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

Risk factors for COVID-19 at primary care level in Europe: generic protocol

MAY 2020

V 02

I-MOVE-COVID-19 Network

WP4 coordinated by Epiconcept

Based on: current literature, I-MOVE primary care generic influenza vaccine effectiveness protocol 2019–2020

Table of contents

1. 6
   1.1. 7
2. 7
   2.1. 7
   2.2. 7
   2.3. 8
3. 8
   3.1. 8
   3.2. 8
   3.3. 8
   3.4. 8
   3.5. 9
   3.5.1. 9
   3.5.2. **Error! Bookmark not defined.**
   3.5.3. 9
   3.5.4. 9
   3.6. 9
   3.7. 10
   3.7.1. 10
   3.7.2. 10
   3.7.3. 11
   3.8. 11
   3.8.1. 11
   3.8.2. 12
   3.8.3. 12
   3.8.4. 13
   3.8.5. 13
   3.8.6. 13
   3.8.7. 13
   3.8.8. 15
   3.8.9. 15
   3.8.10. 16
   3.8.11. 16
   3.8.12. 17
   3.8.13. **Error! Bookmark not defined.**
   3.8.14. 17

3.8.15. 17
3.8.16. 17
3.8.17. 18
3.8.18. 18
3.9. 18
3.9.1. 18
3.9.2. 19
3.9.3. 19
3.9.4. 19
3.10. 19
3.10.1. 19
3.10.2. 20
3.10.3. 20
3.10.4. 20
3.11. 21
3.12. 21
3.13. 22
3.14. 22
3.15. 22
3.16. 23
4. 23
4.1. 23
4.2. 23
4.3. 23
4.4. 23
5. 23
5.1. 23
5.1.1. 23
5.1.2. 23
5.1.3. 24
5.1.4. 24
5.1.5. 24
6. 24
7. 26
Annex 1: List of variables, definitions and coding; I-MOVE-COVID-19 primary care-based risk factor study minimum dataset at site level 24
Annex 3: Genetic and antigenic analysis data (examples) 30
Annex 3. Data flow for pooled dataset 32

Annex 4: Data transfer, frequency of data transfer and data storage at pooled level 33
Annex 5: Detailed analysis plan 34
Descriptive analysis 34
Measure of effect 35
Univariable analysis 35
Stratified analysis 35
Multivariable analysis 35
Controlling for GP effect 36
Variable selection and model specification 36
Model development strategy 36
Continuous variables 36
Output tables presenting ORs 37
Minimum sample size 37
Pooled analysis 38
Annex 6: Study-specific annexes 39
Annex 7: Summary of study on risk factors for hospitalisation and death among COVID-19 patients seen at primary care 40

Version history

Version	Date finalised	Created/modified by	Comments
1.0	2020-04-27	Epiconcept/Nivel	Initial draft sent to partners
2.0	2020-05-31	Epiconcept/Nivel	Next version incorporating partners' comments

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	Real-time 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 causes coronavirus disease 2019 (COVID-19). As of the 7th of April 2020, there were over 1.3 million cases of COVID-19 reported globally, with around half (608500) from the European Union/European Economic Area (EU/EEA) countries and the UK, and over 50000 of them have died. As of the 8th of April, the European Centre for Disease Prevention and Control (ECDC) reports that there is no indication at EU/EEA level that the peak of the epidemic has been reached.[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]

There are several questions to which urgent answers are needed to improve our understanding of SARS-CoV-2, to inform us of the best interventions to prevent or delay the spread of COVID-19 and newly recommended treatment strategies.

One question is which are risk factors and preventive factors for COVID-19 infection. This question can be answered at different levels, including for cases not presenting to health care systems (community cases), cases presenting at primary care level and cases presenting at hospital level. This protocol addresses risk and protective factors for those presenting at primary care level. Additionally, determining which symptoms and symptom combinations are predictive for a COVID-19 infection may help inform a public health response strategy. This protocol also includes two extra questions for a pilot study to understand if within the I-MOVE-COVID-19 data collection system, we can easily collect follow-up data on hospitalisation (and outcome) of a COVID-19 case seen at primary care level. These questions would inform a study on risk factors for hospitalisation among COVID-19 patients seen at primary care level.

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

[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.

