalyses and pooled analyses in epidemiology. Int J Epidemiol 28, 1–9(1999).
2. Lemeshow S, Hosmer D, Klar J. Sample size requirements for studies estimating odds ratios and relative risks. Stat Med 1988 ;(7) :759–64.
3. National Research Council. Combining information – statistical issues and opportunities for research. Washington, DC: National Research Council; 1992. doi: https://doi.org/10.17226/20865.
4. Rodrigues L, Kirkwood BR. Case–control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990 Mar;19(1):205–13.
5. Rothman, K, Greenland, S, Lash, T. Modern Epidemiology (ed.). Philadephia: Lippincott–Williams–Wilkins, 2008.
6. Smith-Warner, S.A, Spiegelman, D, Ritz, J, Albanes, D, Beeson, W.L, Bernstein, L. et al. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. Am J Epidemiol 2006 ;163 :1053–64.
7. Stukel TA, Demidenko E, Dykes J, Karagas MR. Two-stage methods for the analysis of pooled data. Stat Med 2001 Jul 30;20(14):2115–30.

8 Annexes

8.1. Annex 1. List of ICD-9 and ICD-10 codes for pre-existing chronic conditions

Category	ICD-9	ICD-10	Underlying conditions included
Anaemia	280–285	D50-64	Nutritional anaemias, Haemolytic anaemias, Aplastic and other anaemias and other bone marrow failure syndromes
Asplenia	746.87, 759.0	Q89.01, Q20.6, Z90.81	Malposition of heart, Anomalies of spleen, Isomerism of atrial appendages, Acquired and Congenital absence of spleen
Asthma	493.0, 493.1, 493.9	J45	Extrinsic asthma, Intrinsic asthma, Predominantly allergic asthma, Non-allergic asthma, Mixed asthma, Asthma unspecified
Chronic liver disease	571	K70, K72-74, K754, K769	Alcoholic liver disease, Hepatic failure, Chronic hepatitis, Fibrosis and cirrhosis of liver, Other inflammatory liver diseases
Cardiovascular diseases	093, 112.81, 130.3, 391, 393–398, 402, 404, 410–429, 745, 746, 747.1, 747.49, 759.82, 785.2-3	A52.01, B37.6, B58.81, I05-9, I11, I13, I20-25, I26.09, I26.9, I27, I30-51, I97.0-1, R00.1, T81.718A, T81.72XA, T82.817A, T82.818A, Q20-24, Q25.1-2, Q26.0-1, Q26.8, Q87.4, R01.1-2	Syphilitic aneurysm of aorta, Candidal endocarditis, Toxoplasma myocarditis, Chronic rheumatic heart diseases, Ischemic heart diseases, Hypertensive heart and chronic kidney disease, pulmonary embolism with acute cor pulmonale, pulmonary heart diseases, diseases of pulmonary vessels, Other forms of heart disease (including Nonrheumatic valve disorders, pericarditis, endocarditis, myocarditis, cardiomyophathy, heart failure, block, cardiac arrhythmias, heart failure), Complication of other artery / vein following a procedure, Embolism of cardiac/vascular prosthetic devices, implants and grafts, congenital malformations of cardiac chambers and connections or heart, Coarctation or atresia of aorta, Congenital malformations of great veins, Marfan's syndrome, Cardiac murmur
Diabetes	250	E10-11	Type 1 and Type 2 diabetes mellitus
Hypertension	401, 401.0, 401.9, 405, 405.91, 405.99,	I10, I15.8, I15, I15.1, I15.2, I97.3, I27.0	Hypertension (essential and secondary), Secondary to other [renal or endocrine] disorders, Malignant hypertension

Category	ICD-9	ICD-10	Underlying conditions included
Obesity	27800, 278.01, 278.03	E66.01, E66.2, E66.9	Obesity
Immunodeficiency* or organ transplant	042, 279, V08, V42	B20, D80-84, D89.8-9, Z21, Z94	HIV, immune deficiency, organ or tissue replaced by transplant
Neuromuscular disorders	358.00-358.1, 358.8, 358.9, 378.73, 775.2	G70-G70.01, G70.2, G70.80, G70.81, G70.9, G70.89, G73.7,	Myasthenia gravis, Myoneural disorders NEC/NOS, Neuromuscular disease strabism, Congenital and developmental myasthenia, Lambert-Eaton syndrome, Myoneural disorder NOS
Renal disease	274.1, 408, 580–591, 593.71–593.73, 593.9	M10.30, N00-19, N20.0, N28.9	Gout due to renal impairment, Glomerular diseases, Renal tubulo-interstitial diseases, Acute kidney failure and chronic kidney disease, Calculus of kidney, Disorder of kidney and ureter, unspecified
Dementia	290, 294, 331	F01, F03, F05, G30, G31, G91, G94	Vascular dementia, other dementia, Delirium due to known physiological condition, Alzheimer's disease, Other degenerative diseases of nervous system
Stroke	348, 438	G93, I67.83, I69	Brain disorders, Posterior reversible encephalopathy syndrome, Sequelae of cerebrovascular disease
Rheumatologic diseases	446, 710, 714	M30-34, M35.0, M35.5, M35.8-9, M05-06, M08, M12.00	Polyarteritis nodosa and related conditions, Other necrotizing vasculopathies, Systemic lupus erythematosus (SLE), Dermatopolymyositis, Systemic sclerosis, Sicca syndrome, Multifocal fibrosclerosis, other systemic involvement of connective tissue, Rheumatoid arthritis with rheumatoid factor, Other rheumatoid arthritis, Juvenile arthritis, Chronic post-rheumatic arthropathy
Cancer	140–208	C00-96	Malignant neoplasms and neuroendocrine tumours
Lung disease	011, 490–511, 512.8, 513–517, 518.3, 518.8, 519.9, 714.81	A15, J40–47, J60–94, J96, J99, J182,	Respiratory tuberculosis, Bronchitis, not specified as acute or chronic, Chronic bronchitis, Emphysema, Other chronic obstructive pulmonary disease, Asthma, Bronchiectasis, Hypersensitivity pneumonitis due to organic dust, Pneumoconiosis, Airway disease due to specific organic dust, Hypersensitivity pneumonitis due to organic dust, Respiratory conditions due to inhalation

		of chemicals, gases, fumes and vapor, Pneumonitis due to solids and liquids, Respiratory conditions due to other external agents, Acute respiratory distress syndrome, Pulmonary oedema, Pulmonary eosinophilia, not elsewhere classified, Other interstitial pulmonary diseases, Abscess of lung and mediastinum, Pyothorax, Pleural effusion, Pneumothorax and air leak, Other pleural conditions, Intraoperative and postprocedural complications and disorders of respiratory system, not elsewhere classified, Other diseases of the respiratory system, Hypostatic pneumonia, unspecified organism,
	M34.81, M05.10	**Systemic sclerosis with lung involvement, Rheumatoid lung disease with rheumatoid arthritis**
Tuberculosis	A15–A19	Primary respiratory tuberculosis, Respiratory tuberculosis unspecified, Tuberculosis of nervous system, Tuberculosis of other organs, Miliary tuberculosis

*Note: Patients who are only treated with glucocorticoids and have no other immune deficiency, are considered immune suppressed when treated with high-dose corticosteroids (≥ 20 mg/day of prednisone or equivalent for ≥2 weeks) in the last 3 months.

