COVID-19 vaccine effectiveness against severe acute respiratory infections (SARI) hospitalisations associated with laboratory-confirmed SARS-CoV-2

WHO-AFRO Guidance Document

Based on the Draft WHO/EURO guidance (version January 2021)

Version:	1
Date:	28 April-2021

Table of contents

1 Background 5

2 Objectives 8

2.1 Primary objective 8
2.2 Secondary objectives 8

3 Methods 8

3.1 Study design 8
3.2 SARI surveillance population 9
3.3 Surveillance period for CVE analysis 9
3.4 Outcome 9
3.5 SARI Case definition 10
3.5.1 SARI Case definition 10
3.5.2 Inclusion criteria 10
3.5.3 Exclusion criteria 10
3.6 SARI patient identification – algorithm for patient inclusion 10
3.6.1 Recruitment strategies 10
3.7 Laboratory methods 12
3.8 Case and control definitions 13
3.8.1 SARI patients confirmed as COVID-19 (confirmed cases) 13
3.8.2 SARI who are negative for SARS-CoV-2 (primary controls) 13
3.9 Additional information to collect from SARI patients 13
3.9.1 COVID-19 Vaccination history 13
3.9.1.1 COVID-19 vaccination status ascertainment 13
3.9.1.2 Definition of COVID-19 vaccination status 14
3.9.2 Effect modifiers and confounders 14
3.9.2.1 Pre-existing chronic conditions 14
3.9.2.2 Health worker status 15
3.9.2.3 Previous SARS-CoV-2 infection 15
3.9.2.4 Hospitalizations and ambulatory visits in the year prior to hospitalization (optional) 15
3.9.2.5 Previous influenza and pneumococcal vaccinations (optional) 16
3.9.2.6 Functional impairment/frailty (optional) 16
3.9.2.7 Ethnicity (optional) 16
3.9.2.8 Medications for chronic condition(s) (optional) 16
3.9.2.9 Behaviour related to non-pharmaceutical interventions (optional) 16
3.9.2.10 Socioeconomic status or deprivation (optional) 16
3.9.3 Summary of main categories 17
3.10 Ethical considerations 19
3.11 Data management 19
3.12 Data analysis 19

3.12.1 Relevant Definitions for analysis 19
3.12.2 Individual level analysis 19
3.12.2.1 Output tables presenting CVE estimates 21
4 Limitations 22
4.1 Potential biases 22
4.1.1 Negative confounding 22
4.1.2 Positive confounding 22
4.1.3 Unmeasured confounding 22
4.1.4 Previous infection in cases or control; inclusion of asymptomatic controls 23
4.1.5 Validation of exposure 23
4.1.6 Other potential biases 23
4.2 Representativeness of subjects 23
5 Community Engagement 24
6 Annexes 25
6.1 Annex 1: Identifying SARI patients using relevant ICD codes 25
6.2 Annex 2: Other control groups (optional) 27
6.3 Annex 3: List of ICD-9 and ICD-10 codes for pre-existing chronic conditions 28
6.4 Annex 4: Example of a list of variables, definitions and coding; hospital-based COVID-19 vaccine effectiveness 31
6.5 Annex 5. Pooling data 43
6.5.1 Data storage and management 43
6.5.1.1 Missing data 43
6.5.1.2 Data cleaning 43
6.5.2 Pooled analysis outline 44
6.5.3 Bias from pooled estimates 44
6.5.4 Pooled analysis plan 45
6.5.4.1 Descriptive pooled analysis 45
6.5.5 Measure of effect 45
6.5.5.1 Pooled univariable analyses 45
6.5.5.2 Stratified analysis 45
6.5.5.3 Multivariable analysis 46
6.5.5.4 Continuous variables 46
6.5.5.5 Identifying heterogeneity, testing for heterogeneity 47
6.5.5.6 One-stage pooled analysis approach 48
6.5.5.7 Controlling for hospital effect 48
6.5.5.8 Minimum sample size 48
6.5.6 Further analyses 49
6.5.7 Use of propensity scores 50
6.5.8 Data flow for pooled dataset 51
6.6 Annex 6: Additional information for pooled analysis 52

Acknowledgements

The WHO Regional Office for Africa (AFRO) led the adaptation of this document from the WHO/EURO Guidance document (version January 2021) with Hannah Lewis (WHO HQ), Marta Valenciano (Epiconcept), Isabel Bergeri (WHO HQ) and Jason Mwenda (WHO AFRO). WHO AFRO would like to thank WHO EURO (particularly Mark Katz, Pernille Jorgensen, and Richard Pebody) and the members of the original review committee. In addition, thanks to Robert S Heyderman (University College London), Donnie Mategula, Kondwani Jambo and Latif O Ndeketa (Malawi-Liverpool-Wellcome Trust Clinical Research Programme) for their review.

Abbreviations

COVID-19	Coronavirus disease 2019
CVE	COVID-19 vaccine effectiveness
ICD	International classification of diseases
OR	Odds ratio
RT- PCR	Real-time polymerase chain reaction
SARI	Severe acute respiratory infection
SARS-CoV-2	Severe acute respiratory syndrome – coronavirus 2
TND	Test-negative design
VE	Vaccine effectiveness
WHO	World Health Organization

1 Background

Since the emergence of the novel severe acute respiratory syndrome – coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), in late 2019, the COVID-19 pandemic has, as of 27 April 2021, resulted in more than 147 million cases and 3 million deaths worldwide[1].

After the 2009 influenza A(H1N1) pandemic, a number of Member States implemented sentinel surveillance for severe acute respiratory infections (SARI) for influenza. In these surveillance systems, case-based data (on all or a subset of patients) is systematically collected to monitor the severity and the burden of disease, identify viruses associated with severe clinical presentations, and determine risk factors for severe illness.

WHO has advised that countries with existing hospital-based sentinel influenza surveillance adapt these systems to also monitor severe SARS-CoV-2 cases[2] including, if feasible, collecting data that allows for measuring COVID-19 Vaccine Effectiveness (VE). **Systematic testing of SARI patients for SARS-CoV-2 and collection of individual patient data (including on COVID-19 vaccination: number of doses, date, product) in hospitals as part of enhanced routine surveillance allows for the rapid evaluation of any pandemic COVID-19 vaccine post-licensure.** Stakeholders of a SARI surveillance system that can evaluate COVID-19 VE include ministries of health, public health practitioners, clinicians, laboratories, and international public health organizations.

