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

I-MOVE-COVID-19 Network

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

COVID-19 vaccine effectiveness at primary care level in Europe: generic protocol

FEBRUARY 2021

V 2.2

I-MOVE-COVID-19 Network

WP4 coordinated by Epiconcept

Based on: current literature, I-MOVE-COVID-19 protocol on risk factors for COVID-19 at primary care level, I-MOVE primary care generic influenza vaccine effectiveness protocol 2019–2020

Table of contents

Version history

Version	Date finalised	Created/modified by	Comments
1.0	2020-10-09	Epiconcept	Initial draft for internal review
1.1	2020-10-12	Epiconcept	Draft to be sent to I-MOVE-COVID-19 primary care partners
2.0	2020-12-08	Epiconcept	Version with country comments incorporated
2.1	2020-12-09	Epiconcept	Inhouse review comments added
2.2	2020-02-15	Epiconcept	Further comments from I-MOVE-COVID-19 partners included

Abbreviations

COVID-19	Coronavirus disease 2019
EEA	European Economic Area
ECDC	European Centre for Disease Prevention and Control
EU	European Union
GP	General Practitioner
HCW	Healthcare worker
ICD	International classification of diseases
ILI	Influenza-like illness
I-MOVE	Influenza – Monitoring Vaccine Effectiveness in Europe
MS	Member States
OR	Odds ratio
RT- PCR	Reverse-transcriptase polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome – coronavirus 2
VC	Vaccination coverage
VE	Vaccine effectiveness

➢ (The arrow indicates the sections that Member States should adapt and provide details for in their study annexes.)

1. Background

The end of 2019 saw the emergence of a novel severe acute respiratory syndrome – coronavirus 2 (SARS-CoV-2), which can cause coronavirus disease 2019 (COVID-19).

As of the week 5 2021, **20 478 718 cases** and **495 672 deaths** have been reported in the EU/EEA (1).

I-MOVE (Influenza – Monitoring Vaccine Effectiveness in Europe), first established in 2007(2), was the first network to monitor influenza vaccine effectiveness (VE) within and across the seasons in the European Union (EU) and the European Economic Area (EEA). The network has two components, one for primary care practices, recruiting patients with influenza-like illness (ILI) and the other for hospitals, recruiting patients with severe acute respiratory illness (SARI).

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

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

The WP2 primary care surveillance for COVID-19 is coordinated by Nivel (Netherlands institute for health services research). The information for this WP4 study will be collected through the WP2 network. The I-MOVE-COVID-19 primary care network comprises nine sentinel surveillance networks in six European Union (EU) Member States (MS)[2] and in England and Scotland. The laboratory component of the network includes regional and national reference centres from the participating countries. While each of the surveillance sites can analyse their data separately, pooling the data for overall analysis will provide a sample size big enough to answer study questions with reasonable precision.

This document presents the core I-MOVE-COVID-19 European protocol for COVID-19 vaccine effectiveness (VE) at primary care level. The specificities of each site's COVID-19 data collection can be detailed in the individual site protocol annexes.

While control and mitigation strategies such as testing, contact tracing and quarantine procedures help keep COVID-19 in check, having a critical level of immunised people in a population is a further and important method to minimise transmission. Having an effective and safe vaccine against SARS-CoV-2 will help reach this goal while minimising morbidity and mortality among the population. Many vaccines are under development, and as of the 9th of December 2020, one vaccine has already been authorised and used in the UK and many others are in phase 3 clinical trials (3). Post-marketing COVID-19 vaccine effectiveness studies with good precision will be key to determine if vaccines are effective or not among the target group for vaccination. A high sample size is important to ensure a good precision around the

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

[2] France, Ireland, The Netherlands, Portugal, Spain (two sites: the Spanish national system and the Navarra regional system) and Sweden.

point estimate, and for the possibility to rapidly assess the VE. Pooling studies in several European countries may achieve these objectives.

This protocol is an evolving document and will be updated as more information comes in about the types of vaccines used, the doses, the target groups for vaccination and the rollout of the vaccination programmes.

This protocol is written in a generic manner and country-specific details of each study will be outlined in the study annexes (Annex 6).

2. Objectives

2.1. Primary objective

The primary objective will be to measure, for each European primary care surveillance site country and, for pooled analyses, across all participating European primary care surveillance sites, the direct effect (effectiveness) of COVID-19 vaccines by vaccine type and brand against laboratory confirmed SARS-CoV-2 infection using a pooled analysis.

2.2. Secondary objectives

The secondary objectives are to

- measure VE:
  - in each of the participating surveillance networks;
  - by risk groups;
  - by sex;
  - by age groups;
  - by COVID-19 vaccination target group;
  - regularly over calendar time;
  - according to time since vaccination;
  - for one or two doses of vaccine, if applicable;
  - by delay between doses (if two doses received);
- identify key phenotypic or genotypic evolutions that could affect vaccine performance and to estimate VE against specific genetic variants.

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

3. Methods

3.1. Study design

- Test negative, case-control study design.
- Multicentre test-negative case-control study, using pooled data from several countries.

3.2. Study population

The study population comprises community-dwelling individuals with no contraindication for COVID-19 vaccination who consult a participating physician with COVID-like symptoms.

- *Surveillance sites to describe the setting (number of primary care practices included, number of primary care physicians, catchment population if possible)*

3.3.Study period

The study period starts when the COVID-19 vaccine is available in each of the participating countries and when SARS-CoV-2 is circulating. The study period is defined for each priority vaccination group, and begins for each vaccination group, when vaccination campaign in this group begins.

Participating primary care practices carry out the study throughout the year.

- *Study sites to define the beginning of the study period (date/month/year)*
- *Each study site specifies the date of the start of their vaccination campaign for each priority vaccination group.*

3.4. Outcomes

The primary outcome of interest will be PCR laboratory-confirmed COVID-19 in symptomatic patients of all ages consulting at primary care level.

Secondary outcomes of interest, in the same patient group at primary care level, will be genetic variants of COVID-19.

3.5.Case and control definitions

Patients are persons consulting a general practitioner, defined as someone either

- Having a face-to-face consultation with the practitioner (in the practice or at the patient's home)
- Having a telephone/video consultation with the practitioner[3]

A **suspected COVID-19 case** is defined as a patient with at least one of the following:

- Cough
- Fever
- Shortness of breath
- Sudden onset of anosmia, ageusia or dysgeusia

A **confirmed COVID-19 case** will be defined as a suspected COVID-19 case with a respiratory sample PCR-positive for SARS-CoV-2.

[3] We can include these patients if a swab can be taken soon after the consultation (either by the patient self-swabbing, visiting a specific swabbing centre or the practitioner taking a swab, either at patient's home or at the general practitioner's office)

A **COVID-19 negative patient (a test-negative control)** will be defined as a suspected COVID-19 case with a respiratory sample negative for SARS-CoV-2.

3.6. Laboratory methods

Primary care practitioners will collect respiratory specimens from either all or a systematic sample (see section 3.7.1) of eligible patients (suspected COVID-19 cases consulting a practitioner and consenting to take part in the study), respecting safety standards for COVID-19 and following WHO biosafety guidelines.[4] Depending on the setting, some practitioners will refer patients to specific COVID-19 testing centres, or some patients may even be carrying out self-swabbing at home.