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 2020 I-MOVE-COVID-19 European protocol for the risk factor study for COVID-19 at primary care level. The specificities of each site's COVID-19 data collection can be detailed in the individual site protocol annexes. This protocol can be used also for measuring vaccine effectiveness of COVID-19 vaccine, when it is available. A further protocol being developed is for risk factors for COVID-19 in healthcare workers (HCWs).

It is important to note that this risk factor protocol is a constantly evolving document. As more research data becomes available, the protocol will be adapted accordingly.

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

1.1.General principle

The I-MOVE network is nested in sentinel structures carrying out surveillance for influenza and now COVID-19. I-MOVE practitioners are used to carrying out enhanced surveillance to collect data for influenza vaccine effectiveness studies.

The current I-MOVE influenza vaccine effectiveness protocol has been adapted so that I-MOVE-COVID-19 sites can use it for enhanced COVID-19 surveillance. This enhanced surveillance also forms the basis for collecting data for this risk factor study for COVID-19 at primary care level. All information collected as part of the enhanced surveillance will be collected for the risk factor study as well. Additional information will be collected and controls sought for the risk factor study. The same protocol can also be used for influenza vaccine effectiveness (as usual), and also COVID-19 vaccine effectiveness estimation once the vaccine becomes available.

As more research on COVID-19 comes in, the questions in the risk factor study can be modified, deleted or added to (e.g. pre-symptomatic medications and chronic conditions).

2. Objectives

2.1. Primary objective

The primary objective will be to identify key risk factors for and protective factors against COVID-19 among patients presenting at primary care level to a GP including the following categories

- Patient demographics
- Underlying conditions
- Pre-symptomatic medication
- Vaccinations

in order to inform public health response and strategies.

2.2. Secondary objectives

The secondary objectives are to

- describe SARS-CoV-2 genetic clades overall and by time
- identify combinations of signs and symptoms that could identify COVID-19 cases early and increase the performance of case definitions (in terms of sensitivity and positive predictive value).

2.3.Additional objectives

An additional objective is to identify risk and protective factors for a hospitalisation for COVID-19 (more severe disease) among COVID-19 patients, in order to better protect vulnerable people.

This objective will require a modification of the classical I-MOVE test-negative study design as the outcome of hospitalisation may not be known at time of consultation of the patient. Therefore an element of follow-up is needed to provide the information on hospitalisation. This will be piloted as part of the I-MOVE-COVID-19 study, in order to understand if it is feasible to collect this information.

This study will be nested within the test-negative design risk factor study. All methods in the main text below pertain to the risk factor study. Annex 7 outlines briefly the methods of the study to identify risk and protective factors for a hospitalisation for COVID-19 (more severe disease) among COVID-19 patients.

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

3. Methods

3.1. Study design

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

3.2. Study population

The study population comprises community-dwelling individuals with COVID-like symptoms who consult a participating physician.

In the future an extension of this protocol to asymptomatic patients could be considered. Currently the protocol pertains to symptomatic cases. If you are collecting information on asymptomatic cases, please specify this clearly.

- 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 as soon as sites are able to implement the study. If feasible, retrospective data can be collected. Participating primary care practices carry out the study throughout the year.

- Surveillance sites to define the beginning of the study period (date/month/year)

3.4. Outcomes

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

A secondary outcome of interest is the genetic clade of SARS-CoV-2.

3.5.Case 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:

- Acute onset of at least one of the following symptoms: fever or cough or sore throat or shortness of breath or coryza
  or
- a clinician's judgment that illness might be due to a SARS-CoV-2 infection.

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

A **COVID-19 negative patient** 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 surveillance 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.

---

[3] For enhanced surveillance, 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)

[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]

The ECDC-recommended SARS-CoV-2 laboratory confirmation is by viral RNA detection with nucleic acid amplification tests, such as RT-PCR[1], [3]. 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[1].

Information will be collected on type of test, including RT-PCR and point-of-care tests. If point-of-care tests will be used, they should be molecular point-of-care tests only and be validated. A sensitivity analysis will be carried out including and excluding those patients whose results were obtained using a point-of-care 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 3). Processed genetic information, e.g. name of genetic clade, can also be included within the epidemiological database.

- ➢ *Each study site to describe the laboratory procedures (samples taken, storage, transport)*
- ➢ *Each study site to describe the tests and the kits used (and their sensitivity, specificity, PPV) for COVID-19 and, if needed, other respiratory virus detection*
- ➢ *Each study site to describe if the laboratory participates in QA/QC (Quality Assurance/Quality Control) schemes*
- ➢ *Each study site to describe the selection of specimens and the 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 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 case definition and consent to participate (the patient or her/his legal guardian gave consent to participate according to the local ethical review process.