8.2. Annex 2: List of variables, definitions and coding; I-MOVE-COVID-19 hospital-based COVID-19 vaccine effectiveness study minimum dataset

Individual data

- *study sites to list all the variables collected and their coding*
- *study sites to indicate all modifications in the variables collected compared to variables below*
- *New variables for existing partners (not already collected for I-MOVE influenza) or coding changes, highlighted in green*
- *Optional variables shaded in grey (new and old variables)*

	Variable	Type	Values and coding	Definition
Study identifiers	idcountry	Numeric (categorical)	Coded according to international country codes	Identifier uniquely identifying the country (for pooled datasets only)
	idsite	Numeric or text (categorical)	Unique code	Identifier uniquely identifying the site (for countries with multiple sites, each having at least one participating hospital)
	id	Numeric	Unique integer	Unique number for each patient
	hospitalcode	Numeric	Unique integer	Unique number for each hospital
	consent	Numeric	0 = No 1 = Yes 2 = Not required 8 = Do not know	Agreement of patient to participate (where appropriate, i.e. for countries requiring consent)
	consent_sp	Text		Reason provided for non-participation
Target group	target	Numeric (categorical)	0 = not in a target group 1 = age 80+ years 2 = health personnel 3 = in long-term care 4 = other 8 = do not know	Country-defined target groups for vaccination; note these are just examples, countries should provide codes for each target group ("other" option in case of future new target groups)
	target_sp	Text		Specify other target group
Hospital/ward information	admitdate	Date	dd/mm/yyyy	Date of hospital admission
	dischargedate	Date	dd/mm/yyyy	Date of hospital discharge
	deathdate	Date	dd/mm/yyyy	Date of in-hospital death
	icuadmitdate	Date	dd/mm/yyyy	Date of ICU/HDU admission
	icudisdate	Date	dd/mm/yyyy	Date of ICU/HDU discharge

	Variable	Type	Values and coding	Definition
Hospital/ward information continued	hospitalward	Numeric (categorical)	0 = Special COVID-19 ward	First ward of referral (NOTE: this may alternatively be sent as text, as for influenza)
			1 = Lung, pulm/respir.	
			2 = Internal medicine	
			3 = Infectious diseases	
			4 = Emergency or A&E	
			5 = Cardiology	
			6 = Geriatric	
			7 = ICU or HDU	
			9 = Other	
			8 = Do not know	
	hospitalward_oth	Text		Specify other ward
	icu	Numeric (categorical)	0 = No	Admission to intensive care unit (ICU) or high-dependency unit (HDU)
			1 = Yes	
			8 = Do not know	
Patient characteristics	sex	Numeric	0 = female	Sex of patient (Note coding change to match with ECDC coding)
			1 = male	
			3 = other	
			8 = do not know	
	dob	Date	dd/mm/yyyy	Date of birth (only if no age) *(optional)*
	smoking	Numeric (categorical)	0 = Never	Never, former (stopped smoking at least 1 year before inclusion in the study), current smoker (or stopped in the past year)
			1 = Former	
			2 = Current	
			8 = Do not know	
	pregnant	Numeric (categorical)	0 = No	Whether patient is pregnant (for women aged 15 – 55 years)
			1 = Yes	
			8 = Do not know	
	trimester	Numeric (categorical)	1 = first trimester	If patient is pregnant, indicate which trimester (if known)
			2 = second trimester	
			3 – third trimester	
			8 = Do not know	
	hcw	Numeric (categorical)	0 = No	Whether the patient is a healthcare worker
			1 = Yes	
			8 = Do not know	

	Variable	Type	Values and coding	Definition
Patient characteristics (continued)	essential_worker	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether the patient is any other type of essential worker with much human contact (e.g. teacher, police person, supermarket worker) *(optional)*
	esswork_sp	Text		Specify which other type of essential worker *(optional)*
	residence	Numeric (categorical)	0 = at home, not dependent on home support/care 1 = at home, but dependent on home support/care 2 = institutionalised (LTCF) 3 = Other 8 = Do not know	Patient residence at time of SARI onset. Whether patient was living at home or was institutionalised in long-term care facility, or had pre-hospital dependence on home support/care (note: can be used as proxy for frailty) *(optional)*
	residence_sp	Text		Specify other residence (e.g. prison) *(optional)*
	occupation	Text		Patient's occupation (note: this may be collected another way, e.g. by national occupational code, depending on country) *(optional)*
	ethnic	Numeric (categorical)		Patient's ethic group (note: codes will be country-specific) *(optional)*
	ethnic_sp	Text		Other ethnic group not specified in coding above *(optional)*
	ses	Numeric (categorical)		Indicate results from socioeconomic or deprivation index used *(optional)*
Underlying chronic conditions	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
	diabetes	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diabetes

	Variable	Type	Values and coding	Definition
Underlying chronic conditions (continued)	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)
	lungdis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Lung disease (excluding asthma)
	height	Numeric (integer)		Height of patient in metres
	weight	Numeric (integer)		Weight of patient in kg
	bmi	Numeric (1 d.p.)		BMI of patient (only if available in place of missing weight/height)
	obese	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Obesity (only if height, weight and BMI not collected; can be calculated)
	gp_visit	Numeric (integer)		Number of times patient consulted GP for underlying chronic condition in 12 months prior to COVID diagnosis
	hosp_visit	Numeric (integer)		Number of times patient was admitted to hospital for an underlying chronic condition in the 12 months prior to COVID diagnosis
Underlying chronic conditions continued *(optional)*	anaemia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Anaemia/chronic haematologic disease *(optional)*
	asplenia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Asplenia (absence of/damage to spleen) *(optional)*
	dement	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Dementia *(optional)*
	liverdis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chronic liver disease (excluding cancer) *(optional)*

	Variable	Type	Values and coding	Definition
Underlying chronic conditions continued *(optional)*	neuromusc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Neuromuscular disorder *(optional)*
	rendis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Renal disease (excluding cancer and acute renal failure) *(optional)*
	rheumat	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rheumatologic disease *(optional)*
	stroke	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Stroke *(optional)*
	tuberc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Tuberculosis *(optional)*
Pre-symptomatic medication *(optional)*	statin_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was on statins since or from 01 January 2020 *(optional)*
	ace_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	ACE inhibitor (angiotensin converting enzyme inhibitors) *(optional)*
	arb_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	ARB (angiotensin II receptor blockers) *(optional)*
	nsaid_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	NSAID (non-steroidal anti-inflammatory drugs) *(optional)*
	metform_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Metformin *(optional)*
	steroids_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do *not* know	Steroids *(optional)*