A comprehensive generic laboratory protocol will be developed and presented in the future alongside this VE protocol.

- *Each study site to describe the type (nasopharyngeal/oropharyngeal or both) and number of swabs taken for each patient*
- *Each study site to describe where swabbing will be carried out (at practice, at home, in centres, a mixture)*

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

The ECDC-recommended SARS-CoV-2 laboratory confirmation is by viral RNA detection with nucleic acid amplification tests, such as RT-PCR (4). Isolates will undergo molecular analysis for currently circulating SARS-CoV-2 virus. During the influenza season, tests should also be performed for influenza viruses as long as there is circulation of influenza viruses.

Information will be collected on type of test.

Following the procedures outlined by each study, a systematic sample of isolates (or all isolates) will undergo gene sequencing. The sampling procedure can include sequencing all isolates, or a random sample thereof. The sample should be random and thus be representative of cases and be large enough to provide reasonable precision when calculating proportions of virus change over time. Gene sequences should also be uploaded to GISAID's open access EpiCoV platform. Gene sequence information can be provided directly to the I-MOVE-COVID-19 central hub, or the GISAID EpiCoV accession number can be provided alongside the I-MOVE-COVID-19 unique identifier to link these data (see annex 2). If random selection of viruses to sequence is used, the proportion sequenced may vary over time, according to a variety of factors, including resources and incidence of SARS-CoV-2. Study sites should indicate their sampling fraction for sequencing over time (see annex 2, table 5). Processed genetic information, e.g., name of genetic clade, can also be included within the epidemiological database.

- *Each study site to describe the laboratory procedures (samples taken, storage, transport)*

[4]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'.[3]

- *Each study site to describe the tests and the kits used (and their sensitivity, specificity, PPV) for COVID-19 and, if needed, other respiratory virus detection*
- *Each study site to describe if the laboratory participates in QA/QC (Quality Assurance/Quality Control) schemes*
- *Each study site to describe the selection of specimens and the methods for genetic and, when it becomes available, antigenic characterisation*
- *Each study site to describe genetic and, when it becomes available, antigenic analyses and specify sequencing methods*

3.7.Study participant identification

3.7.1. Selection of patients to swab

Study participants are identified among patients presenting to or referred to a participating GP with symptoms compatible with suspected COVID-19.

Following the procedures outlined by each study, all suspected COVID-19 cases are selected and asked to provide a nasal/throat swab specimen for SARS-CoV-2 testing. Sampling all suspected COVID-19 cases is preferred, in particular all patients aged 65 and over. If this is not possible, then a systematic sample with known sampling fraction can be taken, e.g., the first three suspected COVID-19 cases seen each week per GP, including all patients aged 65 and over. SARS-CoV-2-positive suspected COVID-19 patients are considered as lab-confirmed COVID-19 cases. SARS-CoV-2-negative suspected COVID-19 patients are considered as controls.

- *Each study site to describe the procedures to select suspected COVID-19 cases to swab*

3.7.2. Patient inclusion criteria

Patients are eligible if they meet the inclusion definition and consent to participate (the patient or her/his legal guardian gave consent to participate according to the local ethical review process.

- *Each study site to describe country procedures for oral informed consent or written informed consent and specify these in the study annexes.*

3.7.3. Patient exclusion criteria

Patients are **excluded from the primary analysis** if they:

- refuse to participate in the study;
- are not swabbed;
- are unable to give informed consent or follow an interview in their native language because of aphasia, reduced consciousness, or other reasons;
- cannot be swabbed due to severe septum deviation, obstruction or other conditions that contra-indicate swabbing;
- have contraindications for the COVID-19 vaccination;
- are swabbed more than 7 days after symptom onset (to avoid false negatives; the exact cut off will be determined as more research on this comes in);
- have received antivirals ≤14 days prior to swabbing (to avoid false negatives; the exact cut off and types of antivirals will be determined as more research on this comes in);
- are institutionalised (virus exposure and risk factors may be different - specific cohort studies can be undertaken in these groups);

- were vaccinated within 7 days of symptom onset (more information on vaccines needs to be available to determine the cutoff as to when we can consider a patient to be "immunised");
- had an inconclusive RT-PCR test.

Reasons for exclusion are documented.

We will collect information on the exclusion factors and exclude patients according to available evidence (not all available at time of writing) on these factors.

In sensitivity analyses, we will carry out the VE analysis with different cut-offs of numbers of days between onset and swabbing, between vaccination and onset of symptoms. Other sensitivity analyses include if a current control (SARS-CoV-2 negative) and had tested positive (by PCR or serology) to SARS-CoV-2 within a certain number of months of consultation (the exact number of months will be defined when more is known about immunity).
If information is available, we will exclude those positive to a seasonal coronavirus (e.g. HCoV-NL63, HCoV-229E, HCoV-OC43 and HCoV-HKU1) in a sensitivity analysis.

Please see section 3.12.2.

3.7.4. Restriction to priority groups for vaccination

Patients will only be included in the analysis, if they are part of a target group for COVID-19 vaccination, at time of swabbing – and vaccination rollout has begun. This way patients included in the study will have the chance of being vaccinated.

Patients swabbed prior to rollout of the COVID-19 vaccination campaign in their particular target group will not be included, as they are not eligible for vaccination.

3.8.Exposure (vaccination)

3.8.1. Definition of COVID-19 vaccination status

An individual will be considered as vaccinated against COVID-19 with a product-specific vaccine under the following categories:

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

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

In a sensitivity analysis, those fully vaccinated will be analysed separately for those receiving doses at appropriate gaps (days between doses) depending on the vaccine brand, and those receiving doses at gaps not recommended by the vaccine manufacturer.

It is crucial that the vaccination status, doses and date(s) of vaccination variables are collected with the utmost care to ensure data completeness and quality. Additionally, place of vaccination (GP, vaccination centre, etc.) and vaccine status ascertainment should be documented.

3.8.2. Vaccination status ascertainment

The exposure of interest in this study is a vaccination history with COVID-19 vaccine. The vaccination history includes date of administration, type of vaccine and brand name, and the number of doses received. Documenting the batch codes (where this is feasible) will allow identification of the vaccine brand, the vaccine content and the dose. Product type and date are critical variables which must be validated.

An individual is considered as vaccinated against COVID-19 if:

- he or she reports having received a COVID-19 vaccination;
  or
- he or she is registered as vaccinated in the GP information system;
  or
- he or she is registered as vaccinated in a vaccination registry;
  or
- his or her insurance company can show evidence of pharmacy delivery or re-imbursement of COVID-19 vaccine/vaccination
  or
- COVID-19 vaccination has been recorded in his/her vaccination card/vaccination booklet.

☑ Each study site to document:

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

3.9. Data to be collected, including potential effect modifiers and confounding factors

3.9.1. Patient characteristics

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

- Age in years
- Sex
- GP code (in order to account for clustering by GP)
- Smoking history (never smoked, former smoker (stopped smoking for at least one year), current smoker (including stopped smoking less than one year ago). Smoking refers to any type of smoking (cigarettes, cigars, vaping, etc.)
- Pregnancy (yes/no)
- Healthcare worker (yes/no)

Healthcare worker

The definition of a healthcare worker for the purposes of this study is a person who is working ((paid or on a regular voluntary basis) in healthcare AND has contact with patients (any type of patient) during his/her work. This includes: doctors, nurses, emergency medical personnel, medical and nursing students having contact with patients, as well as porters and cleaners.