- ➢ *Oral informed consent or written informed consent according to country procedures, as specified in the study annexes.*

3.7.3. Patient exclusion criteria

Patients are excluded in general 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;
- are swabbed more than a certain number of 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 ≤xx 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);
- had an inconclusive RT-PCR test;
- are a current control (SARS-CoV-2 negative) and had tested positive (by PCR or serology) to SARS-CoV-2 in the past xx months (the exact number of months will be determined as more research comes in).

Reasons for exclusion are documented.

Further reasons for exclusion may include:

- having received certain medication/treatments prior to swabbing (that may lead to false negatives, e.g. Plaquenil)

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.

For specific sub-analyses on vaccination as a risk and preventive factor, we also collect information on contra-indications for vaccination. Those with contra-indications for vaccination will be excluded prior to any analysis of vaccination as a risk or preventive factor.

Additionally, as some patients may have had a previous SARS-CoV-2 infection, but were not tested, we will ask about having been ever been recommended to quarantine/self-isolate due to being a contact of the case in the past and at what date (see next section). A sensitivity analysis will be carried out including and excluding these patients.

3.8.Study data to be collected

The following data will be collected as part of the study. Many of these data are collected already as part of the I-MOVE-COVID-19 surveillance in WP2. The additional data are highlighted in purple.

Some of the following data that are collected are potential risk and preventive factors. Others are potential confounders or effect modifiers. Some are both risk/preventive factor and potential confounder/effect modifiers. Some data collected are related to exclusion criteria.

3.8.1. Patient characteristics

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

- Age in years
- Sex

- Level of urbanisation of place of residence of patient, based on GP practice as a proxy, see paragraph below
- Deprivation score, based on GP practice as a proxy, see paragraph below
- 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 surveillance 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 with contact to patients, porters and cleaners.

Level of urbanisation

Level of urbanisation of place of residence may be a risk factor for COVID-19. We suggest collecting urban/rural information according to the Eurostat NUTS3 regions (https://ec.europa.eu/eurostat/web/rural-development/methodology), with three classifications: predominantly rural, intermediate, predominantly urban.

Study sites can collect this information in several ways:

- Asking the GP to provide the classification of urban/rural directly, upon specifications.
- Obtaining the GP practice postcode and then translating it to the appropriate urban/rural classification.

Deprivation score

Deprivation may also be a risk factor for COVID-19. Where available, we suggest countries to abstract from the GP postcode the European deprivation index for easier standardisation across countries (https://academic.oup.com/eurpub/article/28/suppl_4/cky213.625/5191925).

If this is not available, then a country-specific "indicator of deprivation" can be used. This will be discussed as the protocol further evolves.

3.8.2. Information on consultation

- Type of consultation: In practice, video, telephone, home
- Date of consultation

3.8.3. Clinical signs and symptoms

We will collect information on symptoms to better understand the clinical spectrum of disease. It is also important for the VE studies to collect symptoms as completely as possible, in order to have a stricter case definition in terms of symptoms. As a minimum:

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

- sore throat
- fatigue
- myalgia
- malaise
- coryza, rhinitis
- chest pain
- chills

If possible, sites could also collect:

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

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

3.8.4. 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

3.8.5. Referral to hospital

In order to get a measure of severity of the case, we will include the question:

- Did you recommend the patient be referred to hospital?

3.8.6. New pilot questions on outcome/severity of illness episode

The following questions differ from the "referral to hospital" question in 3.8.5, as a component of follow-up is needed for these questions. Study sites can pilot these questions, if possible:

- As part of this illness episode, was this patient hospitalised?
- Did the patient die as part of this illness episode (within 28 days of symptom onset; the number of days of follow-up to be reviewed regularly)?

3.8.7. 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 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 (optionally sites can 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, this is preferred. If not possible, we suggest categories (BMI: 30–39 and ≥40).

3.8.8. Pre-symptomatic vaccination status

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

- Seasonal influenza vaccination from the most recent influenza season (with date, if possible)
- Latest pneumococcal vaccination type (with year if possible)
- Bacille Calmette Guérin (BCG) vaccination (ever; with approximate year, if possible)

In addition, once available, information will be collected on COVID-19 pandemic vaccination including number of doses, date, and product.