	Variable	Type	Values and coding	Definition
Pre-symptomatic medication *(optional)*	corticost_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Corticosteroids *(optional)*
	dmards_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Biological disease-modifying anti-rheumatic drugs (DMARDs) e.g. rituximab, tocilizumab, etc. *(optional)*
	chemo_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chemotherapy (within 6 months or currently) for cancer *(optional)*
	gliclaz_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Gliclazides (for diabetes or heart failure) *(optional)*
	psychotrop_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Psychotropic drugs (including benzodiazepine, etc.) *(optional)*
	antivir_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Antivirals *(optional)*
	chloroq_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chloroquine *(optional)*
	hydroxychloroq_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Hydroxychloroquine *(optional)*
	other1_pre_sp	Text		Other pre-symptomatic medication #1 *(optional)*
	other2_pre_sp	Text		Other pre-symptomatic medication #2 *(optional)*
	other3_pre_sp	Text		Other pre-symptomatic medication #3 *(optional)*
Vaccination (other than against COVID-19)	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

	Variable	Type	Values and coding	Definition
Vaccination (other than against COVID-19), continued	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 8 = Do not know	Received PCV7/10 or 13 vaccination
	pcv_vaccdate	Date	dd/mm/yyyy	Date of last PCV7/10 or 13 vaccination
Symptoms at or prior to admission (for SARI case definition)	feverish	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 > 38°C
	suddenonset	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sudden onset
	headache	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Headache
	sorethroat	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sore throat
	cough	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cough
	sob	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Shortness of breath
	malaise	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Malaise

	Variable	Type	Values and coding	Definition
Symptoms at or prior to admission, continued	general_deter	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Deterioration of general condition (asthenia or loss of weight or anorexia)
	myalgia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Myalgia
	anosmia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Loss of sense of smell
	ageusia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Loss of sense of taste
	dysgeusia	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Alteration of sense of taste
	onsetdate	Date	dd/mm/yyyy	Date of onset of first symptom
Symptoms at or prior to admission, continued *(optional)*	chills	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	"Chills", or shivering *(optional)*
	tach	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Tachypnoea or signs of low oxygen saturation *(optional)*
	coryza	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Coryza *(optional)*
	confusion	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Confusion *(optional)*
	dizzy	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Dizziness *(optional)*

	Variable	Type	Values and coding	Definition
Symptoms at or prior to admission, continued *(optional)*	chest	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chest pain *(optional)*
	palp	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Heart palpitations *(optional)*
	diarr	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diarrhoea *(optional)*
	nausea	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Nausea *(optional)*
	vomit	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Vomiting *(optional)*
	abdopain	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Abdominal pain *(optional)*
	dermato	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rash or other dermatological manifestation of COVID-19 *(optional)*
Laboratory tests (SARS-CoV-2)	swabdate	Date	dd/mm/yyyy	Respiratory specimen collection date
	lab_covtest	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient was tested for SARS-CoV-2 (during hospitalisation)
	lab_covtesttype	Numeric (categorical)	1 = RT-PCR 2 = Serology 3 = Rapid test 4 = Other 8 = Do not know	Type of lab test used
	lab_covtesttype_sp	Text		Specify other type of lab test

	Variable	Type	Values and coding	Definition
Laboratory tests (SARS-CoV-2) *continued*	lab_covid	Numeric (categorical)	0 = Negative 1 = Positive 2 = inconclusive / undetermined 8 = Do not know	Laboratory result: virus type SARS-CoV-2
	lab_ctvalue	Numeric		Specify cycle threshold (Ct) value
	seq	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient sample was sequenced/sent for sequencing
	genetic_group	Text		Laboratory result: genetic group
Prior lab tests	prev_labcovid	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient had a prior positive COVID test
	prev_labcovid_type	Numeric (categorical)	1 = RT-PCR 2 = Serology 3 = Rapid test 4 = Other 8 = Do not know	Type of prior COVID test used (for positive result above)
	prev_labcovid_sp	Text		Specify other type of test used
	prev_labcovid_date	Date		Date of prior positive COVID test
Laboratory tests (SARS-CoV-2) *(optional)*	pcr2	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether a second PCR was done (if first PCR was negative) *(optional)*
	lab_covidpcr2	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Second PCR result for virus type SARS-COV-2 *(optional)*
Laboratory tests (other respiratory viruses)	lab_fluany	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result: any influenza virus type
	lab_flu_type	Numeric (categorical)	1 = influenza A(H1N1) 2 = influenza A(H3N2) 3 = influenza A (untyped) 4 = influenza B/Yamagata 5 = influenza B/Victoria 6 = influenza B (untyped) 8 = Do not know	If positive for influenza, indicate which type and subtype, if known

	Variable	Type	Values and coding	Definition
	lab_mers	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result: virus type MERS-CoV
	lab_othcov	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result: virus type other coronavirus
	resp_virus	Numeric (categorical)	0 = None 1 = RSV 2 = Metapneumovirus 3 = Other respiratory infection 4 = Adenovirus 8 = Do not know	Which other non-influenza, non-coronavirus patient tests positive for
	resp_virus_oth	Text		Specify other respiratory virus
Frailty assessments	frailty_any	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether any type of clinical frailty score was used at admission to assess patient
	frailty_type	Numeric (categorical)	1 = Barthel Index 2 = Clinical Frailty Score (CFS) 3 = Other 8 = Do not know	Indicate which type of clinical frailty score was used
Frailty assessments (continued)	frailty_sp	Text		Specify which other clinical frailty scale was used and the score
	frailty_barthel	Text		Total Barthel score at admission (if used)
	frailty_cfs	Text		CFS score at admission (if used)
	frailty_walk	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Indicate whether patient can walk unaided across the room (if no official clinical scales used above)
	frailty_toilet	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Indicate whether patient can use the toilet independently (if no official clinical scales used above)
	frailty_bathe	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Indicate whether patient can bathe independently (if no official clinical scales used above)

	Variable	Type	Values and coding	Definition
Case definitions and outcomes	severity_vent	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient was ventilated (invasive ventilation) (this classification will be done by re-coding after data collection)
Case definitions and outcomes (continued)	severity_ox	Numeric (categorical)	0 = No 1 = Yes 3 = Do not know	Whether patient received oxygen (e.g. high-flow), but was not intubated/ ventilated (this classification will be done by re-coding after data collection)
	covid	Numeric (categorical)	0 = No 1 = Confirmed 2 = Probable 3 = Other coronavirus 4 = Suspected 8 = Do not know 8 = Do not know	Whether patient is a case of COVID-19 or not (this classification will be done by re-coding after data collection)
	outcome	Numeric (categorical)	1 = died 2 = discharged from hospital 4 = still on treatment 8 = unknown outcome	Indicate the outcome of the patient known at the time of data collection (note: this may be updated later)
	deathcause	Numeric (categorical)	1 = died from COVID-19 2 = died other cause 8 = died unknown cause	Cause of death
Pandemic COVID-19 vaccination	panvacc1dose	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received pandemic COVID-19 vaccination, first dose
	panvacc1date	Date	dd/mm/yyyy	Vaccination date, first dose
	panvacc1type	Text		Type of vaccine (product name)
	panvacc1batch	Text		Vaccine batch number (if known)
	panvacc2dose	Numeric	0 = No 1 = Yes 8 = Do not know 3 = second dose NA	Received pandemic COVID-19 vaccination, second dose
	panvacc2date	Date	dd/mm/yyyy	Vaccination date second dose