3.9.2. Information on consultation

- Type of consultation: in practice, video, telephone, home, at a COVID-19 centre
- Date of consultation

3.9.3. Clinical signs and symptoms

Collection of good quality symptom information is crucial for the VE study in order to be able to validate the case definition used. As a minimum:

- fever/feverishness
    - if fever: measured fever (with temperature)
- cough
- shortness of breath
- anosmia
- ageusia
- dysgeusia

As many study sites will use this protocol to measure influenza vaccine effectiveness, the following variables should be collected to be able to determine which patients meet the EU ILI case definition:

- headache
- sore throat
- myalgia
- malaise

As part of the I-MOVE-COVID-19 risk factor study, many studies also collect the following symptoms in order to better understand the clinical spectrum of disease:

- coryza, rhinitis
- chest pain
- chills
- fatigue

- nausea
- vomiting
- diarrhoea
- stomach ache (abdominal pain)
- conjunctivitis
- dizziness
- cyanosis or associated pulse oximetry
- rash or other dermatological manifestation
- palpitations

We will collect the **date of symptom onset**.

3.9.4. Information on swabbing and test results

For each patient we will collect information on:

- date of swabbing
- place of swabbing (GP practice, COVID centre, self-swabbing)
- type of swab (nasopharyngeal, oropharyngeal, both)
- type of COVID-19 test (PCR, point-of-care)
- result of COVID-19 test

Some studies will be carrying out testing for other respiratory viruses. We will collect:

- test results from any other respiratory viruses (e.g. rhinovirus, RSV, enterovirus, adenovirus, human metapnemovirus, seasonal coronaviruses, etc.)

3.9.5. Pre-existing chronic conditions

If physicians are recruiting cases and controls using electronic medical records, the list of ICD codes can be used to document a study participant's chronic diseases (see Table 1):

> The list below is very comprehensive. A suggested minimum number of chronic diseases is specified below.

Table 1: ICD-9, ICD-10 and ICPC-2 codes for chronic diseases

Category	ICD-9	ICD-10	ICPC-2 (to be confirmed)
Anaemia	280–285	D50-64	B78, B80- B82
Asplenia	746.87, 759.0	Q89.01, Q20.6, Z90.81	*(to be completed)*
Asthma	493.0, 493.1, 493.9	J45	R96
Chronic liver disease	571	K70, K72-74, K754, K769	
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	K73, K83, K77, K74-K76, K78-K80
Diabetes	250	E10-11	T90
Hypertension	401, 401.0, 401.9, 405, 405.91, 405.99,	I10, I15.8, I15, I15.1, I15.2, I97.3, I27.0	K86-K87
Obesity	27800, 278.01, 278.03	E66.01, E66.2, E66.9	T82

Immunodeficiency* or organ transplant	042, 279, V08, V42	B20, D80-84, D89.8-9, Z21, Z94	B99
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,	*(to be completed)*
Renal disease	274.1, 408, 580–591, 593.71–593.73, 593.9	M10.30, N00-19, N20.0, N28.9	U99
Dementia	290, 294, 331	F01, F03, F05, G30, G31, G91, G94	P70
Stroke	348, 438	G93, I67.83, I69	K89-K90
Rheumatologic diseases	446, 710, 714	M30-34, M35.0, M35.5, M35.8-9, M05-06, M08, M12.00	L88
Cancer	140–208	C00-96	A79, B72, B74, D74-D78, F74, H75, K72, L71, N74, N76, R84, R85, S77, S79, T71, T73, U75-U77, U79, W72-W73, X75-X77, X81, Y77-Y78
Lung disease excluding asthma)	011, 490–511 (exclude asthma), 512.8, 513–517, 518.3, 518.8, 519.9, 714.81	A15, J40–44 J46–47, J60–94, J96, J99, J182, M34.81, M05.10	R83, R79, R91, R95, R99
Tuberculosis		A15–A19	A70

*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.

If ICD/ICPC codes are not available, a list of underlying conditions should be prepared by using a short questionnaire.

The list of underlying conditions in the questionnaire should include if possible:

- diabetes (sites are encouraged to distinguish between type 1 and type2);
- cardiovascular disease: myocardial infarction, angioplasty, coronary artery bypass surgery, stroke, transient ischemic attacks, treated hypercholesterolemia, not including hypertension;
- hypertension;
- chronic pulmonary disease (not including asthma);
- asthma;
- cancer;
- renal disease;
- chronic liver disease;
- rheumatologic diseases
- obesity (see paragraph below)
- immunodeficiency.

For obesity, we will collect body mass index (BMI). If it is possible to collect the actual BMI or height and weight (in metric units), this is preferred. If not possible, we suggest categories (BMI: 30–39 and ≥40).

3.9.6. Pre-symptomatic vaccination status (other than COVID-19 vaccine)

We will collect information on influenza, pneumococcal and BCG vaccination:

- Seasonal influenza vaccination from the most recent influenza season (with date of vaccination)
- Latest pneumococcal vaccination type (with year if possible)

See section 3.8.3 on vaccination status ascertainment for information on methods for influenza vaccination and pneumococcal vaccination ascertainment.

3.9.7. Pre-symptomatic medication status (optional)

We will document whether the patients were prescribed any of the listed medications in the 2 weeks preceding symptom onset.

The three main medications to be included are angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARBs) and non-steroidal anti-inflammatory drugs (NSAIDs). Additional medications include antivirals, statins and other anti-hypertensive medication. For each of these:

- An individual will be considered as "on" the medication if she/he was prescribed/was on treatment before onset of symptoms
- An individual will be considered as "not on" the medication if s/he was only prescribed/was only on treatment **after** symptom onset.

- angiotensin-converting enzyme (ACE) inhibitors;
- angiotensin II receptor blockers (ARBs);
- non-steroidal anti-inflammatory drugs (NSAIDs);
- statins;
- corticosteroids;
- biological disease-modifying anti-rheumatic drugs (DMARDs);
- current/recent cancer chemotherapy;
- antithrombotic/ platelet aggregation inhibitors;
- metformin;

The minimum information, if not otherwise specified, for the medication status of these medications is "Was the patient on the drug in the 2 weeks preceding symptom onset? yes/no".

3.9.8. Antiviral use before swabbing

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

3.9.9. Information on previous SARS-CoV-2 infection

Among those patients consulting their GP with COVID-19-like symptoms, some may have already had a SARS-CoV-2 infection in the past. Collecting information on previous infection gives information on those who have had more than one SARS-CoV-2 infection and also helps with the control selection. If possible, we will collect the following information:

- whether the patient had a previous positive SARS-CoV-2test (yes/no/unknown)
- type of test: PCR, point-of-care test, serology
- date of test (in case of multiple positive test results, the most recent)

- history of COVID-19, e.g. clinical confirmation, contact to positive case

In the future, we may also include results of antibody tests here.