Vaccination status ascertainment

Any prior BCG vaccination, as well as 2019/20 vaccination against influenza and vaccination against pneumococcal diseases will be collected. This may not be possible for BCG vaccination in some sites, but instead these could indicate if there was no BCG vaccination policy at all, or if e.g. “assume that those born after 1956 would be vaccinated”, etc.

The sources of information for vaccination may include:

- vaccination registry
- consultation of the patient’s vaccination card
- GP electronic medical records
- self-report
- data from the patient’s insurance company showing evidence of pharmacy delivery or reimbursement of influenza vaccine during the current influenza season.

3.8.9. Pre-symptomatic medication status

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".

3.8.10. 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.

3.8.11. Information on previous SARS-CoV-2 infection

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

- patient had previous positive COVID-19 test (yes/no/unknown)
- type of test: PCR, point-of-care test, serology
- date of test

Some patients consulting their GP may have had a SARS-CoV-2 infection in the past without having had a test. We will also ask:

- Have you ever been quarantined/asked to self-isolate due to being a contact of a confirmed COVID-19 case?
- If yes, date.

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

3.8.12. Functional status

Low functional status is defined as needing help to bathe or to walk.

3.8.13. Health care utilization 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 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

3.8.14. Travel

To determine potential place of infection, we will ask about the patient's recent travel (this question will be more relevant after social distancing measures end):

- Have you travelled outside this country in the past 14 days?
- If yes, which country?

3.8.15. Ethnicity

Ethnicity has been identified as a risk factor for COVID-19 in some studies. Collecting ethnicity data can be sensitive to collect and difficult to harmonise. This will be included as a pilot variable and revised in the future.

➢ *Sites to describe if and how they will collect ethnicity data.*

The following exposure-related questions (from contact to a confirmed case to protective factors) are entirely optional:

3.8.16. Contact to a confirmed case

Have you had close contact with a confirmed COVID-19 case (<2 metres) in the past 14 days?

If yes, in which setting?

- home
- education/school/day care
- workplace
- healthcare setting (GP, hospital)
- other (please specify)

3.8.17. Exposures in the past 14 days

- In the last 14 days have you been out of the house and in environments of 2+ people not part of your household?
- If yes, which setting?
  - education/school/day care
  - workplace
  - public transport
  - healthcare setting (GP, hospital)
  - at a leisure setting (shopping, walk in parks, restaurants/bars/pubs, social gathering, sport, etc.)
  - Other

3.8.18. Household setting

- How many members are in your household?

3.8.19. Protective factors

- Which precautions have you taken in the past 14 days (select as many as are appropriate)?
  - face/nose/mouth protection (mask/scarf wearing)
  - extra handwashing
  - use of hand disinfectant
  - home disinfection
  - home isolation (I have not left my home)
  - social distancing
  - other

3.9.Data

3.9.1. Sample size

The number of individuals included in the risk factor 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 COVID-19. Sample size will also depend on length of time in the study.

A sufficient sample size should be obtained in order to ensure enough cases and controls in each stratum for a precise estimate. The following sample size calculation provides an estimate of the sample size required to obtain a statistical significance at the 5% level with 80% power. First, the proportion of controls with the risk/protective factor in question should be estimated), and a minimum expected odds ratio (OR) to be determined (or determined from previous literature).

For simplicity's sake, the sample size calculation below assumes a ratio of 1:1 of cases and controls (Table 2).

For example, if the prevalence of a given risk factor is 50% in controls, we would need at least 137 COVID-19 cases and 136 controls in each of the strata to be able to detect an OR of at least 2 with 80% statistical power and a 95% confidence level, with a ratio of 1:1 of cases to controls (Table 2). Note that Table 2 provides the sample size for a univariable analysis; for a multivariable analysis, the sample size would be greater.

The pooled analyses should not prevent study teams from including a big enough sample size to obtain exact estimates for each separate study.

Note that sample size is calculated differently for the vaccine effectiveness study.