	Variable	Type	Values and coding	Definition
Pandemic COVID-19 vaccination (continued)	panvacc2type	Text		Type of vaccine (product name)
	panvacc2batch	Text		Vaccine batch number (if known)
	vacc_stat_ascert	Numeric	1 = vaccination card 2 = vaccine registry 3 = hospital notes 4 = GP notes/verbal 5 = pharmacy record/verbal 6 = patient (verbal) 7 = family of patient (verbal) 8 = do not know 9 = insurance claims record 10 = other	Ascertainment of vaccination status and vaccine product name. Note that GP records may be from direct viewing of the notes or discussion with the GP; similarly the pharmacy record may be viewed directly or this information confirmed by the pharmacist, or by the patient's family. If more than one, select all that apply **This may require two variable fields: one for vaccination status ascertainment, and one for vaccine product determination**
	vacc_stat_ascert_sp	Text		Specify other vaccine status ascertainment
Practice of non-pharmaceutical interventions *(optional)*	risk1_mask	Numeric (categorical)	0 = Never 1 = Sometimes 2 = Usually 3 = Always 8 = Not applicable	Wore mask in public (e.g. in shops, on public transport, in the company of those not living in their home) (optional)
	risk2_handwash	Numeric (categorical)	0 = Never 1 = Sometimes 2 = Usually 3 = Always 8 = Not applicable	Frequently washed hands with soap and water for at least 20 seconds (optional)
	risk3_sanitiser	Numeric (categorical)	0 = Never 1 = Sometimes 2 = Usually 3 = Always 8 = Not applicable	Used hand sanitiser when soap and water not available (optional)
	risk4_socialdist	Numeric (categorical)	0 = Never 1 = Sometimes 2 = Usually 3 = Always 8 = Not applicable	Ensured physical distancing in public (e.g. remaining at least 1 m* away from others: sites to adjust to match national recommendations *(optional)*

8.3. Annex 3: Detailed analysis plan

8.3.1. Data storage and management

Study databases can be submitted to Epiconcept in any format. A country (or study) identifier will be included in each record (e.g. ES for Spain, UK for the United Kingdom), and a hospital code will be included (e.g. a unique number). Data management will follow the basic principles outlined below and in section 3.13.7 (Data management). A country-specific flowchart of exclusions and restrictions will be shared with each of the participating countries. Variables will be recoded and new variables generated. The recoded data will be stored separately from the crude data and recoding will be documented.

Summary and frequency tables and graphic displays of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of respiratory specimen collection before date of onset of symptoms). Any improbable, illegal or missing values should be investigated.

Any subsequent changes to the data will be fully documented and stored separately from the crude database, to ensure reproducibility and transparency of data management.

Missing data

Any missing data will be described. If there is much missing data with no evidence of bias in the missing data, and variables that are considered good predictors of the missing data are available, multiple imputation methods at study level will be used to replace missing values. A sensitivity analysis will be carried out comparing results from the complete case analysis (where records with missing data will be dropped) and the full set analysis (with imputed data).

Data cleaning

Summary and frequency tables as well as visual representations of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies should be carried out (e.g. date of discharge from hospital before date of onset of symptoms). Ideally, these checks could be included as warnings if using an electronic questionnaire, in order to avoid inconsistencies in the data entry. These values will be checked against the questionnaires or queried with the hospitals. Any changes to the data will be documented and stored separately from the crude database. Any recoding of data (e.g. age) will be documented.

- *Countries to specify the data checking and cleaning process*

8.3.2. Pooled analysis outline

Epiconcept will conduct the pooled analysis. The higher sample size for this analysis will provide more power (and precision). Data can be coded as outlined in Annex 1, or a codebook can be provided by the study teams to Epiconcept that includes the variable names, descriptions and coding. Epiconcept will perform additional data cleaning, and will document and share any further data cleaning and analysis with all country coordinators to ensure it can be reproduced.

For the pooled data, interim analyses will be conducted in different periods if appropriate and according to the available sample size. The timing to conduct each interim analysis will depend on the time needed to reach the appropriate sample size. This will depend mainly on the incidence of hospitalisation, COVID-19 incidence, vaccination coverage, the recruitment strategy within hospitals and the number of participating hospitals/services per hospital in each country. The pooled analysis will be carried out in a similar way to the country-specific analysis. Country will be included potentially as a fixed effect or as a random effect in a multilevel model. Statistical heterogeneity between study sites will be determined, using Q-test and the I2 index.(13)

Briefly, cases and controls will be described by baseline characteristics, and uni- and multivariable analyses performed as described in section 3.14.1 for individual analysis.

8.3.3. Bias from pooled estimates

With any multi-centre study, there is always the potential for heterogeneity among sites. With data from a number of different hospitals from different countries being pooled, any bias in the individual studies will influence the pooled estimate. The power of the test for the presence of heterogeneity between individual studies is low if there are few sites/countries. In this case, the test may not be able to detect heterogeneity between them, despite it being present. It is important that heterogeneity is also assessed using qualitative knowledge about differences between studies. Depending on the nature of the bias, the inclusion of biased studies in the pooled estimate could lead to over- or underestimation of the true association between COVID-19 vaccination and the outcome.

There are many conditions which could lead to bias in a single site or hospital. With this new virus, there are new and evolving surveillance systems and strategies in each participating country. There are not only different tests being used, but a variation in the number of tests used to declare an individual negative, for example. Another example is that, when under high pressure (e.g. high volume of patients to be admitted during a peak in the epidemic for any site), it is possible that some hospitals may switch to admitting only suspected COVID-19 patients, while others focus on non-COVID-19 patients. In the event of the former type of hospital being a participating hospital in this study, this could affect the recruitment of controls and result in cases being predominantly recruited from one hospital over another. If a participating site only has one

hospital providing data, this could mean they are only able to provide information on cases. Conversely, if the single participating hospital was designated a non-COVID-19 admitting hospital, this site would only be able to provide information on controls. *These hospitals would not be included in the study.*

To allow for complete assessment of heterogeneity, sites need to document all changes in their COVID-19 surveillance system during the study period.

8.3.4. Pooled analysis plan

8.3.4.1. Descriptive pooled analysis

The proportion of eligible hospitalised cases and controls who accepted to participate in the study will be calculated. The proportion of patients not consenting will be documented, along with reasons for no participation. Patients excluded will be described in a study flowchart.

Cases and controls will be described by baseline characteristics.

The main characteristics of each study will be summarised individually, including:

- Number of hospitals participating and catchment population
- Beginning of vaccination campaigns for pandemic vaccine
  - Beginning of the study
  - End of the study
  - Vaccine product(s) used
  - Estimated vaccine coverage in the country/region by vaccine brand, by target vaccine group
- Number of patients screened
- Number of patients excluded per reasons for exclusion

8.3.4.2. Measure of effect

This study is a case control study (test-negative design). The measure of association is an odds ratio (OR). This can be measured by logistic regression. An OR = 1 indicates no association between an exposure and the outcome. An OR>1 indicates a potential risk factor, an OR < 1 indicates a potential protective factor, noting that the confidence interval around the OR helps with its interpretation.

For vaccination as preventive factor, the CVE can be computed as CVE = (1 – OR)*100. A 95 % confidence interval is computed around the point estimate.