3.9.10. Health care utilisation in the previous 12 months

In order to document and control for healthcare seeking behaviour in the control groups and the severity of underlying conditions, we will collect:

- the number of GP visits made (face-to-face, or telephone consultations) in the past 12 months before inclusion in the study
- the number of hospital admissions due to underlying conditions in the 12 months prior to inclusion in the study

3.10. Data

3.10.1. Sample size

The number of individuals included in the VE study will depend on the number of patients consulting at primary care level with COVID-19-like symptoms and the number of patients laboratory-confirmed with SARS-Cov-2. Sample size will also depend on length of time in the study.

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

The following sample size estimates focus on the precision of the VE estimate (Table 2). As mathematically the lower confidence interval boundary is always larger than the upper confidence interval boundary, we focus on a precision of the lower confidence interval, ranging between 10 and 30%. We also assume a case to control ratio of 1:4. We include varying vaccine coverage among the source population between 30% and 50%, varying vaccine effectiveness with the OR between 0.2 and 0.7.

A dynamic version of this table in Excel sheet format is available for study sites on request.

Table 2: Sample size calculations

Precision of lower CI boundary	Controls /	Detectable OR	Vaccine coverage in source population/controls	Number of cases	Number of controls	VE	CI

	case						
0.1	4:1	0.1	0.5	105	419	90	80–95
0.1	4:1	0.2	0.5	192	766	80	70–87
0.1	4:1	0.3	0.5	308	1232	70	60–78
0.1	4:1	0.4	0.5	455	1821	60	50–68
0.1	4:1	0.5	0.5	636	2542	50	40–58
0.2	4:1	0.1	0.5	42	167	90	70-97
0.2	4:1	0.2	0.5	66	262	80	60-90
0.2	4:1	0.3	0.5	98	391	70	50-82
0.2	4:1	0.4	0.5	138	551	60	40-73
0.2	4:1	0.5	0.5	187	746	50	30-64
0.3	4:1	0.1	0.5	26	105	90	60-98
0.3	4:1	0.2	0.5	38	150	80	50-92
0.3	4:1	0.3	0.5	53	212	70	40-85
0.3	4:1	0.4	0.5	72	289	60	30-77
0.3	4:1	0.5	0.5	96	363	50	20-69

* Sample sizes calculated using Stata™ 's power functionality

3.10.2. Datasets and coding

Some study sites may not be able to collect all information proposed above. Study sites can indicate which variables they can collect and which data source they will use in the table below. The collected information can use the coding as in **Annex 1: List of variables collected, definition and coding**.

3.10.3. Data collection instruments

Data will be collected using a standardised questionnaire/data collection form. Some information may require follow-up. The source(s) of data may include:

- face-to-face/telephone interview
- electronic medical records
- interview with patient or his/her family
- vaccination and other registries
- laboratory

➢ *Each surveillance site to define the sources of information used for each variable collected (see also Annex 1)*

3.10.4. Data collection validation

A sample of paper questionnaires will be checked against the study database to validate data entry.

For GPs using electronic medical records, a sample of questionnaires are checked against the medical records and against the study database.

➢ *The specific validation procedures, including sample size calculation for questionnaire validation (if applicable) should be specified in the study annexes. Vaccination status, date, dose(s) if relevant and vaccine brand should be collected carefully and validated.*

3.11. Data management

3.11.1. Data collection, entry and storage at site level

Web-based data collection methods or paper-based methods can be used. Double data entry is recommended if paper forms are used.

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

Epiconcept provides the option of web-based data collection methods, if so desired by the sites: the Voozanoo web-based data entry platform, which is a secure system. These data can be accessed by the study site and the coordinating hub only. These methods can also be combined with paper-based methods.

If the Epiconcept web-based data collection methods are not used, data can be coded as outlined in Annex 1, but it is not required.

Information on antigenic, when available, and genetic analyses can be stored separately on an Excel spreadsheet (see Annex 2).

All data should be stored and processed in a way compliant with GDPR.

- ➢ *Study sites to specify procedures of data collection and entry*
- ➢ *Study sites to specify methods of data storage and their compliance with the GDPR requirements*
- ➢ *Study sites to provide a codebook that includes the variable names, variable descriptions, and the coding of variable values (see also Annex 1).*

3.11.2. Data anonymisation and persistent unique identifier

All data sent from the sites should be anonymised. This means that the case-based data sent to the coordinating hub and the data on the Voozanoo data entry web platform (for sites using it) should not include

- Any names of patients
- Any addresses of patients
- Any medical registration numbers
- Any telephone numbers, email addresses or other contact details of patients
- Any dates of births (age in years is OK)
- Any other (combination of) information that increases the risk of identification

If these types of data are included in the data, the coordinating hub will not use them and will delete them.

Each case-based record should have a unique identifier that the coordinating hub can use to identify a record when asking any questions to sites about data completeness or quality. This identifier should be persistent over the whole course of the surveillance/study (it should not change).

- ➢ *Surveillance sites to describe how and who performs the database anonymisation prior to local data analysis*

3.11.3. Data transfer, frequency of data transfer/reporting and storage at coordinating level

The frequency of reporting new data from study sites to the coordinating hub will initially be determined as more is known about vaccine availability and rollout. And the frequency may be revised according to COVID-19 incidence among sites participating and the recruitment strategy within primary care sites. This frequency will be reviewed on a regular basis.

For more information on data transfer, frequency of data transfer/reporting and storage at coordinating level, please see Annex 4.

3.11.4. Data checking and cleaning

Data checking will be carried out at site level, and also at pooled level by the coordinating team. Summary and frequency tables as well as visual representations of appropriate variables are used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies are carried out (e.g. date of swabbing before date of onset of symptoms). These values should be checked against the questionnaires or queried with the GP. Any missing data will be described.
Any changes or recoding (e.g. age to age groups) to the data during the cleaning process are documented and stored separately from the crude database. A guide and/or an example Stata do-file for data cleaning is provided if so desired.
At pooled level, questions arising after data checking will be queried with the sites using the unique identifiers, so records can be traced back whilst maintaining anonymity. Data checking is an iterative process (see Annex 3). Data cleaning (recoding) will only take place in agreement with the site.

3.12. Analysis

Each individual study site can analyse their data. The coordinating hub can provide example scripts if desired or carry out the site-specific data analysis at the site's request.

In a second step, a pooled analysis will be carried out. The higher sample size in the pooled analysis will provide more power (and precision).

Please see the detailed plan of analysis for site-specific and pooled analyses in Annex 5.

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

- sex
- age group
- health care worker status
- time: month of symptom onset
- vaccination status
- symptoms
- absence, presence of at least one, presence of more than one high-risk condition
- specific chronic conditions (e.g. respiratory, cardiovascular diseases)
- pregnancy
- influenza, pneumococcal and BCG vaccination status
- respiratory co-infections

In a second step, a univariable analysis will be carried out to measure the association between vaccination and being a laboratory-confirmed COVID-19 case.

A stratified analysis (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 (see also annex 5). Creating direct acyclical graphs may help better understand how the variables relate to each other and the outcome. In a final step, a multivariable analysis will be carried out to take confounding factors and potential effect modifiers into account. Please see annex 5.