[1] WHO Coronavirus Disease Dashboard. Available at: https://covid19.who.int/?gclid=CjwKCAiAgJWABhArEiwAmNVTB_X_NfQL7xLVFp6Bmat80f0SYb2GvJZFUJCMwpPi0BaEzcj30O6Q0hoCWnMQAvD_BwE (accessed 27 April 2021)

[2] World Health Organization. Maintaining surveillance of influenza and monitoring SARS-CoV-2 – adapting Global Influenza surveillance and Response System (GISRS) and sentinel systems during the COVID-19 pandemic. Available at: https://www.who.int/publications/i/item/maintaining-surveillance-of-influenza-and-monitoring-sars-cov-2-adapting-global-influenza-surveillance-and-response-system-(gisrs)-and-sentinel-systems-during-the-covid-19-pandemic (accessed 27 April 2021)

Evaluating the performance COVID-19 vaccines post-licensure is critical as a number of factors can impact real-world VE, including transportation and storage conditions, how vaccines are administered, advanced age, presence of underlying medical conditions, and previous SARS-CoV-2 infection. In addition, post-licensure evaluations of pandemic vaccines will allow public health authorities to a) understand the duration of protection of vaccines and thus the need (and frequency) for re-vaccination, b) estimate the level of protection against severe disease and death, c) assess the relative effectiveness of different vaccine types and of single doses, and d) evaluate VE of new emerging virus variants.

This document describes how existing hospital-based surveillance systems for SARI can be adapted to enhance data collection to inform estimates of vaccine effectiveness against COVID-19 in persons of all ages. It outlines in detail methods for collecting and analysing data on vaccinated and unvaccinated patients based on the "test negative design" (TND). The methods are aligned with WHO interim guidance [3,4]. Should a country intend to set-up a new SARI surveillance site, this protocol should be used in conjunction with relevant global and regional surveillance guidance.[5678]

Since this document only outlines general principles, country-specific guidelines related to COVID-19 SARI surveillance to allow estimation of VE should be detailed in national guidelines. SARI surveillance work being conducted on vaccine effectiveness should be in accordance with the WHO guidelines regarding the ethical issues in public health surveillance.[9] Formal ethical committee approvals or waivers should be

[3] World Health Organization. COVID-19 vaccine effectiveness against hospitalised SARI patients with laboratory-confirmed SARS-CoV-2 Generic WHO/Euro protocol (under development)

[4]World Health Organization. Interim guidance: Evaluation of COVID-19 Vaccine Effectiveness. March 2021. Available at: https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccine_effectiveness-measurement-2021.1 (accessed 27 April 2021)

[5] World Health Organization. Maintaining surveillance of influenza and monitoring SARS-CoV-2 – adapting Global Influenza surveillance and Response System (GISRS) and sentinel systems during the COVID-19 pandemic. Available at: https://www.who.int/publications/i/item/maintaining-surveillance-of-influenza-and-monitoring-sars-cov-2-adapting-global-influenza-surveillance-and-response-system-(gisrs)-and-sentinel-systems-during-the-covid-19-pandemic (accessed 27 April 2021)

[6] World Health Organization. Global epidemiological Standards for Influenza. 2013. Available at: https://www.who.int/influenza/resources/documents/WHO_Epidemiological_Influenza_Surveillance_Standards_2014.pdf (accessed 27 April 2021)

[7] World Health Organization. WHO guidance for surveillance during an influenza pandemic. 2017 Update. https://www.who.int/influenza/preparedness/pandemic/WHO_Guidance_for_surveillance_during_an_influenza_pandemic_082017.pdf (accessed 27 April 2021)

[8] World Health Organization Regional Office for Africa. Technical Guidelines for Integrated Disease Surveillance and Response in the African Region: Third edition. 2019. Available at: *https://www.afro.who.int/publications/technical-guidelines-integrated-disease-surveillance-and-response-african-region-third* (accessed 27 April 2021)

[9] WHO guidelines on ethical issues in public health surveillance. World Health Organization. 2017. Available at: https://apps.who.int/iris/bitstream/10665/255721/1/9789241512657-eng.pdf?ua=1 (accessed 27 April 2021)

obtained similar to how they were obtained for existing SARI surveillance systems. Surveillance procedures should protect the interests of patients and mitigate any foreseeable risks and harms to patients.

2 Objectives

2.1 Primary objective

The primary objective is to measure overall and product-specific COVID-19 vaccine effectiveness (CVE) against laboratory-confirmed SARS-CoV-2 in hospitalised SARI patients belonging to the target group(s) for COVID-19 vaccination.

2.2 Secondary objectives

- To estimate overall and product-specific CVE against laboratory-confirmed SARS-CoV-2 requiring hospitalisation among SARI patients in vaccination target groups by:
    - age group
    - sex
    - risk group (e.g. specific chronic conditions; pregnancy)
    - time since vaccination
    - in persons who have received one dose of vaccine compared to those who have received two doses (where there is a two-dose regimen)
    - specific SARS-CoV-2 genetic variant(s)

- To estimate CVE for more severe outcomes (e.g. oxygen therapy, ICU admission, or in-hospital mortality, see below under 3.4)
- To estimate CVE by presence of chronic conditions
- When applicable and depending on availability of these data in the participating country, to identify potential factors that may modify CVE: influenza vaccination, living in a long-term care facility (LTCFs), receiving statins or other long-term medications and higher risk of exposure SARS-CoV-2.

3 Methods

3.1 Study design

- Case-control, test-negative design (TND) among hospitals recruiting Severe Acute Respiratory Infection patients
- The study design is a test-negative, case–control design, which has been used in the past 10 years for estimating annual influenza VE. The principle behind this design is to evaluate SARS-CoV-2 laboratory results among persons who meet the standard SARI case definition and categorize those who test positive as “cases” and those with a negative test results as

"controls". Where possible, CVE analysis is based on information gathered via existing SARI surveillance systems.

3.2 SARI surveillance population

- The surveillance population will consist of individuals of all ages who are hospitalised with SARI symptoms in a hospital that is part of the SARI surveillance network
- The CVE analysis should only include SARI patients who belong to a country's target group(s) for COVID vaccination and have no contra-indications to COVID-19 vaccination.