8.3.4.3. Pooled univariable analyses

Baseline characteristics of cases and controls will be compared using the chi-square test, Fisher's exact test, t-test or the Mann-Whitney test (depending on the nature of the variable and the sample size). The association (OR) between vaccination status and baseline characteristics will be measured for both case and control groups.

8.3.4.4. Stratified analysis

The analysis by vaccine product will be further stratified according to (depending on sample size):

- sex
- age groups, e.g. 0–14 years, 15–49 years, 50–64 years, 65–79 years, 80+ years
- specific chronic conditions (e.g. respiratory, diabetes, obesity)
  - absence, presence of at least one, presence of more than one high-risk condition
- time: this will depend on timing of the pandemic in sites/countries and may just include one period at the start of the study once vaccines are available, and a specified period later on
- swab delay (0–3 days, 4–7 days; 8+ days)
- vaccination delay (<8 days, 8–14 days, >14 days, etc.)
- hospital admission delay (0–4 days, 5–9 days, 10 days +, onset after hospitalisation)
- previous vaccination against influenza and pneumococcal disease
- prior infection with influenza or COVID-19 (prior to hospital admission for SARI)
- current co-infection with influenza or other respiratory viruses
- severity (ICU admission, ventilation/oxygen, death)
- for the various groups of vaccines (if available/applicable), mode of injection (intradermal vs intramuscular)
- use of medications for chronic conditions (e.g. statins)

Virus type-specific outcomes will be used, if available and feasible at the time of analysis.

A sufficient sample size should be planned in order to ensure enough individuals in each stratum for a precise estimate. Effect modification will be assessed comparing the OR across the strata of the potential effect modifiers. Confounding will be assessed by comparing crude and adjusted OR for each potential confounder.

8.3.4.5. Multivariable analysis

A multivariable logistic regression analysis will be conducted to control for negative and positive confounding. Odds ratios and standard errors will be obtained. Variables will be tested for multicollinearity. Interactions will be tested using the likelihood ratio test or Wald's test and will be included in the model if significant at the 5 % level. Factors other than statistical significance (prevalence of exposure, magnitude of

OR) will also be used as criteria for inclusion of a variable or an interaction term. If possible, a variable for sex, age and for onset time should always be included in the model.

Continuous variables

Continuous variables in the COVID-19 datasets include age, time of onset of symptoms, GP visits in the previous 3 months and hospitalisations in the past 12 months. These variables can be coded as categories, e.g. age group, week of symptom onset, etc. However, when coding continuous variables as categories, you may lose information, introduce residual confounding and increase the standard error of your model. Tests will be carried out to see if these variables could be coded as a linear term, polynomial or a spline. In addition, a balance will be sought between simplicity of a model (so a non-expert can understand what is going on), precision and a model that estimates the vaccine effect with the least bias.

Identifying heterogeneity, testing for heterogeneity

Country-specific crude and adjusted ORs and their confidence intervals will be plotted in separate forest plots. Following the core protocol minimises heterogeneity between studies. However, adherence to the protocol and study design and study quality characteristics will also be checked. Other study site characteristics will be assessed where feasible, such as types of circulating virus, information on health care use, organisation of the vaccination campaign. Then a qualitative decision will be taken if one or more studies are substantially different from the other and should be excluded from the pooled analysis.
Statistical heterogeneity between studies will be tested using Q-test and the I^2 index (see boxes for formulae below). The Q statistic follows a Chi^2 distribution (with k-1 degrees of freedom). The Q-test reports presence or absence of heterogeneity, while the I^2 index (based on the Q-statistic) quantifies the extent of the heterogeneity. According to the Higgens and Thompson classification, an I^2 index of around 25% indicates low, 50% indicates medium and 75% indicated high heterogeneity between studies.

$$Q = \sum w_i \ (\log(OR_i) - \log(OR_F))^2$$

Where:

$w_i = 1/v_i$

v_i is the inverse variance of the estimated log odds ratio of study *i*

$$\log(OR_F) = \frac{\sum w_i \times \log(OR_i)}{\sum w_i}$$

$$I^2 = \frac{Q-(k-1)}{Q} \times 100\% \quad \text{for } Q > (k-1)$$

$$I^2 = 0 \quad \text{for } Q \leq (k-1)$$

Formulae are given here for completeness, in practice these measures are automatically calculated by many statistical software packages as part of the meta-analysis commands.

8.3.4.6. One-stage pooled analysis approach

If sample sizes are too small to measure vaccine effectiveness controlling for all potential confounders for each individual study site, a 1-stage pooled approach will be used for analysis.

Individual study data will be pooled into one dataset and analysed as a 1-stage model with study site as a fixed effect. This could provide a large enough sample size to obtain (for example) an estimate of CVE early in the study with reasonable precision. The results of this analysis should be interpreted with caution, though, as it assumes not only that the underlying true exposure effect is the same in all studies, but also that the association of all covariates with the outcome is the same in all studies.

Formal tests of interaction between study site and covariates will be carried out to determine if the effect of each covariate differs across studies, to test the assumptions of the 1-stage pooled fixed effect analysis.

The significance of interaction terms are themselves influenced by sample size and should be interpreted also with caution. Particular care needs to be taken if heterogeneity is found between study sites when using a 1-stage fixed effects approach (see above section). Reasons for heterogeneity need to be thoroughly investigated and the assumptions underlying the 1-stage pooling approach need to be revisited.

8.3.4.7. Controlling for hospital effect

Primary analysis will be carried out using simple logistic regression to obtain the individual study estimates. However, there could be an effect of the hospital that is related both to the exposure (propensity to vaccinate) and the outcome (in terms of swabbing behaviour). To adjust for this cluster effect, a multi-level logistic regression with each hospital as a random effect will be carried out when using 1-stage pooled analysis.

Multi-level logistic regression can also be carried out for each individual study with hospital as a random effect. Then the 2-stage model as outlined above will be used to obtain a summary CVE measure, using these estimates.

The same applies to stratified analyses. The point estimates and confidence intervals from the multi-level and simple logistic regression will be compared in a sensitivity analysis.

8.3.4.8. Two-stage pooled analysis approach

If adequate sample size by study is achieved to obtain an adjusted OR, then a 2-stage approach to pooled analysis will be taken.

Country-specific adjusted ORs and standard errors for the effect of COVID-19 vaccination obtained from the individual studies, will be combined in a model that incorporates random effects of the studies, to account for unmeasured country- and hospital-specific factors that differ between countries.

The country-specific exposure-disease effects (ORs) are then weighted by the inverse of their marginal variances. The marginal variance is the sum of the individual study-specific variances and the variance of the random study effects (τ2). This will give the pooled odds ratio and standard error.

$$\log(OR_R) = \frac{\sum w_i * \times \log(ORi)}{\sum w_i *}$$

$$wi* = \frac{1}{vi + \tau^2}$$

The country-specific ORs and their confidence intervals, along with the pooled OR, will be presented graphically in a forest plot. This model will also be compared against a 2-stage analysis with fixed study effects, to assess the effects of model assumptions.