3.12.1. Analysis with different control groups

Additionally, an analysis with different control groups will be carried out. The default control group will be those patients testing negative to SARS-CoV-2. Other control groups will include (where this information is available):

- Patients testing negative to all respiratory viruses
- Patients testing positive to respiratory viruses other than SARS-CoV-2 (e.g. rhinovirus, RSV, etc.; noting that patients with seasonal coronaviruses are excluded)

- Patients attending GP practices for reasons other than a respiratory infection (a traditional case control study) could also be considered.

3.12.2. Sensitivity analyses

As sensitivity analyses, we will measure VE

- with different cut-offs of numbers of days between onset and swabbing
- with different cut-offs of numbers of days between vaccination and onset of symptoms
- including and excluding those with previous positive tests, and also with different delays between previous test and enrolment in the current study.

- using only controls positive to other respiratory viruses

- excluding controls positive to seasonal coronaviruses

- including and excluding those fully vaccinated but with inappropriate gaps between doses.

3.13. Ethical considerations

Each surveillance site will comply with national ethics committee requirements. Where required, informed consent will be sought from all participants or legal tutors. The national ethics committees will specify whether oral, written, or no consent will be required. A copy of the ethical approvals should be sent to the coordinating centre.

- *Each site to describe the procedures to comply with the national ethics committee requirements and the type of informed consent needed as well as whether consent can be obtained for a legal tutor*

- *Each site to send a copy of the ethical approval to the coordinating centre*

3.14. Safety

During consultations and during the swabbing procedure, the safety of the practitioners is paramount. Any person swabbing, handling swabs and swabbing material, also in laboratories, should ensure that adequate personal protective equipment is used and hygiene measures followed.

- *Each surveillance site to state the safety measures carried out.*

3.15. Dissemination of results

Initial estimates will be disseminated as soon as possible at regular intervals and will be updated as frequently as possible and provided to key stakeholders at national, European and international level.

The results will be placed on the I-MOVE-COVID-19 website (https://www.imoveflu.org/i-move-covid-19/) with unrestricted access.

Reports and publications (in PDF) will also be uploaded onto the Zenodo platform as open access. Zenodo is a research repository launched in 2013 and hosted by CERN. It is GDPR-compliant and different access levels exist (https://about.zenodo.org/).

3.16. Data sharing

Anonymised data underpinning the reports will also be made publicly available on the Zenodo platform, along with a data codebook and scripts where possible. This will enable validation of the reports and ensure transparency and reproducibility. It will also enable other researchers to access and use the data for COVID-19 research. Site-specific data will only be shared openly with the site's consent.

The I-MOVE-COVID-19 data will also be made available on the EC data sharing platform, once the platform becomes more established. The EC data sharing platform has restricted access.

3.17. Publications, scientific communication

Results will only be published in open-source journals (this is a requirement of the European Commission's H2020 funding received for this surveillance project). Each study site is responsible for and free to publish their own results in open-source journals. Study site coordinators can decide which scientific conferences will be attended in order to present the results. An article presenting the results of the pooled analysis and will be submitted to an open-source, peer-reviewed journal.

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

I-MOVE-COVID-19 results will be shared widely with other H2020 project teams and the public, as required by the European Commission's H2020 "open data" policy.

3.18. Training

Investigators and data collectors will be trained on the study protocol before the start of the study. They will receive the protocol and questionnaires.

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

4. Logistical aspects

4.1.Study site leader

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

The National Centre for Microbiology, Instituto de Salud Carlos III, Madrid, Spain is in charge of compiling and summarising the genetic data from the study sites.

4.2.Human resources

In each site, an investigator will be in charge of monitoring data collection at the GP office level. GPs will collect the information among consulting patients. The specific human resources needed in each country are detailed in the study annexes. Epiconcept ensures the overall coordination of the various surveillance sites.

4.3.Supervision

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

5. Limitations

5.1.Potential biases

5.1.1. Unmeasured confounding

Observational studies can be hampered by confounding. The test-negative design used here may help overcome some of the unmeasured/difficult to measure confounding. Unmeasured confounding could include heterogeneity of exposures among unvaccinated and vaccinated (those groups may be different to each other in terms of high-risk behaviours). Statistical techniques to overcome the bias of unmeasured confounding in any exposure–outcome association in this analysis will be considered.

5.1.2. Representativeness of subjects included in the study

The study includes cases that are consulting GPs for COVID-19-like symptoms. Containment and mitigation strategies for the COVID-19 pandemic may differ by country depending on the case management strategy (e.g. recommendation of contacting a specific COVID-19 helpline, or consulting a GP or health centre by telephone first). In some cases, the management strategy will have an impact on which patients consult a GP and are swabbed. This also may have an impact on the time lag between onset and respiratory specimen collection, and currently we do not know if this may affect false negativity rates. Beside the collection of the aforementioned data in the protocol, case-containment/ mitigation / health care seeking strategies should be described for each country. Note that the test-negative design adjusts for case management strategies, e.g., patients with contact to a confirmed case, as both cases and controls come from this population.

- *Each site to describe the potential limitations in terms of representativeness of the subjects included*

5.1.3. Controls who are no longer at risk of disease

In this test-negative design, cases and controls are selected concomitantly. Controls may go on to be future cases, however at the time they are selected to be controls, they should be at risk of the disease. Patients presenting to the GPs with COVID-19-like symptoms and are thus swabbed, may test negative to SARS-CoV-2, but have had SARS-CoV-2 infection in the past. If this is the case, the control is no longer at risk of disease and should not be included in the study.

This study attempts to ascertain which controls may have had a past SARS-CoV-2 infection, by asking about previous SARS-CoV-2 tests and test results, as well as asking about previous guidance to quarantine/self-isolate if they had had contact with a case. However, among the controls, there could potentially be several patients with prior SARS-CoV-2 infection. The results will be interpreted in light of this and an estimate of a range of potential bias will be calculated around the VE estimates.

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

5.1.4. Performance of PCR tests

The clinical sensitivity of SARS-CoV-2 PCR tests approaches 80% [6], resulting in misclassification of SARS-CoV-2 infections, particularly in false negatives. If the misclassification is not differential by vaccination status, the VE will tend to be biased towards the null. If vaccination reduces viral load, then there is potential for a misclassification of SARS-CoV-2 cases, with a proportion of vaccinated cases falsely identified as vaccinated controls, biasing the VE away from the null.

In order to determine the impact on our VE estimates, we will consider simulation studies taking the clinical sensitivity of SARS-CoV-2 into account and assuming a proportion of vaccinated cases misclassified as controls due to potential effects of vaccination on viral load.

5.1.5. Pooled estimate and its bias

Any bias in the individual studies influences the pooled estimate. The power of the test for the presence of heterogeneity between individual studies is low if there are few studies. In this case, the test may not be able to detect heterogeneity between studies, despite it being present. It is important that heterogeneity is 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. Given the disruption of routine sentinel surveillance in many countries during the pandemic, surveillance systems and strategies are evolving over time at country level. Heterogeneity may be greater in this pandemic context than previously, not only including heterogeneity between sites, but over time within a single site.