3.3 Surveillance period for CVE analysis

The relevant surveillance period should begin when COVID-19 vaccines are available in the country. It should continue until transmission levels decline to community transmission level 1 (CT1)[10] for a sustained period. Due to the constantly evolving COVID-19 situation (globally and locally) and the unpredictability of the intensity of transmission at the time when vaccines become available, it is recommended to plan for at least a 6-month surveillance period, although this period would ideally be longer.

3.4 Outcome

The outcome of interest for the primary analysis will be SARS-CoV-2 detection in patients of all ages, eligible for vaccination and hospitalised with SARI symptoms. SARS-CoV-2 infection should be laboratory-confirmed by PCR documented either on admission to hospital or within 10 days before admission.

Secondary outcomes include genetic variants of SARS-CoV-2 in hospitalised SARI patients of all ages.

Additional outcomes include markers of severity of disease during hospitalization (according to what is feasible in country setting), including: length of stay (LOS), oxygen therapy, intensive care unit (ICU) admission, mechanical ventilation, in-hospital death, clinical signs of pneumonia, severe respiratory rate > 30 breaths/min, severe respiratory distress, acute respiratory distress syndrome (ARDS), oxygen saturation <90% on room air, sepsis and septic shock.

[10] CT1: defined as a low incidence of locally acquired widely dispersed cases detected in the past 14 days

3.5 SARI Case definition

3.5.1 SARI Case definition

A SARI patient will be defined using the WHO SARI case definition[11] as follows:

- a hospitalised person with acute respiratory infection, with
- a history of fever or measured fever of ≥ 38 C°
- **and** cough
- with symptom onset within the last 10 days.

Where a minimum of 24 hours in-hospital is required to be considered hospitalized.

3.5.2 Inclusion criteria

- Meets the SARI case definition
- Part of a target group the COVID-19 vaccine on the date of hospital admission
- No contraindications for COVID vaccination

3.5.3 Exclusion criteria

SARI patients will not be enrolled in the CVE analysis if one of more of the following conditions are met:

- has a contraindication for the COVID-19 vaccine
- is ineligible for COVID-19 vaccine at the time of hospital admission (i.e. does not belong to a vaccine target group)
- has a history of hospitalisation within the 14 days prior to this admission (including transfers from another hospital)

3.6 SARI patient identification – algorithm for patient inclusion

SARI patients will be identified among patients hospitalised for at least 24 hours in one of the participating hospitals, according to current country protocols for SARI sentinel surveillance. Patients who meet the inclusion criteria should be enrolled and swabbed at admission or within 48 hours of hospital admission.

3.6.1 Recruitment strategies

Recruitment strategies should follow current recruitment strategies for existing SARI surveillance. In one approach, surveillance team members, which could include doctors, nurses or other health professionals, can systematically screen hospitalised patients (in participating wards) which meet the SARI case definition

[11]WHO . WHO Surveillance definitions for ILI and SARI. Available at: https://www.who.int/influenza/surveillance_monitoring/ili_sari_surveillance_case_definition/en/ (accessed 27 April 2021)

to participate in the surveillance (Figure 1). Like in routine SARI surveillance, it should be made clear to patients which aspects of data collection are related to surveillance and which aspects are related to their clinical care in the hospital.

For SARI surveillance systems that include electronic patient records where ICD codes are routinely used, potential SARI patients can be identified using relevant ICD codes (see **Annex 1**). Laboratory registries can also be used to identify SARI patients with a respiratory sample tested.

Any patient-related information that is collected from hospital medical records, vaccine registries, or other sources should be collected in accordance with procedures outlined and approved in the protocol approved or granted waiver by the local ethical review committee, in accordance with WHO guidelines on ethical issues in public health surveillance.

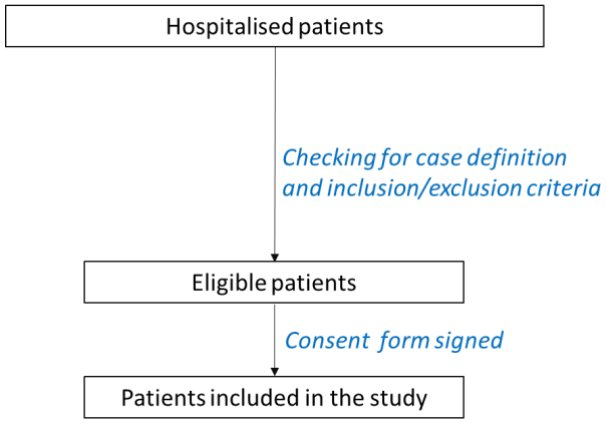


Figure 1: Proposed inclusion algorithm for hospitals/services systematic screening of all admitted patients.

Following local procedures for SARI sentinel surveillance, patients meeting the SARI case definition will be asked to consent to participate in the SARI surveillance. Patients who meet the SARI case definition should have a nasopharyngeal and/or oropharyngeal (throat) specimen taken for SARS-CoV-2 testing. In participating SARI surveillance hospitals, respiratory samples may be routinely collected from all hospitalized patients admitted with respiratory symptoms and tested for SARS-CoV-2, and therefore additional specimens for surveillance purposes will not need to be collected.

If resources are limited, SARI patients can be identified only on certain days of the week (e.g. recruitment of patients every second day, or only on 2 days of each week). Specific approaches to SARI patient recruitment should be planned in advance and well documented. Convenience sampling is not recommended as this can introduce bias.

3.7 Laboratory methods

As is the practice for routine SARI surveillance systems, trained health personnel should collect respiratory specimens (see Section 4.4) from all eligible patients, respecting safety standards for COVID-19 and following WHO biosafety guidelines.[12,13]

Samples should be tested by RT-PCR for SARS-CoV-2.

For countries with sequencing (or other genomic characterization) capacity, according to national testing guidelines and capacity, a systematic sample of PCR-positive samples should undergo genetic sequencing. If only a subset of the sample is sequenced, efforts should be made to ensure that these samples are representative of the SARI cases included in the surveillance (irrespective of vaccination status or clinical presentation).

Genetic sequences, if sequencing is performed, should also be uploaded to an open access epidemic and pandemic virus data sharing platform. Processed genetic information, e.g. name of genetic clade, can also be included within the epidemiological database.