If, despite the common protocol, covariates were not uniformly collected in the different studies, then an analysis will be carried out excluding certain studies and a comparison to the analysis including all studies will be made. In a different scenario, analyses can also be carried out excluding certain study participants for whom variables were collected differently.

8.3.5. Further analyses

Where sample size allows, further analyses will be carried out. These include:

- CVE at different time points in calendar time, e.g. CVE by week or group of weeks (e.g. CVE for weeks 2–3, 4–5, 6–7, etc.)

- CVE by time since vaccination. Time since vaccination can be calculated by subtracting the date of vaccination from the date of onset. Time since vaccination can then be modelled as a continuous variable, including correction for either stable or increased rate of COVID-19 illness over time; cumulative risk of COVID-19 illness
- CVE for patients with previous influenza vaccination (current influenza season) vs no previous influenza vaccination
- If negative CVE is found in some target groups
    - assess possibility of vaccine-mediated enhanced disease (VMED), which could manifest as negative CVE, by comparing severity in vaccinated and unvaccinated patients. Results should show reduced severity among vaccinated patients; findings of increased severity in vaccinated patients could suggest VMED
- As a sensitivity analysis, CVE will be calculated
    - considering those vaccinated <X days before onset of symptoms as unvaccinated (in the main analysis these records will be excluded)
    - including in the control group, SARI patients testing positive for influenza
    - including in the case group, SARI patients testing positive for influenza
    - including in the control group, SARI patients whose influenza vaccine status is unknown
    - using, as a control group, only SARI patients testing positive for at least one non-influenza respiratory virus
    - considering different restrictions according to swabbing delay (e.g. <14 days, <10 days, etc.)
    - considering the sensitivity and specificity of PCR
    - based on assumptions of previous infections
    - excluding participants who received antivirals ≤14 days prior to swabbing
    - excluding all participants with lab-confirmed influenza at any time after COVID-19 onset, to reduce bias
        - this can then be repeated using RSV as a sham outcome (if multiplex results are available for any sites); there should be no association between COVID-19 vaccination and RSV-positivity in the absence of confounding

We can also put time as a variable in the model. As time may be an effect modifier (there may be different CVE at different times of the pandemic), then we can add an interaction term or perform the proposed stratified analysis.

8.3.6. Minimum sample size

Sample sizes may be very small for some sub-analyses. Different criteria can be used to determine whether the sample size is large enough to obtain a valid measure of CVE:

- There are at least 10–15 cases (or controls, whichever is smaller) in the sub-analysis for crude analyses and more for adjusted analyses (e.g. at least 10 for each parameter in the model)
- There are ≥5 records in each cell of the two-by-two table of case and vaccination status
- The precision of the estimate does not span both -200% and 90% (uninformative).

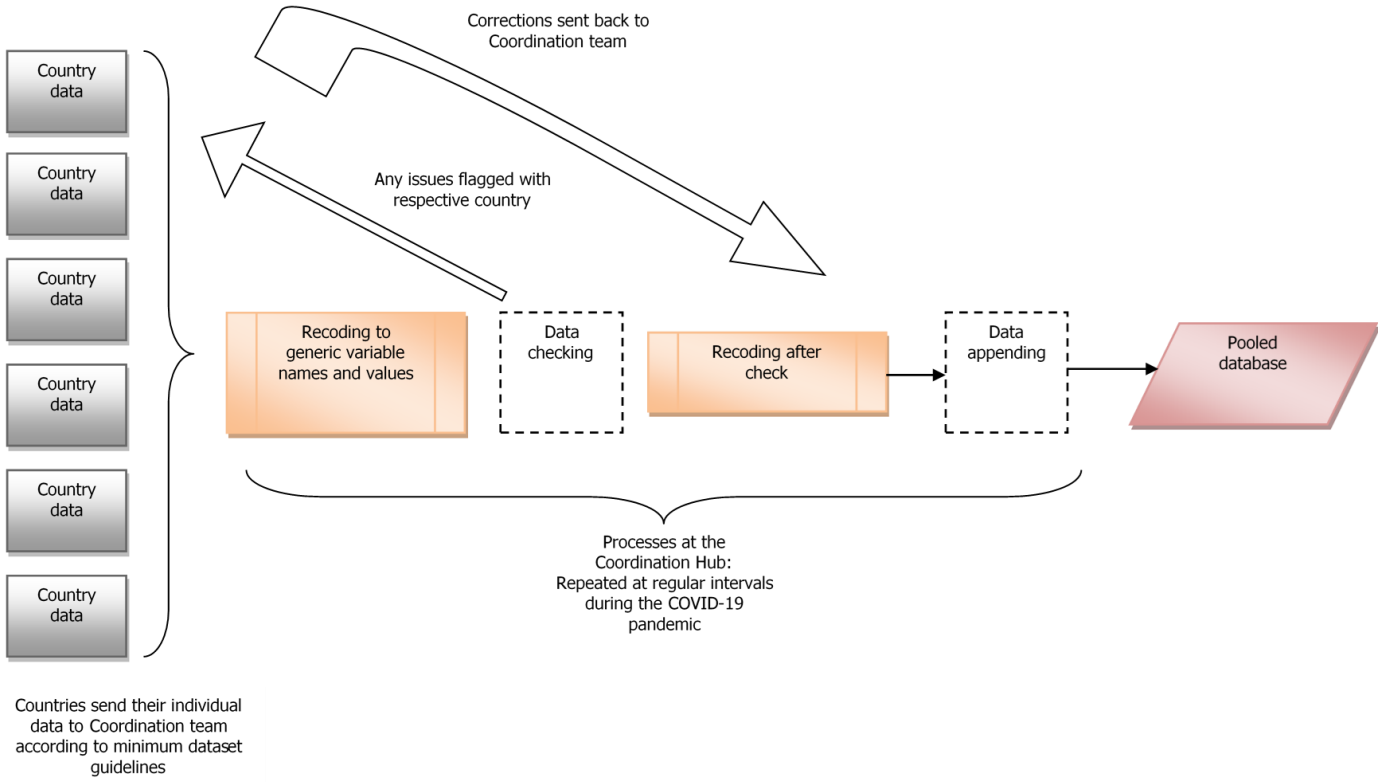
With low sample size, we should consider collapsing categories, modelling continuous variables in a different way (if applicable); sensitivity analyses can be carried out using penalised logistic regression.

8.3.7. Use of propensity scores

To limit the number of co-variables to include in the multivariable model, **if sample size allows,** estimates will be built and adjusted based on propensity scores. Propensity scores can be defined as the conditional probability of receiving the vaccine given a number of observed covariables.

In propensity score matching, a propensity score for vaccination is calculated for cases and controls. Cases and controls are then matched by propensity score and all non-matched patients are discarded. Variables used to calculate the propensity score will include variables related to the vaccination and outcome. Care will be taken to avoid correlation and overmatching.

8.4. Annex 4. Data flow for pooled dataset



8.5. Annex 5: Genetic and antigenic analysis data (example)

The minimum amount of data needed to obtain genetic data from GISAID (sequences of all viruses should be sent to GISAID's open access EpiCoV platform) is country, I-MOVE-COVID-19 ID number and GISAID accession number. Additional information on CT value and selection for characterisation and reasons for not characterising can be additionally collected (see Table 5 below).