6. References

1. European Centre for Disease Prevention and Control. COVID-19 situation update worldwide, as of week 5, updated 11 February 2021 [Internet]. 2021 [cited 2021 Feb 15]. Available from: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases

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

3. European Centre for Disease Prevention and Control. Vaccines and treatment of COVID-19 [Internet]. 2020 [cited 2020 Oct 12]. Available from: https://www.ecdc.europa.eu/en/covid-19/latest-evidence/vaccines-and-treatment

4. World Health Organisation. World Health Organization. Laboratory biosafety guidance related to coronavirus disease 2019 (COVID-19): Interim guidance [Internet]. 2020 [cited 2020 Apr 24]. Available from: https://apps.who.int/iris/bitstream/handle/10665/331138/WHO-WPE-GIH-2020.1-eng.pdf

5. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006 Jun;11(2):193–206.

6. He JL, Luo L, Luo ZD, et al. Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. *Respir Med*. 2020;168:105980. doi:10.1016/j.rmed.2020.105980

7. Annexes

Annex 1: List of variables, definitions and coding; I-MOVE-COVID-19 primary care-based COVID-19 VE study minimum dataset at site level

The following list of variables constitutes the proposed dataset for COVID-19 VE study at primary care level. Sites may not be able to collect all the proposed data and can list the variables collected in the study-specific annex.
Sites can follow this variable naming and coding, or are welcome to code variables and values in their own way and send a codebook along with their data.

- *VE sites can use the table below to indicate which variables they are collected and data sources*
- *VE sites to indicate all modifications in the variables collected and coding compared to variables below*

Variable name	Collected by study site? Please indicate also data source if not patient interview	Type	Values and coding	Definition
Study-related variables				
participate	☑	Numeric (binary)	0 = No 1 = Yes	Agrees to participate
refuse	☐	Text		Reasons for refusal to participate
id	☑	Type of variable at discretion of site	[needs to be unique]	Unique and persistent identifier for each record
gpcode	☑	Type of variable at discretion of site	[needs to be unique]	Unique identifier for each GP
Demographics				
age	☐	Numeric (continuous)	Integer	Age of each participant in years
sex	☐	Numeric (binary)	0 = female 1 = male	Sex of study participant
hcw	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient is a healthcare worker
height	☐	Numeric (continuous)		Height in cm (if BMI is not collected)
weight	☐	Numeric (continuous)		Weight in kg (if BMI is not collected)
Signs and symptoms				
onsetdate	☐	Date	dd/mm/yyyy	Date of onset of symptoms
fever	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Fever or feverishness
temp	☐	Numeric (up to one decimal)		Measured temperature
cough	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cough
shortbreath	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Weakness
anosmia	☐	Numeric (categorical)	0 = No 1 = Yes	Anosmia (Loss of sense of smell)

			8 = Do not know	
ageusia	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Aguesia/dysgeusia (Loss or distortion of sense of taste)
malaise	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Malaise
myalgia	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Myalgia
sorethroat	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sore throat
suddenonset	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Sudden onset
headache	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Headache
fatigue	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Fatigue
coryza	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Coryza or rhinitis
nausea	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Nausea
vomiting	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Vomiting
diarrhoea	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diarrhoea
chills	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chills/feverishness
chestpain	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Chest pain
lossapp	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Loss of appetite
stomache	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Stomach ache
conjunct	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Conjunctivitis
dizziness	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Dizziness
cyanosis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cyanosis
rash	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rash or other dermatological manifestation
palpitations	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Palpitations
no_symp	☐	Numeric (categorical)	0 = No 1 = Yes	The patient has no symptoms. This question is

				important if asymptomatic cases are included.
Swabbing/testing information				
swabdate	☐	Date	dd/mm/yyyy	Swabbing date
swabplace	☐	Numeric (categorical)	1 = GP practice 2 = COVID-19 centre 3 = Self-swabbing 4 = Swab at home by HCW 8 = Do not know	Place of swabbing
swab_type	☐	Numeric (categorical)	1 = Nose 2 = Throat 3 = Both nose and throat 8 = Do not know	Type of swab taken
test_type	☐	Numeric (categorical)	1 = PCR 2 = Point of care 3 = Other 8 = Do not know	Type of test used (if other, please specify)
lab_res	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Inconclusive 8 = Do not know	Laboratory result for SARS-CoV-2 (positive/negative)
geneticvariant	☐	Text		Genetic variant of SARS-CoV-2 virus (can be collected separately at different date)
Results for other respiratory pathogens				
lab_flu	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for influenza (positive/negative)
lab_rsv	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for RSV (positive/negative)
lab_metapneum	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for human metapneumovirus (positive/negative)
lab_rhinovirus	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for rhinovirus (positive/negative)
lab_adenovirus	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for adenovirus (positive/negative)
lab_bocavirus	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for bocavirus (positive/negative)
lab_seascorona	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for seasonal coronavirus (positive/negative)
lab_enterovirus	☐	Numeric (categorical)	0 = Negative 1 = Positive 2 = Not done 8 = Do not know	Laboratory result for enterovirus (positive/negative)
Vaccination variables				
covvaccany	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	COVID-19 vaccination status

covvaccdoses	☐	Numeric (categorical)	0 = 0 doses 1 = 1 dose 2 = 2 doses 8 = Do not know	COVID-19 vaccine doses
covvaccdate_firstdose	☐	Date	dd/mm/yyyy	COVID-19 vaccination date of first dose
covvaccdate_seconddose	☐	Date	dd/mm/yyyy	COVID-19 vaccination date of second dose
covvaccbrand_firstdose	☐	Text		Brand name of first dose COVID-19 vaccine
covvaccbrand_seconddose	☐	Text		Brand name of second dose COVID-19 vaccine
targetvacc	☐	Numeric (categorical)	1 = 1 dose 2 = 2 doses 8 = Do not know	Belongs to COVID-19 vaccine target group at time of swabbing
fluvaccany	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received flu vaccination in current season
fluvaccdate	☐	Date	dd/mm/yyyy	Influenza vaccination date
fluvacctype	☐	Text		Type of vaccine (brand name)
pneumovacc	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received pneumococcal vaccination
pneumotype	☐	Numeric (categorical)	1 = PPSV23 2 = PCV13 3 = Other (pls specify) 8 = Do not know	Type of pneumococcal vaccine
pneumotype_other	☐	Text		Other type of pneumococcal vaccine if not PPSV23 or PCV13
pneumoyear	☐	Number		Year of receipt of pneumococcal vaccination
bcgvacc	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Ever received BCG vaccine
bcgyear	☐	Number		Year of receipt of BCG vaccination
Underlying chronic conditions				
diabetes	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diabetes and endocrine disease
heart_dis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Heart disease (excluding hypertension)
hyperten	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Hypertension
immuno	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Immunodeficiency and organ transplant
lungdis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Lung disease
asthma	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Asthma
cancer	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cancer
obese	☐	Numeric (categorical)	0 = No 1 = BMI ≥30-39 2 = BMI ≥40 8 = Do not know	If height and weight are not collected: BMI ≥30-39; ≥40