Biological materials and related data should only be collected and stored in collaboration with hospitals and local health authorities according to current procedures for SARI/SARS-CoV-2 surveillance policies.

Most respiratory samples that will be collected for SARI surveillance to inform vaccine effectiveness estimates will be done so as part of routine clinical practice. It should be explained that management of these samples should be in accordance with hospital policies on management of clinical samples. For all respiratory samples that are collected for surveillance and not for clinical purposes, an explanation should be given to the participant about how these samples should be managed; policies should be in accordance with existing protocols for management of samples related to SARI surveillance.

Results of PCR tests for SARS-CoV-2 should be shared with participants as soon as they are available. Because the surveillance is being carried out within hospitals, dissemination of results should be carried out in accordance with hospital policies.

---

[12]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'.

[13]World Health Organization. Laboratory Biosafety Manual, 4th Edition. Available at: https://www.who.int/publications/i/item/9789240011311

3.8 Case and control definitions

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

A confirmed COVID-19 case will be defined as a hospitalised patient that meets the SARI case definition (incl. inclusion/exclusion criteria, see 3.5), with a respiratory sample positive for SARS-CoV-2 by PCR only, either on admission to hospital or documented within 10 days prior to hospital admission.

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

The primary control group will be defined as hospitalised patients that meet the SARI case definition (incl. inclusion/exclusion criteria, see 3.5), with a respiratory sample **negative** for SARS-CoV-2 by PCR at the time of hospital admission (or within 48 hours of admission and with no previous positive test).

During periods of low circulation of respiratory viruses, the number of SARI patients testing negative for SARS-CoV-2 may be limited. Therefore, some sites may wish to include other control groups, either in addition to SARI controls or as an alternative, which must be recruited throughout the whole study period. Options for these controls are outlined in **Annex 2**.

3.9 Additional information to collect from SARI patients

Data should be collected using a standardised questionnaire/data collection form. Existing SARI questionnaires should be adapted to collect the additional information outlined below. **Of importance is the inclusion of vaccination history, effect modifiers/confounders and key dates (see below)**. The questionnaire should be adapted to the country setting.

3.9.1 COVID-19 Vaccination history

3.9.1.1 COVID-19 vaccination status ascertainment

For every SARI patient, dates of vaccine administration, type of vaccine and brand name, and batch code for every dose of COVID-19 vaccine should be recorded, along with the source of this information. A documented source of vaccine history is preferred over a verbal report however lack of vaccination will likely only be available by verbal report.

The sources of information for COVID-19 vaccination status and date of vaccination(s) may include:

- vaccination registry (preferred option)
- consultation of the patient's vaccination card or patient's hospital notes
- interview with the patient's primary care clinician
- interview with facility that administrated the vaccine

3.9.1.2 Definition of COVID-19 vaccination status

- An individual will be considered as vaccinated after receiving one dose against COVID-19 with a product-specific vaccine (see section "Vaccination status ascertainment") during the current pandemic if s/he has received one dose of the named vaccine product more than 14 days before SARI symptom onset.
- Fully vaccinated (one-dose vaccine): to be defined according to vaccine product recommendations. Most likely patients will be considered fully vaccinated if they have received one dose at least 14 days before symptom onset.
- Fully vaccinated (two-dose vaccine): to be defined according to vaccine product recommendations. Most likely patients will be considered fully vaccinated if they have received the second (final) dose at least 14 days before symptom onset.
- Partially vaccinated (two-dose vaccine): to be defined according to vaccine product recommendations. Most likely a patient will be considered partially vaccinated if they have received only the first 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.
- Sensitivity analyses evaluating CVE at different periods of time post-vaccination may be conducted.

3.9.2 Effect modifiers and confounders

It is important to collect variables that may be effect modifiers or confounding factors of the CVE estimates. Effect modifiers are variables that define sub-groups in which VE is truly different Confounders are those variables that are related to both COVID-19 and vaccination status, and not in the pathway between vaccination and disease, which can potentially distort the VE estimate. Listed below are variables that maybe potential effect modifiers or confounding factors.

3.9.2.1 Pre-existing chronic conditions

Information about underlying medical conditions which are risk factors for severe Covid-19 infection should be collected (See Table 2 below). This list should be adapted to the country context depending on the risk groups targeted by vaccination. The list of ICD-9 and ICD-10 codes for pre-existing chronic conditions can be found in **Annex 3.**

Table 2. List of underlying medical conditions that have been identified as risk factors for severe COVID-19 infection

Chronic Medical Condition
Cancer
Chronic cardiac disease, excluding hypertension
Hypertension
Chronic kidney disease
Chronic liver disease
Chronic respiratory disease
Asthma
Diabetes
HIV
Immunocompromised, including solid organ transplant
Neurological disease, including cerebrovascular disease
Obesity
Rheumatologic disease
Anemia or other blood disorder

3.9.2.2 Health worker status

Information on health worker status should be collected. Health workers can be defined according to previously published WHO guidance documents[14] or according to national guidelines used to define the target population of health workers eligible for COVID-19 vaccination.

Because essential or frontline workers have been shown to be at increased risk for severe COVID-19 infection, information on essential worker status could be collected as well.

3.9.2.3 Previous SARS-CoV-2 infection

Previous infection with SARS-CoV-2 may result in immunity to subsequent infections. SARI patients should be asked about prior *SARS-CoV-2 infection* (confirmed or suspected) to elucidate possibility of prior infection in order to distinguish between the effect of vaccination and previous infection.

3.9.2.4 Hospitalizations and ambulatory visits in the year prior to hospitalization (optional)

In order to document and control for healthcare-seeking behaviour and the severity of the underlying conditions, information on the number of hospital admissions in the previous 12 months prior to admission to the hospital should be collected.

[14] WHO. World Health Report 2006. Health Workers: A global profile. Available at: https://www.who.int/whr/2006/06_chap1_en.pdf (accessed 27 April 2021)

3.9.2.5 Previous influenza and pneumococcal vaccinations (optional)

When applicable, information on vaccinations received for influenza and pneumococcal disease [(pneumococcal polysaccharide vaccine (PPV) and pneumococcal conjugate vaccine (PCV)] including date of vaccination should be collected, where available.

3.9.2.6 Functional impairment/frailty (optional)

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[15].