Table 2. Calculated sample sizes* for different levels of expected odds ratio (OR) by varying prevalence of exposure in controls, assuming 95% confidence and 80% power

	Prevalence of exposure in controls						
OR	**5%**	**10%**	**25%**	**50%**	**75%**	**90%**	**95%**
0.4	619	314	134	81	85	151	272
0.6	1615	833	372	247	288	554	1020
0.8	7357	3841	1785	1267	1597	3213	6018
1.5	1687	911	466	3887	571	1260	2433

2	516	283	152	137	215	492	962
2.5	272	151	85	81	134	314	619
3	177	100	58	58	100	240	475
3.5	129	74	44	46	81	199	397
4	101	58	36	39	71	175	349
4.5	83	48	31	34	64	158	316
5	70	41	27	31	58	146	293

* Sample sizes calculated using Stata™ 's power functionality

3.9.2. 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. Variables highlighted in purple are new to the risk factor study (not collected as part of the surveillance). The collected information can use the coding as in **Annex 1: List of variables collected, definition and coding**.

3.9.3. 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)*

3.9.4. 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) are specified in the study annexes.*

3.10. Data management

3.10.1. 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 unless electronic medical records 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).*

3.10.2. 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 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*

3.10.3. 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 for surveillance data will initially be monthly. This maybe revised to less frequent reporting 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 5.

3.10.4. 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.

3.11. 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 and potential risk and preventive factors. 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
- follow-up status (if available)
- travel and close contact exposures

In a second step, a univariable analysis will be carried out to measure the association between an exposure 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.

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 determine risk and protective factors taking confounding factors and potential effect modifiers into account. Please see annex 5.

3.12. 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*

3.13. 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.*

3.14. Dissemination of results

Initial OR estimates will be disseminated at regular intervals (see section 4.4) and will be updated as frequently as possible. A final overall OR when the pandemic is over. (Note that this may be revised depending on how the pandemic progresses.)

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

This report (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/).

3.15. Data sharing

The data underpinning the reports will alsobe 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.

In the future, ECDC will also create a record type in "The European Surveillance System" (TESSy) specifically for I-MOVE-COVID-19 data. This will be for surveillance purposes, and will include cases only (not controls). ECDC and Epiconcept will draw up and sign an agreement setting out the mutual roles and responsibilities of Epiconcept and ECDC in relation to the management of the case-base data (Article 29 of Regulation 2018/1725). Epiconcept will upload the I-MOVE-COVID-19 data on behalf of all sites, and sites will have access to the I-MOVE-COVID-19 data on TESSy.

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 is a platform with restricted access.

3.16. Publications, scientific communication

Results of the individual studies should only be published in open-source journals (this is a requirement of the European Commission's H2020 funding received for this surveillance project). 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.

3.17. 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.

4.4.Reports

There are two study reports that are deliverables for the I-MOVE-COVID-19 project. These should include data from all sites. They are due for submission at month 12 and 24 of the project, corresponding to the 15th of March 2021 and the 15th of March 2022.

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. 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 cases 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 negativityrates. Beside the collection of the aforementioned data in the protocol, case-containment/ mitigation / health care seeking strategies should be described for each country.

- *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 salient risk/protective factors.

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

5.1.4. Swabbing of asymptomatic COVID-19 cases

This protocol does not include asymptomatic COVID-19 cases. It is recommended that GPs swab those patients presenting with COVID-19-like symptoms. An extension of this study can be included in a future version to include also asymptomatic cases.

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 a risk/protective factor and the outcome.

6. References

[1] European Centre for Disease Prevention and Control (ECDC), World Health Organization Regional Office for Europe (WHO/Europe), 'Rapid risk assessment: Coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK – eighth update', European Centre for Disease Prevention and Control. Accessed: Apr. 22, 2020. [Online]. Available: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-eighth-update.

[2] M. Valenciano, B. Ciancio, and I-MOVE study team, 'I-MOVE: a European network to measure the effectiveness of influenza vaccines', *Euro Surveill.*, vol. 17, no. 39, Sep. 2012.

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

[4] T. B. Huedo-Medina, J. Sánchez-Meca, F. Marín-Martínez, and J. Botella, ‘Assessing heterogeneity in meta-analysis: Q statistic or I2 index?’, *Psychol Methods*, vol. 11, no. 2, pp. 193–206, Jun. 2006, doi: 10.1037/1082-989X.11.2.193.

7. Annexes

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

The following list of variables constitutes the proposed dataset for I-MOVE-COVID-19 risk factor study at primary care level. Sites may not be able to collect all the proposed data and/or may wish to collect other data. Sites 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.
The additional variables compared to those suggested for the surveillance are highlighted in purple.