renal_dis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Renal disease
liver_dis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Liver disease
rheum_dis	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rheumatological disease
Optional: Presymptomatic medication (medication taken at least 14 days before symptom onset)				
statin	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took statins
ace	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took angiotensin-converting enzyme inhibitors
arb	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took angiotensin II receptor blockers
nsaids	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took non-steroidal anti-inflammatory drugs
corticosteroids	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took corticosteroids
dmards	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient took biological disease-modifying anti-rheumatic drugs
chemo	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient has had current/recent cancer chemotherapy
antithrom	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	antithrombotic/ platelet aggregation inhibitors
metformin	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	metformin
Possible exclusion criteria				
antivir	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Administration of antivirals prior to swabbing
antivirdate	☐	Date	dd/mm/yyyy	Date administration antiviral
antivirtype	☐	Text		Type of antiviral (brand name)
res_home	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Living in a residential home
contra	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Contra-indication for COVID-19 vaccination
prevtest	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Has the patient had a positive SARS-CoV-2 test prior to this illness episode?
type_prev_test				
date_prev_test	☐	Date	dd/mm/yyyy	Date of positive SARS-CoV-2 test prior to this illness episode?
result_prev_test				
Other variables				
severity	☐	Numeric (count)	integer	Number of hospitalisations previous 12 months for the chronic disease

gpvisit	☐	Numeric (count)	integer	Number of GP consultations previous 12 months
pregnant	☐	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Pregnancy status
smoking	☐		0 = Never 1 = Former 2 = Current 9 = Do not know	Never, former (stopped smoking at least 1 year before inclusion in the study), current smoker (Any smoking can be included: cigarettes, cigars, vaping, etc.)

Annex 2: Genetic and antigenic analysis data (examples)

The minimum amount of data needed to obtain genetic data from GISAID (sequences of all viruses should be sent to GISAID's open access EpiCoV platform) is country, I-MOVE-COVID-19 ID number and GISAID accession number. Additional information on CT value and selection for characterisation and reasons for not characterising can be additionally collected (see Table 4).

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

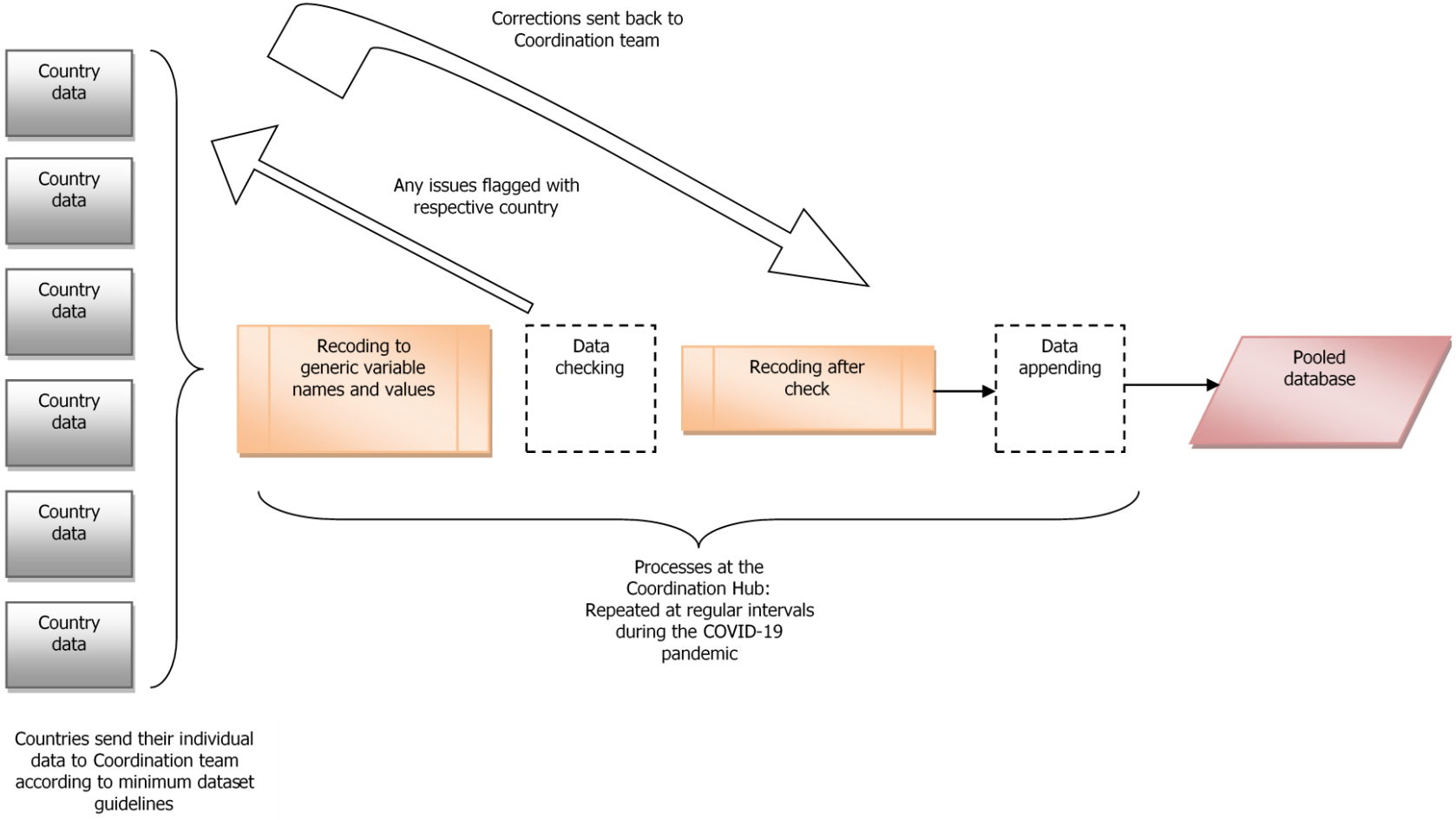
Table 4: Example of a data collection form for genetic data.

Where not all viruses were attempted to be sequenced, but only a random selection of them, additional information on sampling fraction should be provided, In order to better understand how viruses were selected for sequencing over time. An example can be seen in table 5.

Time period	**First date of time period**	**Last date of time period**	**Sampling fraction used**	**Date used for definition of time unit (onset date, swab date, other)**	**Comments**
1					
2					
Example 1	*01/10/2020*	*31/12/2020*	*1*	*Date of onset*	*(this is only an example; all specimens were characterised)*
Example 2	*01/01/2021*	*15/02/2021*	*0.2*	*Date of onset*	*(this is only an example; 20% of all specimens were characterised)*

Table 5: Example of documenting outlining how viruses were selected for sequencing over time

Annex 3. Data flow for pooled dataset



Annex 4: Data transfer, frequency of data transfer and data storage at pooled level

Software

For the multi-centre pooled analysis, study sites will send an anonymised database to the coordinating team through the secure data transfer system EpiFiles (https://epifiles.voozanoo.net), which is a web platform which allows secure file exchanges between entities. Each site has a login and password for the EpiFiles system. Only the coordinating hub will be able to access the site-specific files.

Frequency

The frequency of reporting new data from study sites to the coordinating hub for surveillance data will initially be monthly for the individual level enhanced surveillance. This will be revised to less frequent reporting according to COVID-19 incidence and the recruitment strategy within primary care sites. This frequency will be reviewed on a regular basis.

For sites using the Voozanoo platform, data will be downloaded on a monthly basis.

Study period of data to be transferred for individual level enhanced surveillance

Sites can send only new data to the coordinating hub each month, which will then be appended to previous data, or, if they prefer, they can send all data from study start.