3.9.2.7 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. Information on ethnic group could be also collected from SARI patients.

3.9.2.8 Medications for chronic condition(s) (optional)

The SARI patient's medications for chronic diseases (medication name and amount and frequency of dose) in the previous 3 months should be recorded.

3.9.2.9 Behaviour related to non-pharmaceutical interventions (optional)

No or inadequate adherence to standard protective measures (such as social distancing, mask-wearing, or hand hygiene), might influence viral load and thereby disease severity. Individuals who are not consistently observing NPIs may be less likely to demand vaccination against COVID-19, and so they might be over-represented in the unvaccinated population. As a result, they are also more likely to be COVID-19 positive. compliance with physical distancing, mask-wearing and hand hygiene should be documented when possible.

3.9.2.10 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.

[15] Where possible, the Barthel Index should be used: D. T. Wade & C. Collin (1988) The Barthel ADL Index: A standard measure of physical disability? International Disability Studies, 10:2, 64-67, DOI: 10.3109/09638288809164105

3.9.3 Summary of main categories

Collected information falls under the following main categories (see **Annex 4** for example list of variables and definitions):

- Hospital information
  - country, hospital
  - first ward of referral
  - ICU / other ward admission

- Patient characteristics:
  - age
  - sex
  - chronic conditions
  - smoking history *(optional)*
  - pregnancy, including gestation period (weeks)
  - health worker
  - occupation *(optional)*
  - clinical frailty score at admission *(optional)*
  - ethnic group *(optional)*
  - risk-of exposure
    - mask use, hand hygiene, social distancing *(optional)*
    - SES/deprivation *(optional)*

- Clinical information
  - clinical characteristics that comprise the SARI case definition and the suspected COVID-19 case definition[16] should be collected.
  - severity
    - length of stay
    - oxygen use
    - ICU admission
    - invasive ventilation
    - death
    - clinical signs of pneumonia (fever, cough, dyspnoea, fast breathing)
    - respiratory rate > 30 breaths/min *(optional)*
    - severe respiratory distress *(optional)*
    - oxygen saturation <90% on room air *(optional)*
    - acute respiratory distress syndrome (ARDS) *(optional)*
    - sepsis *(optional)*

[16] WHO COVID-19 and SARI case definition. Available at: https://apps.who.int/iris/bitstream/handle/10665/333912/WHO-2019-nCoV-Surveillance_Case_Definition-2020.1-eng.pdf?sequence=1&isAllowed=y (accessed 27 April 2021)

        - septic shock *(optional)*

    - previous clinical symptoms (if no prior tests done)

- Laboratory
    - type of swab (nasopharyngeal, oropharyngeal etc.) *(optional)*
    - type of test (e.g., PCR)
    - results (including information on antigen and genetic analysis, where available)
    - previous positive PCR or antigen test for SARS-CoV-2, if feasible (for sensitivity analyses)
    - sequencing results *(optional)*
    - presence of influenza and other respiratory viruses
    - blood culture result (if taken)

- Underlying chronic conditions, including HIV and obesity
    - use of medications for chronic conditions *(optional)*
    - number of hospitalisations for chronic conditions in the previous 12 months *(optional)*
    - number of primary care clinician visits in the previous 3 months *(optional)*

- Vaccination and antivirals for respiratory infections
    - COVID-19 vaccination including number of doses, date, product
    - current influenza vaccination *(optional)*
    - pneumococcal vaccination status, type of vaccine and either date or year of vaccination *(optional)*
    - antiviral administration *(optional)*

- Functional status or proxy by residence type *(optional)*

- SES/deprivation *(optional)*

3.9.4 Key dates

The following dates should be collected:

- **date of first key symptom onset (**date of onset of first symptom included within SARI case definition)
- **date of vaccination (COVID-19),** (for sites collecting influenza and pneumococcal vaccination status, date of last 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)
- **date of discharge/death** (to allow measurement of length of hospital stay)

3.10 Ethical considerations

Data collection for this vaccine effectiveness analysis will be conducted within the framework of ongoing SARI surveillance, which constitutes routine public health practice. However, approvals or ethical review committee waivers should be sought from the appropriate local ethical review committee(s). Patients should be informed that participation is voluntary and that he/she will be able to withdraw from the study, without justification, at any time during the study and without consequences.

3.11 Data management

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

Data validation, cleaning and verification should be carried out periodically. Specific scripts should be developed for that purpose.

All data management procedures should comply with applicable national laws on data protection to ensure the confidentiality of patients).

3.12 Data analysis

3.12.1 Relevant Definitions for analysis

See **Annex 4** for an example of a list of variables, definitions and coding for hospital-based COVID-19 vaccine effectiveness.

3.12.2 Individual level analysis

Cases and controls will first be described by baseline characteristics. Patients will be described according to:

- sex
- age group
- occupation (health worker, other essential worker)
- date 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. HIV, TB, respiratory, cardiovascular diseases)
- pregnancy
- smoking status
- BMI
- influenza and pneumococcal vaccination status (where applicable and available)
- respiratory co-infections (where available)
- severity

An example layout of this descriptive analysis is provided in Table 3 below.

Table 3. Example of descriptive table for cases and controls

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	x/x (x)	x/x (x)
Missing	X	X
Sex		
Female	x/x (x)	x/x (x)
Missing	X	X
Health 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 by documentation	x/x (x)	x/x (x)
Unvaccinated	x/x (x)	x/x (x)
Fully Vaccinated	x/x (x)	x/x (x)
Partially vaccinated	x/x (x)	x/x (x)
Unknown	x/x (x)	x/x (x)
Etc.		

*IQR: inter-quartile range

This analysis is a test-negative design. The measure of association is an odds ratio (OR). This can be measured by logistic regression. An OR = 1 suggests 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. Because it is a test-negative design, time (onset or testing) should be included in all analysis.

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.

Prior to multivariable analysis, a model development strategy should 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 final CVE can be estimated using the formula above.

3.12.2.1 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 4 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 4. Example of table showing vaccine effectiveness against COVID-19 adjusted for various covariables by sex and age group

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		

In sensitivity analyses, the CVE analysis will be performed with different cut-offs of numbers of days between onset and swabbing, onset and hospitalisation, and between vaccination and onset of symptoms. Other sensitivity analyses, where possible, include exclusions of prior test positives and those positive to a seasonal coronavirus (e.g. HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1).