- *Surveillance sites can use the table below to indicate which variables they are collected and data sources*
- *Surveillance 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
gppostcode	☐	Text		Gp postcode. This can be used for determining deprivation score and urban/rural. If not available, the "urban" variable below can be used. Sites should translate gppostcode to urban/rural and deprivation score at site level and should not send GP postcode information to the central hub.
urban	☐	Numeric (categorical)	0 = Predominantly rural 1 = Intermediate 2 = Predominantly urban	This variable can code for level of urbanisation, using GP postcode as a proxy.

			8 = Do not know	
hcw	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient is a healthcare worker
ethnicity	□	Numeric (categorical)		Collection to be determined according to country-specific guidelines
height	□	Numeric (continuous)		Height in cm
weight	□	Numeric (continuous)		Weight in kg
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
malaise	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Malaise
myalgia	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Myalgia
cough	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cough
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
shortbreath	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Weakness
anosmia	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Anosmia (Loss of sense of smell)
ageusia	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Aguesia/dysgeusia (Loss or distortion of sense of taste)
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	Diarrhoea

			8 = Do not know	
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.
Outcome				
hosp_refer	□	Numeric (binary)	0 = No 1 = Yes	Referral of patient to hospital for COVID-19 as part of consultation
hosp	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Pilot question: Among those positive for COVID-19, were they hospitalised as part of this disease episode?
death	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Pilot question: Among those positive for COVID-19, did they die as part of this disease episode?
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)
clade	□	Text		Genetic clade 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)
Vaccination variables				
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
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 influenza 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?
prevtest_when	□	Date	dd/mm/yyyy	Date of positive SARS-CoV-2 test prior to this illness episode?
everquar	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Has the patient ever been asked to quarantine/self-isolate due to being a contact of a confirmed case?
everquar_when	□	Date	dd/mm/yyyy	Date of quarantining/self-isolation due to being a contact of a confirmed case
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
fs_bath	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Requires assistance to bath
fs_walk	□	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Requires assistance to walk
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.)
Exposure-related variables				
travel		Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Has the patient travelled in the past 14 days?
travel_country		Text		Country of travel
contact_case		Numeric (categorical)	1=home setting 2=education/school/daycare 3=workplace 4=healthcare setting (GP, hospital) 5=other setting 8=unknown	Has the patient had close contact with a COVID-19 case (<2 metres) in the 14 days before symptom onset?
exposure		Numeric (categorical)	1=education/school 2=workplace 3=public transport 4= healthcare setting (GP, hospital) 5=leisure setting 6=other setting 8=unknown	In the last 14 days have has the patient been out of the house and in environments of 2+ people?
hh_number		Numeric (integer)	Integer	How many people are in the patient's household?
protective		Numeric (categorical)	1=face/nose/mouth protection (mask/scarf wearing) 2=extra handwashing 3=use of hand disinfectant 4=home disinfection 5=home isolation (I have not left my home) 6=social distancing 7=other 8=unknown	Which precautions has the patient taken in the 14 days preceding symptom onset?