For some study data there may be some changes to previous data (e.g. missing data completed, changes after data quality checks), therefore we recommend sending all data from surveillance start with each monthly transfer.

Data storage at pooled level

Please see also the I-MOVE-COVID-19 data management plan for more information (https://docs.google.com/document/d/1uflXrwOLIdIr_Y7jzGKCF-BX4SanuWI1OAVe2pfZJBc/edit).

All anonymised data received from study sites will be stored in a GDPR-compliant manner. Work package leaders and the coordinators will have access to the pooled data. The pooled data will be stored in G Suite (provided by Google). This environment is GDPR-compliant and secure and private: https://gsuite.google.com/security/?secure-by-design_activeEl=data-centers

Annex 5: Detailed analysis plan

Each individual study site can analyse their data. The coordinating hub can provide example scripts if desired or carry out the site-specific data analysis at the site's request.

In a second step, a pooled analysis will be carried out. The higher sample size in the pooled analysis will provide more power (and precision).

Descriptive analysis

The proportion of patients not consenting will be documented. Patients excluded will be described in a study flowchart.

Cases and controls will be described by baseline characteristics. An example layout of this is in table 6 below.

Table 6: 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
Healthcare worker	x/x (x)	x/x (x)
Missing	x	x
Days between onset of symptoms and swabbing		
0	x/x (x)	x/x (x)
1	x/x (x)	x/x (x)
2	x/x (x)	x/x (x)
3	x/x (x)	x/x (x)
4-7	x/x (x)	x/x (x)
COVID-19 vaccination	x/x (x)	x/x (x)
Missing	x	x
Etc.		

Patients will be described according to:

- sex
- age groups
- health care worker status
- urban/rural residence
- time: month of symptom onset
- symptoms
- absence, presence of at least one, presence of more than one high-risk condition
- specific chronic conditions (e.g. respiratory, cardiovascular diseases)
- pregnancy
- presymptomatic medication
- influenza, pneumococcal and BCG vaccination status
- respiratory co-infections
- referral to hospital or not
- travel and other exposures

Measure of effect

This study is a case control study (test-negative design). The measure of association is an odds ratio. This can be measured by logistic regression.

Vaccine effectiveness is computed as VE = (1 – OR)*100. A 95 % confidence interval is computed around the point estimate.

Stratified analysis

The analysis can be stratified according to (if sample size allows):

- age groups
- sex
- presence of at least one chronic condition;
- calendar time and time since vaccination

A sufficient sample size should be planned in order to ensure enough individuals in each stratum for a precise estimate. Effect modification should be assessed comparing the VE across the strata of the baseline characteristics. Confounding should be assessed by comparing crude and adjusted VE for each baseline characteristic.

Multivariable analysis

A multivariable logistic regression analysis will be conducted to estimate VE and 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, onset time (we select cases and controls concomitantly) should always be included in the model.

Note for the pooled analysis, as this is a multicentre study, study site should be included in each model, including the "crude" model.

Controlling for GP effect

Primary analysis will be carried out using standard logistic regression to obtain the individual study VE estimates. However, there could be variability between GPs. To adjust for this possible cluster effect, a multi-level logistic regression with each GP as a random effect will be carried out and compared to the single level analysis.

Variable selection and model specification

Model development strategy

To find a suitable model, we will consider very carefully the variables collected and determine which are GP level variables, which are individual level variables, which variables are intermediaries of each other and which variables are potential confounders and effect modifiers. Variables will also be checked for collinearity, and decisions will be made to include the group of collinear variables in the model or select amongst them.

The above considerations are particularly important for this study, as some of the medication collected and the chronic conditions of the patients may be strongly correlated.

Creating a direct acyclical graph may help better understand the relation between all variables collected and the outcomes.

Some variables will be a priori variables. These are variables that we want to keep in the model, as previous studies have shown them to be potential confounders or effect modifiers. These could include age and sex, but also potentially others.

If the model is not overfitted and variables are included that are not collinear or intermediaries, then there may be less concern for parsimony, as including insignificant variables may result in more accurate p-values for tests for variables of interest.

However, if sample size is low and the model is overfitted, then a backwards step-down variable selection procedure could be considered.

Interaction terms should be included cautiously, factors other than statistical significance (prevalence of exposure, magnitude of OR) will also be used as criteria for inclusion of an interaction term.

Several different models may have to be presented and considered.

Continuous variables

Continuous variables in the I-MOVE-COVID-19 datasets include age, date of onset of symptoms and number of GP visits and hospitalisations. 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 VE with the least bias.

If using restricted cubic splines to model continuous variables, the Stata programme “mkspline” can be used.

Output tables presenting ORs

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 7 can be used (variables presented just as an example of the output format). Useful information includes numbers of cases and controls and presentation of results for different models.

Table 7. Example table of odds ratios for different risk and predictive factors for COVID-19, primary care-based COVID-19 risk factor study, I-MOVE-COVID-19, 2020.

	VE	**(95%CI)**

Analysis scenarios, population included			
All ages	N (cases/ vaccinated; controls/ vaccinated)		
	Crude *		
	Adjusted for onset week		
	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 *		
0–14 years	N (cases/ vaccinated; controls/ vaccinated)		
	Crude *		
	Adjusted for onset week		
	Adjusted for sex*		
	Adjusted for chronic condition*		
	Adjusted for age (cubic spline)*		
	Adjusted for onset week, age (cubic spline)*		
	Etc		

* If pooled analysis, study site included as fixed effect.

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 odds:

- 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 exposure status

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.

- *Each study site to specify criteria used to determine minimum sample size if desired.*

Pooled analysis

For the pooled data, interim analyses will be conducted in different periods 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 COVID-19 consultations at primary care (stage of the pandemic and challenges at primary care level), the sampling strategy among GPs and the number of participating GPs in the study.

The pooled analysis will be carried out in a similar way to the site-specific analysis. Country or study site will be included potentially as a fixed effect or as a random effect in a multilevel model.

For key risk and preventive factors, heterogeneity between study sites will be determined. Any bias in the individual studies influences the pooled estimate. The power of the test for the presence of heterogeneity between individual studies will be low if the sample size per study site is small. In this case, the test may not detect the presence of heterogeneity, even if present. It is important that heterogeneity will also be 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 OR.

Statistical heterogeneity between studies will be tested using Q-test and the I^2 index (5). 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.

Study-specific crude and adjusted ORs and their confidence intervals will be plotted in separate forest plots. Study site characteristics will be assessed where feasible, such as information on health care use, organisation of the pandemic strategy. 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.

Annex 6: Study-specific annexes

Study specifications for each country are summarised in the annexes. Each surveillance site annex should include:

- description of the primary care practices participating in the study (number of GP practices, number of GPs, sampling strategy (all, systematic sample), information on sampling (face-to-face, self-swabbing, use of point-of-care tests, lack of PPE), catchment population)
- definition of beginning of pandemic
- list of variables collected (and coding if different from suggested coding)
- pandemic vaccines used
- vaccine status ascertainment method
- details on methods for data collection, data entry and data transmission
- data validation procedures
- laboratory issues (laboratory performing tests; tests used: PCR, culture, 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 GPs.