Controls who are negative by PCR but have chest imaging test results suggestive of COVID-19, and those with prior SARS-CoV-2 infection in the 3 months prior to admission, may be excluded as controls in sensitivity analyses.

Please see section on further analyses including the possibilities for pooled analysis, in **Annexes 5 and 6.**

4 Limitations

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. Potential limitations to the VE estimates for COVID-19 are discussed below.

4.1 Potential biases

4.1.1 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 (e.g. presence of chronic diseases) will minimise negative confounding.

4.1.2 Positive confounding

Positive confounding refers to biases that reflect a 'healthy vaccinee effect'. People with a healthy lifestyle will be more likely to accept vaccination, thus leading to an increase of estimated 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 who have a higher risk of infection (through increased exposure – e.g. not routinely wearing facemasks where recommended) may also be less likely to be vaccinated which may also increase their exposure to disease. If positive confounding is present, CVE will be overestimated.

4.1.3 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 surveillance population for which no information is collected in the questionnaire and that therefore could lead to positive or negative confounding. Therefore, some residual unmeasured confounding may remain.

4.1.4 *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.

4.1.5 *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. Vaccination status of cases and controls using an independent source (i.e. vaccination register, primary care clinicians).

4.1.6 *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).

4.2 Representativeness of subjects

SARI surveillance 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, only seeing a primary care clinician 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. Beside the collection of dates of onset/admission/respiratory specimen collection, health-seeking behaviour and case-management strategies should be described for each SARI surveillance system and it should be noted how these may affect the CVE estimates.

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

5 Community Engagement

This study will engage with the community in so far as it will be conducted in hospitals, which serve their local communities. Efforts should be made to be transparent about the objectives of the study with interested community members.

6 Annexes

6.1 Annex 1: Identifying SARI patients using relevant ICD codes

For SARI surveillance systems that include electronic patient records where ICD codes are routinely used, potential SARI patients can be identified using relevant ICD codes (Table 1) for further screening (Figure 1).

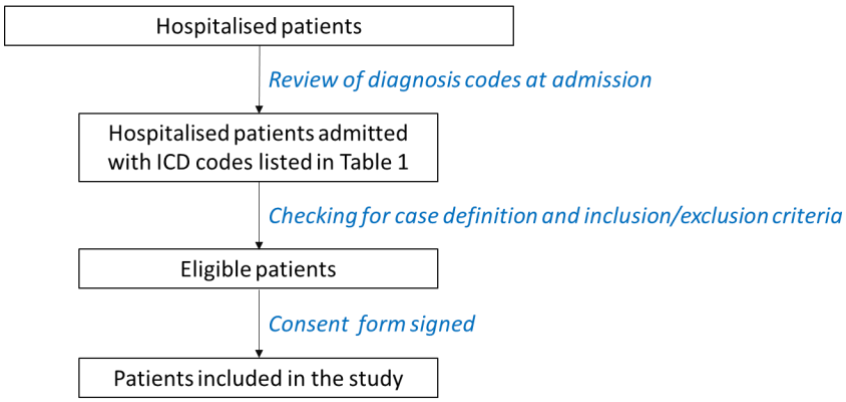


Figure 1: Algorithm to identify SARI patients using ICD codes.

Table 1. List of diagnosis codes for which patients could be screened for presence of SARI symptoms according to the case definition.

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
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

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
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
Other	Anosmia, ageusia, myalgia	781.1, 729.1	R43.0, R43.2, M79.1

**SIRS: Systemic inflammatory response syndrome*

6.2 Annex 2: Other control groups (optional)

During periods of low circulation of respiratory viruses, the number of SARI patients testing negative for SARS-CoV-2 may be limited. Therefore, some sites may wish to include other control groups, either in addition to SARI controls or as an alternative, which must be recruited throughout the whole study period.

Example of other control groups include:

- Patients hospitalised with non-SARI related symptoms matched by time, age group and, if possible, underlying conditions
    - Example of source of hospitalised non-SARI cases: hospital wards admitting patients without COVID-19
- Primary care: selection of primary care patients belonging to the hospital catchment area and vaccination target group matched by time and age-group.
    - Example of source of primary care patients: contact the primary care provider of the case, and select patients from his/her list (matching by primary care provider)
- Community controls
    - Random selection of community controls belonging to the vaccination target group matched by time and age group (e.g. vaccine registry, telephone random survey, other planned survey)
    - Vaccination coverage in cases will be compared to vaccination coverage in the vaccination target population (screening method). Vaccination coverage should be available by time, age group and comorbidities.
        - Random sample of primary care providers in the hospital catchment area to compute the proportion of primary care patients who are vaccinated.
        - Vaccination coverage using vaccination centres in the hospital catchment area: the vaccination coverage can be computed by dividing the number of individuals vaccinated (by age group, target group) by the number of individuals in the hospital catchment area (by age group, target group). Several methods can be used to estimate the population in the hospital catchment area.

All control groups should represent the vaccination coverage of the population giving rise to the cases. As the circulation of SARS-CoV-2 and vaccination coverage changes over time, it is recommended to match cases and controls by time (e.g. onset of SARI symptoms) or adjust by time in the analysis.

6.3 Annex 3: List of ICD-9 and ICD-10 codes for pre-existing chronic conditions

To be adapted to each country.

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

	M34.81, M05.10	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, **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.