Table 5: Example of a data collection form for genetic data.

	Country	I-MOVE-COVID-19 ID number	GISAID accession ID number	Selected for character-isation?	Reasons for not character-ising?	CT value	Type of sample (primary specimen or isolate)
strain 1							
strain 2							

Where not all viruses were attempted to be sequenced, but only a random selection of them, additional information on sampling fraction should be provided, In order to better understand how viruses were selected for sequencing over time. An example can be seen in Table 6 below.

Table 6: Example of documenting outlining how viruses were selected for sequencing over time

Period	First date of period	Last date of period	Sampling fraction	Date used for definition of time unit (onset date, swab date, other)	Comments
1					
2					
Example1	*01/10/2020*	*31/12/2020*	*1*	*Date of onset*	*For example: all specimens were characterised*
Example2	*01/01/2021*	*15/02/2021*	*0.2*	*Date of onset*	*For example: 20% of all specimens were characterised*

8.6. Annex 6: Study-specific annexes

Study specifications for each country should be summarised in this annex. Each study site annex should include:

- description of the hospitals participating in the study (wards involved, bed capacity, catchment population, detailed mode of recruitment including the use of computerised system to identify SARI patients)
- definition of beginning of pandemic
- pandemic (when applicable) vaccines used
- vaccine status ascertainment method
- details on methods for data collection, data entry and data transmission
- list of variables collected (and coding if different from suggested coding)
- data validation procedures
- laboratory issues (laboratory performing tests; tests used: PCR, antigen test, strain characterisation; methods for specimen collection, storage, transport; selection procedures for strain characterisation)
- consent, ethical procedures (oral/written consent; submission to ethics committee)
- human resources needed
- provisions to train hospitals.

8.7. Annex 7: Stata syntax or R scripts

Study sites may request Stata or R syntax/scripts from Epiconcept.

8.8. Annex 8: History of changes to the generic protocol

The broad adaptation and use of this generic protocol led to identifying potential points of improvement. This paragraph aims at listing the changes brought to the protocol throughout its use. Changes are displayed in red text.

8.8.1. Version 04 to versions 05 and 06 (January 2021)

1. This version was updated after the latest review of the Generic WHO/Euro SARI VE protocol version 1 (23 December 2020). This additional reference has been added to the title page.
2. The following notice was added to page 5

   ➢ *Changes in this protocol relative to the previous version are indicated in yellow highlight*

3. Primary objective page 8

   The primary objective will be to measure, for each European participating site country and, for pooled analyses, across all participating sites, the direct effect (effectiveness) of overall and product-specific COVID-19 vaccines against laboratory-confirmed SARS-CoV-2 in hospitalised SARI patients of all ages, in order to provide up-to-date information for the public health response at hospital level by informing prevention measures and highlighting target groups at risk.

4. Secondary objectives page 8

   1. To measure overall and product-specific CVE against laboratory-confirmed SARS-CoV-2 in hospitalised SARI patients by:

5. Secondary objectives page 9

   2. To measure overall and product-specific CVE among SARI patients requiring hospitalisation against

6. Study population pages 9–10

   Version 04

   The study population for the CVE study will therefore consist of individuals of all ages likely to be hospitalised

   Version 05

   The study population for the CVE study will therefore consist of individuals of all ages hospitalised

7. Study period page 10

   Participating hospitals carry out the study throughout the year. However, the study will end if SARS-CoV-2 is no longer circulating in the community (likely to be a 12-month period initially).

8. Outcome page 10

Version 04

The primary outcome of interest will be SARS-CoV-2 in patients of all ages hospitalised with SARI symptoms, laboratory-confirmed by PCR documented either on admission to hospital or within 14 days before admission.

Version 05

The outcome of interest for the primary analysis will be SARS-CoV-2 in patients of all ages hospitalised with SARI symptoms, laboratory-confirmed by PCR documented either on admission to hospital or within 14 days before admission.

9. Exclusions and sensitivity analyses, page 12

   One exclusion has been removed to avoid confusion, as it appears also in the sensitivity analyses:

   - was swabbed more than 7 days after symptom onset (to avoid false negatives; the exact cut off will be determined as more research on this comes in)

   One has been changed:

   - … history of hospitalisation within the 14 days immediately prior to this admission …

   And one sensitivity analysis has been added:

   - excluding those who have received antivirals ≤14 days prior to swabbing (to avoid false negatives; the exact cut-off and types of antivirals will be determined as more research becomes available

10. Control definition page 12

    A control will be defined as a patient hospitalised with SARI symptoms, with a respiratory sample **negative** for SARS-CoV-2 by PCR only within 14 days prior to hospital admission.

    Controls who are negative by PCR but have CT results suggestive of COVID, and those with prior SARS-CoV-2 infection in the 3 months prior to admission, may be excluded as controls in sensitivity analyses (see section 3.5.6 Exclusion criteria below).

11. Exclusions page 12
12. SARI patient identification page 14

    Version 04

    SARI patients should be enrolled and swabbed (if not previously swabbed) within 48 hours of hospital admission,

    Version 05

    SARI patients should be enrolled and swabbed within 48 hours of hospital admission,

13. Definition of vaccination status page 19

    Version 04

- ***Fully vaccinated*** (two-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received both doses** at least 7 days* before onset
- ***Fully vaccinated*** (single-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received one dose** at least 7 days* before onset
- ***Partially vaccinated*** (two-dose vaccine only): to be defined according to vaccine product recommendations, but most likely a patient will be considered partially vaccinated if they have **received one of two doses** at least 7 days* before onset

Version 05

- ***Fully vaccinated*** (two-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received both doses** at least 14 days* before onset
- ***Fully vaccinated*** (single-dose vaccine): to be defined according to vaccine product recommendations, but most likely patients will be considered fully vaccinated if they have **received one dose** at least 14 days* before onset
- ***Partially vaccinated*** (two-dose vaccine only): to be defined according to vaccine product recommendations, but most likely a patient will be considered partially vaccinated if they have **received one of two doses** at least 14 days* before onset

14. Sources of information for vaccination status page 20

- *Each study site to document*
  - *vaccine products used*
  - *places of vaccination (GPs, specific vaccination centres, etc.)*
  - *precise mode of vaccine ascertainment (self-report, card, registry, etc.)*
  - *if no precise dates of vaccination collected, the variable allowing a patient to be defined as vaccinated or unvaccinated*
  - *vaccine status ascertainment validation*

15. Table 2 page 21

immunodeficiency (including HIV) and organ transplant

- *Each study site to update the mandatory list to include pre-existing conditions defining target groups for vaccination in your country*

- *Each study site to define the list of chronic conditions to be included and state whether they are used to define target groups for vaccination, as well as any pre-existing medications being taken, and describe what the sources of information for these will be*

16. Healthcare worker definition page 24

The definition of a healthcare worker for the purposes of this study is anyone working (paid or on a regular voluntary basis) in healthcare who has contact with any type of patient) during his/her work. This includes: doctors, nurses, emergency medical personnel, medical and nursing students having contact with patients, as well as porters and cleaners. It also includes anybody working with resident contact in a nursing/residential home for the elderly. Study sites should collect information on healthcare worker status where possible.