6.4 Annex 4: Example of a list of variables, definitions and coding; hospital-based COVID-19 vaccine effectiveness

This list should be adapted to the variables collected at country level and to the coding used for SARI surveillance
Individual data

- *Countries to list all the variables collected and their coding*
- *Countries to indicate all modifications in the variables collected compared to variables below*
- *Optional variables shaded in grey*

	Variable	Type	Values and coding	Definition
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	Agreement of patient to participate (where appropriate, i.e. for countries requiring consent)
			1 = Yes	
			2 = Not required	
			8 = Do not know	
	consent_sp	Text		Reason provided for non-participation
Target group	target	Numeric (categorical)	0 = none	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)
			1 = age 80+ years	
			2 = health personnel	
			3 = other at-risk occupation	
			4 = in long-term care	
			5 = other	
			8 = do not know	
	target_sp	Text		Specify other target group
Hospital/ward information	hospitalward	Numeric (categorical)	0 = Special COVID-19 ward	First ward of referral
			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	

	Variable	Type	Values and coding	Definition
Hospital/ward information continued	hospitalward_oth	Text		Specify other ward
	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
	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
			1 = male	
			3 = other	
			8 = do not know	
	dob	Date	dd/mm/yyyy	Date of birth (only if no age; once age calculated from dob this will be dropped) *(optional)*
	smoking	Numeric (categorical)	0 = Never	Never, former (stopped smoking at least 1 year before hospital admission, current smoker *(optional)*
			1 = Former	
			2 = Current	
			8 = Do not know	
	pregnant	Numeric (categorical)	0 = No	Whether patient is pregnant
			1 = Yes	
			8 = Do not know	
	gestation	Numeric (integer)		Number of weeks of gestation of pregnancy *(optional)*
	hcw	Numeric (categorical)	0 = No	Whether the patient is a health worker
			1 = Yes	
			8 = Do not know	
	residence	Numeric (categorical)	0 = at home, not dependent on home support/care	Patient residence at time of SARI onset. Whether patient was living at home or was institutionalised, or had pre-hospital dependence on home support/care
			1 = at home, but dependent on home support/care	
			2 = institutionalised (LTCF)	
			3 = other	
			8 = Do not know	
	residence_sp	Text		Specify other residence (e.g. prison)
	occupation	Text		Patient's occupation (note: this may be collected another way, e.g. by national occupational code, depending on country) *(optional)*

	Variable	Type	Values and coding	Definition
Patient characteristics (continued)	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)*
Risk-taking behaviour (optional)	risk1_mask	Numeric (categorical)		Frequency of mask use
	risk2_hand	Numeric (categorical)		Frequency of hand hygiene
	risk3_socialdist	Numeric (categorical)		Frequency of social distancing
	risk4_view	Numeric (categorical)		Pandemic perception
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
	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)

	Variable	Type	Values and coding	Definition
Underlying chronic conditions (continued)	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 visited their primary care clinician in the 3 months prior to COVID diagnosis for an underlying chronic condition
	hosp_visit	Numeric (integer)		Number of times patient was admitted to hospital in the 12 months prior to COVID diagnosis for an underlying chronic condition
Optional underlying chronic conditions	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)*
	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)*
	tuberc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Tuberculosis *(optional)*
	stroke	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Stroke *(optional)*

	Variable	Type	Values and coding	Definition
Optional underlying chronic conditions continued	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)*
	art_pre	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	ART (anti-retroviral therapy) *(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)*
	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)*

	Variable	Type	Values and coding	Definition
Pre-symptomatic medication *(optional)* continued	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 *for respiratory infections (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) *(optional)*	flu_vacc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received current seasonal influenza vaccination
	flu_vaccdate	Date	dd/mm/yyyy	Date of last influenza vaccination
	ppv_vacc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received PPV23 vaccination
	ppv_vaccdate	Date	dd/mm/yyyy	Date of last PPV23 vaccination

	Variable	Type	Values and coding	Definition
Vaccination (other than against COVID-19), continued *(optional)*	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	History of fever (undocumented temperature)
	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 of symptoms
	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 (dyspnoea)
	malaise	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Malaise
	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

	Variable	Type	Values and coding	Definition
Symptoms at or prior to admission, continued	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 (rapid breathing) 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)*
	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)*

	Variable	Type	Values and coding	Definition
Symptoms at or prior to admission, continued *(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
	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
	lab_covid	Numeric (categorical)	0 = Negative 1 = Positive 2 = Inconclusive/undetermined 8 = Do not know	Laboratory result: virus type SARS-CoV-2
	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

	Variable	Type	Values and coding	Definition
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 2 = Inconclusive/undetermined 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_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	Laboratory result: virus type other coronavirus
	resp_virus	Numeric (categorical)	0 = None 1 = RSV 2 = Metapneumovirus 3 = Other respiratory infection 8 = Do not know	Which other non-influenza, non-coronavirus patient tests positive for
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_sp	Text		Specify which other clinical frailty score was used
	frailty_barthel	Text		Total Barthel score at admission (if used)
	frailty_cfs	Text		CFS score at admission (if used)

	Variable	Type	Values and coding	Definition
Case definitions and outcomes	Covid	Numeric (categorical)	0 = No 1 = Confirmed 2 = Other coronavirus 3 = Suspected 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)
	severity_ox	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient received oxygen (e.g. high-flow), but was not intubated/ ventilated
	severity_vent	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Whether patient was ventilated (invasive ventilation)
	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
Outcomes (optional)	severity_ards	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	acute respiratory distress syndrome (ARDS)
	severity_resp	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Severe respiratory distress
	severity_rate	Numeric (integer)		Respiratory rate (breaths/minute) at admission or enrolment
	severity_SpO2	Numeric (percent)		Blood oxygen level (SpO2) on room air at admission or enrolment
	severity_sepsis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sepsi§s
	severity_shock	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Septic shock

Pandemic COVID-19 vaccination	panvaccany	Numeric (categorical)	0 = No	Received pandemic COVID-19 vaccination
			1 = Yes	
			8 = Do not know	
	panvaccdose	Numeric	0, 1, 2	Number of doses received
	panvaccdate1	Date	dd/mm/yyyy	Vaccination date, first dose
	panvaccdate2	Date	dd/mm/yyyy	Vaccination date, second dose
	panvacctype1	Text		Type of vaccine (product name), dose 1
	panvacctype2	Text		Type of vaccine (product name), dose 2
	panvaccbatch1	Text		Batch number of vaccine product, dose 1
	panvaccbatch2	Text		Batch number of vaccine product, dose 2

6.5 Annex 5. Pooling data

This is a summary of a plan of analysis if data from different sites/countries using the same protocol are pooled.

6.5.1 Data storage and management

Anonymised data should be transferred to the central hub using a secure data platform.

A country (identifier will be included in each record, and a hospital code will be included (e.g. a unique number). Data management will follow the basic principles outlined below and in section 3.11 (Data management). A country-specific flowchart of exclusions and restrictions will be shared with each of the participating countries. Variables will be re-coded 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.

6.5.1.1 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).

6.5.1.2 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.