17. Questionnaire page 52

New codes added for target group

0 = not in a target group

8 = do not know

New variables added for ICU admission and discharge dates: icuadmitdate and icudisdate

18. Questionnaire page 53

New variables added for other essential workers: essential_worker and esswork_sp

19. Questionnaire page 61

New variables added for previous positive COVID test: prev_labcovid, prev_labcovid_type and prev_labcovid_date

20. Questionnaire page 62

Two variables severity_ox and severity_vent now shaded pink as they can be created through re-coding from other variables and do not need to be entered separately

21. Further analyses page 70

An additional sensitivity analysis has been included

- Excluding participants who received antivirals ≤ 14 days prior to swabbing

8.8.2. Version 06 to version 07 (February 2021)

1. This version was updated after the latest review of the WHO/Euro SARI COVID VE questionnaire v7 (03 February 2021). This additional reference has been added to the title page.
2. Abbreviations page 4

Two new abbreviations have been added: Ct (cycle threshold value) and SES (socioeconomic status).

3. Study population and study period page 9
   Two typing errors have been corrected to COVID-19 and SARS-CoV-2

4. Case definitions page 11
   Paragraph on SARI patients with onset 14 days prior has been updated to:

SARI patients with onset of symptoms within 14 days prior to hospital admission will be included in the study. Note that hospitals already participating in SARI surveillance systems should not modify the SARI inclusion criteria for surveillance. However, for the CVE analysis, we will only include those patients with onset of symptoms 14 days prior to hospital admission.

In addition the following footnote has been added for SARI controls (in Section 3.5.5): Note: controls **must** have a negative PCR at admission.

5. Exclusion criteria page 12
   - excluding those who are a current control (SARS-CoV-2 negative) but were positive by PCR or serology in the previous year before the current hospitalisation or reported clinically confirmed COVID-19, so as to determine the best cut-off period for having had a previous positive test during the previous year vs "any previous positive test" regardless of date

6. Recruitment strategies page 13
   For hospitals with electronic patient records and/or diagnosis codes commonly displayed, SARI-related ICD codes (or other codes used for SARI surveillance) will be sought. Patients admitted with any of the ICD codes listed in Table 1 will be approached; those meeting the SARI case definition and the inclusion criteria will be invited to be part of the study and sign informed consent (Figure 1).

7. Medication status for chronic conditions page 21
   The use of specific types of chronic medications prior to vaccination or illness may modify or confound the effect of the vaccine.

      - onset of SARI symptoms or

8. Pregnancy status page 22
   3.10.6 Pregnancy status

   Pregnancy status will be collected and coded for women aged 15–55 years as follows: pregnant (yes/no/do not know), and if yes: trimester (1/2/3/do not know)

9. Smoking history page 22

Smoking history will be collected and coded as follows: never-smoker, former smoker (stopped smoking at least one year before inclusion in the study), current smoker (or stopped in the past year).

10. Practice of NPIs, page 25

3.10.17 Practice of non-pharmaceutical interventions (optional)

Individuals not practising routine use of NPIs will be at greater risk of infection, and may also be less likely to access vaccination services. Stratifying by NPI use, if collected, will allow comparison of CVE between those regularly using and not using these key NPIs. Key NPIs proposed are: use of a mask in public, frequently washed hands with soap and water for at least 20 seconds, used sanitiser when soap and water were unavailable, ensured physical distancing in public (remaining at least 1 m away from others), with the following options: always, usually, sometimes, never, not applicable.

- *Each study site to decide on the best way to collect NPI use in their population and to describe the measure(s) used to capture use of NPIs. In particular it is important that these questions are asked without judgement*
- *Sites to adjust distance etc. to match national recommendations*

11. Collected information page 28
    - other symptoms

12. Practice of non-pharmaceutical interventions (optional) page 29
    - non-pharmaceutical interventions *(optional)* (section 3.10.17)

13. Clinical characteristics (symptoms and markers of severity) page 31

The following three symptoms have been associated with COVID-19 illness and are part of the ECDC COVID-19 case definition. They should also be collected where possible:

It would also be helpful to collect whether these symptoms appeared with **sudden onset.**

14. Case definition (minimum key variables) page 32

As a minimum, we need to collect data on the symptoms required for the WHO or ECDC case definitions. The following variables are imperative for application of the WHO SARI case definition:

- fever or feverishness
    - if fever: measured fever (with temperature), or feverishness
- cough
- onset date
- admission date

The following additional variables will also be imperative if the study site is using the ECDC case definition:

- shortness of breath
- sudden onset
- anosmia
- ageusia
- dysgeusia

15. Data management page 33

There are two methods for data transfer for the COVID-19 VE study:

16. Tables 4 and 5 titles pages 35 and 36

Table 4: Example of descriptive table for cases and controls; I-MOVE-COVID-19 hospital-based vaccine effectiveness study, Europe, 2021.

Table 5: Example of table showing vaccine effectiveness against COVID-19 adjusted for various covariables by sex and age group, hospital-based I-MOVE-COVID-19 vaccine effectiveness study, Europe, 2021.

17. Dissemination of results page 38

The enrolment of COVID-19 cases will be regularly updated by each study coordinator on a website developed for the multicentre study. Initial CVE estimates will be disseminated as soon as possible, in the second quarter of 2021, with a later CVE in the third quarter of 2021. (Note that this may be revised depending on how the pandemic progresses.)

18. Representativeness of subjects included in the study page 44

In some sites (those where the CVE study is not part of routine SARI surveillance), very severely ill patients will not be able to give informed consent. If this is the case, some very severe cases may not be included.

19. Questionnaire page 51

Criterion added for current smoker: (or stopped in the past year)

Criterion added for pregnancy: (for women aged 15–55 years) and new variable added for trimester of pregnancy: trimester

20. Questionnaire page 58

Criterion added for COVID lab test variable (lab_covtest): (during hospitalisation)

21. Questionnaire page 59

New category added for COVID lab test result variable (lab_covid): 2 = inconclusive/undetermined

New variable added for COVID lab test cycle threshold (Ct) value: lab_ctvalue

New variable added for influenza lab test results (to indicate type and subtype, where known): lab_flu_type

22. Questionnaire page 60

New category added for non-influenza lab test results: 4 = Adenovirus

New variable added for specifying other respiratory virus found: resp_virus_oth

Clarification added for variable frailty_sp, which now reads:

Specify which other clinical frailty scale was used and the score

New variables added for frailty, to be used if no official clinical frailty scales were used: frailty_walk, frailty_toilet, frailty_bathe

23. Questionnaire page 61

The variables for pandemic vaccine have been changed so that there is a completely separate variable for second dose and each dose has four associated variables: dose number (panvacc1dose, panvacc2dose), date (panvacc1date, panvacc2date), type/product name (panvacc1type, panvacc2type) and batch number, if known (panvacc1batch, panvacc2batch).

24. Questionnaire page 62

New variable added for specifying other vaccine status ascertainment: vacc_stat_ascert_sp

New (optional) variables added for non-pharmaceutical interventions: risk1_mask risk2_handwash risk3_sanitiser risk4_socialdist