6.5.2 Pooled analysis outline

If pooling across sites is conducted, a central hub will be responsible for the pooled analysis. The higher sample size for this analysis will provide more power (and precision). Data can be coded as outlined in Annex 3, or a codebook can be provided by the surveillance teams to the central hub hat includes the variable names, descriptions and coding. The central hub 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 surveillance sites will be determined, using Q-test and the I2 index (see boxes for formulae below).

Briefly, cases and controls will be described by baseline characteristics, and uni- and multivariable analyses performed as described in section 3.13.1 for individual analysis.

6.5.3 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, 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 analysis.*

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

6.5.4 Pooled analysis plan

6.5.4.1 Descriptive pooled analysis

The proportion of eligible hospitalised cases and controls who were included in the surveillance 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 flowchart.

Cases and controls will be described by baseline characteristics.

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

- Number of hospitals participating and catchment population
- Beginning of vaccination campaigns for pandemic vaccine
  - Beginning of the analysis period
  - End of the analysis period
  - 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

6.5.5 Measure of effect

This analysis 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.

6.5.5.1 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.

6.5.5.2 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 analysis period 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.

6.5.5.3 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.

6.5.5.4 Continuous variables

Continuous variables in the COVID-19 datasets include age, time of onset of symptoms, primary care clinician 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.

6.5.5.5 *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 design and quality characteristics will also be checked. Other surveillance 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.

6.5.5.6 One-stage pooled analysis approach

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

Individual surveillance system data will be pooled into one dataset and analysed as a 1-stage model with surveillance site as a fixed effect. This could provide a large enough sample size to obtain (for example) an estimate of CVE early in the analysis 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 surveillance 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 surveillance 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.

6.5.5.7 Controlling for hospital effect

Primary analysis will be carried out using simple logistic regression to obtain the individual site 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 surveillance site 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.

6.5.5.8 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.

Two-stage pooled analysis approach

If adequate sample size by site 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 site-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 surveillance participants for whom variables were collected differently.

6.5.6 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 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

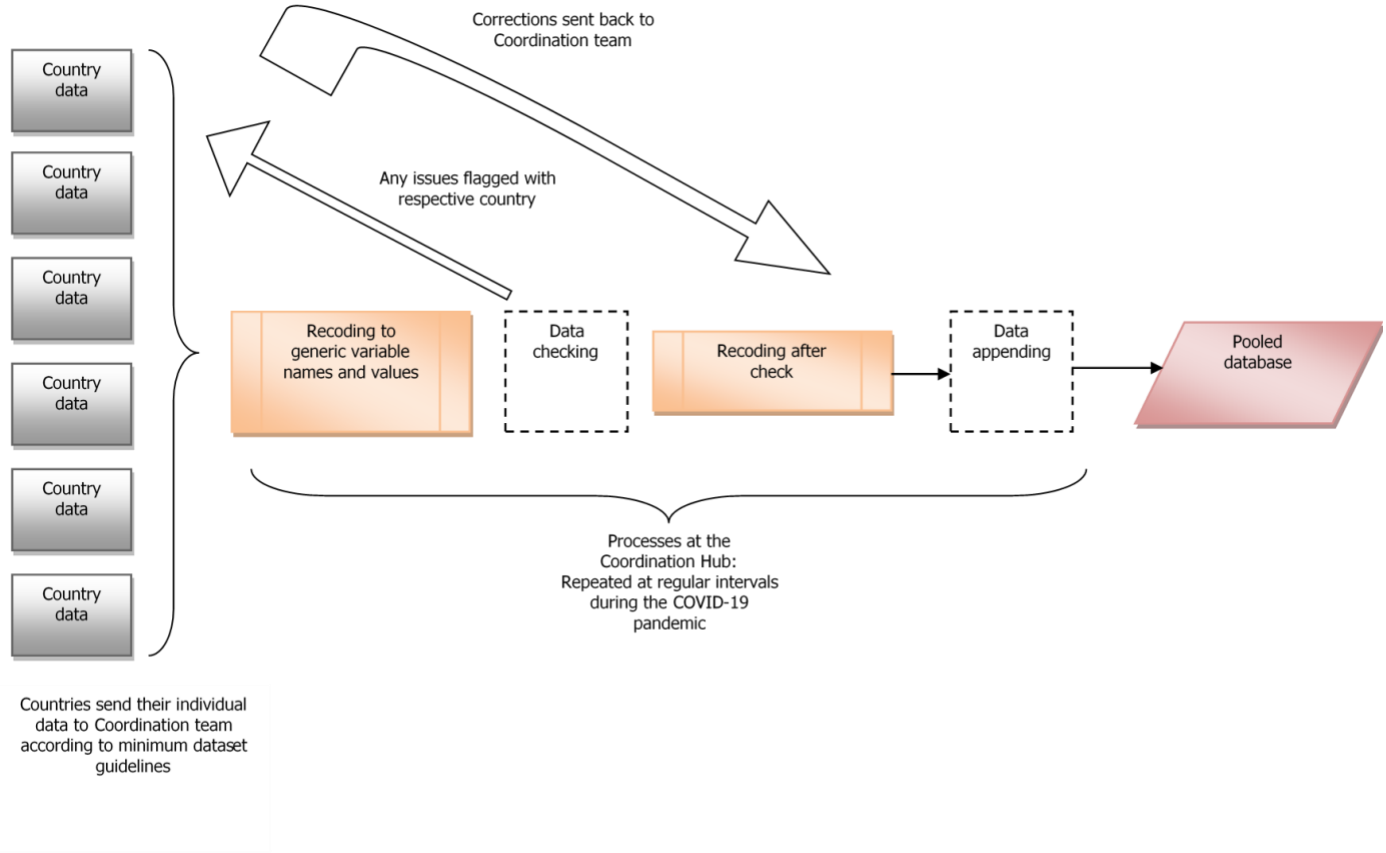
We can also put time as a variable in the model. As time may be an effect modifier (there may be different CVE at different times of the pandemic), then we can add an interaction term or perform the proposed stratified analysis.

6.5.7 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.

6.5.8 *Data flow for pooled dataset*



6.6 Annex 6: Additional information for pooled analysis

If data are to be pooled, surveillance specifications for each country should be summarised in this annex. Each country annex should include:

- description of the hospitals participating in the surveillance system (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
- target groups for vaccination, with vaccination timelines if possible
- 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.