Annex 3: 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							

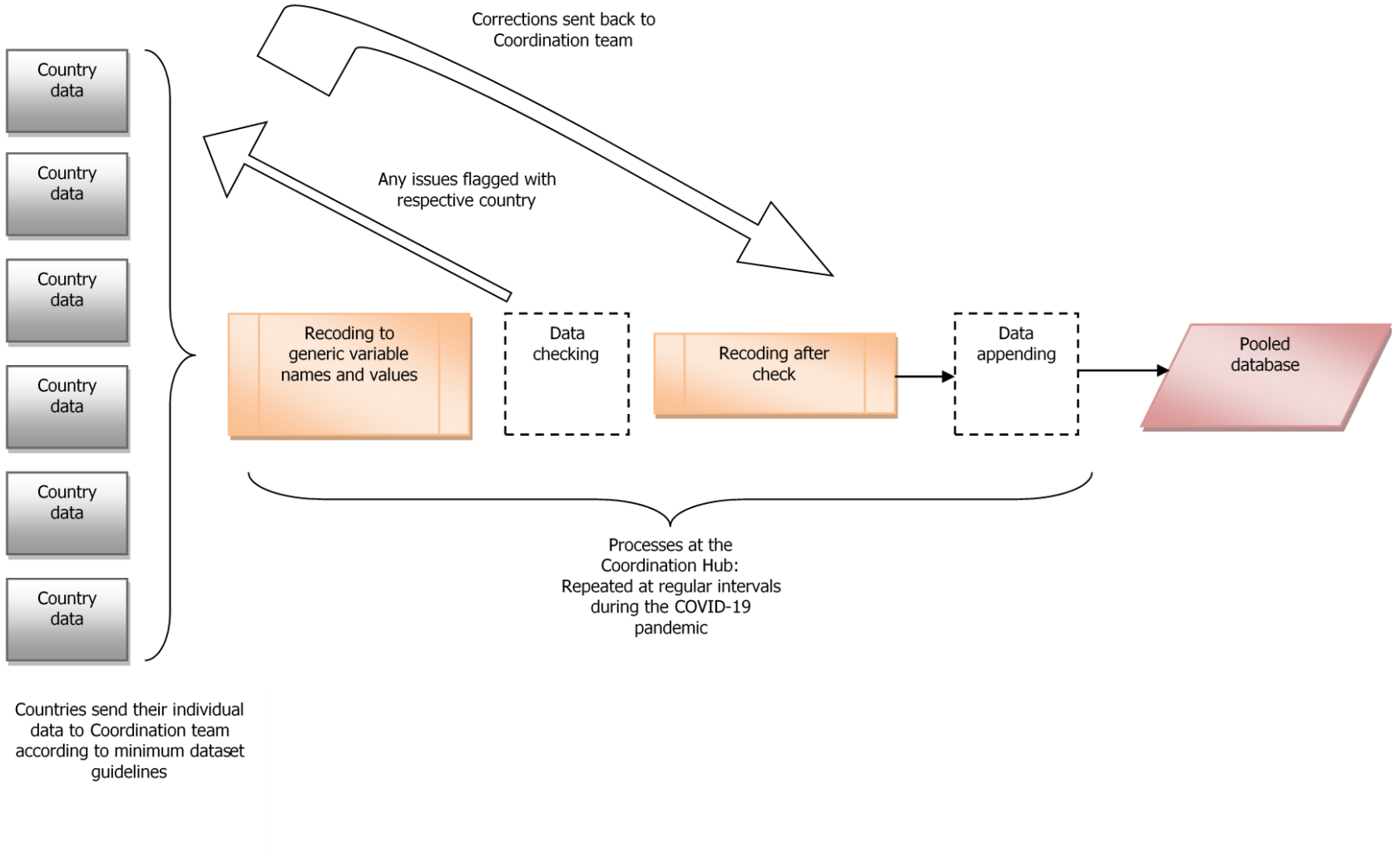
Table 4: Example of a data collection form for genetic data.

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)

Current season influenza 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. An odds ratio of 1 indicates no association between an exposure and the outcome. An odds ratio greater than one indicates a potential risk factor, an odds ratio lower than one indicates a potential protective factor, noting that the confidence interval around the odds ratio helps interpret the odds ratio.

For vaccination as preventive factors, a vaccine effectiveness can be computed as VE = (1 – OR)*100. A 95 % confidence interval is computed around the point estimate.

Univariable analysis

For each potential risk and protective factor the odds ratio and its 95% CI will be calculated.

Stratified analysis

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

- age groups
- sex
- presence of at least one chronic condition;
- time: by month, groups of months or early, middle and late in the pandemic

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 OR across the strata of the baseline characteristics. Confounding should be assessed by comparing crude and adjusted OR for each baseline characteristic.

Ideally this should be done for every relevant risk and protective factor.

Multivariable analysis

A multivariable logistic regression analysis will be conducted to estimate the associations between risk/preventive factors 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, for onset time (we select cases and controls concomitantly) should always be included in the model.

Controlling for GP effect

Primary analysis will be carried out using standard logistic regression to obtain the individual study 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 will 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. Potential risk/preventive factors that are insignificant are of interest themselves.

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 date of admission to hospital. 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 OR with the least bias. If the continuous variable is itself a risk factor or preventive factor, interpretation may be easier if coded in categories.

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.

Population included	**Potential risk factors/preventive factors**		**OR**	**(95%CI)**
All ages		N = xxxx		
	Age group	0-14 y	ref	
		15-44y		
		45-64y		
		65+y		
	Sex	Male		
	Influenza vaccination	Yes		
	HCW	Yes		
	Travel in past 14 days before symptoms.	Yes		
	Etc.			
Aged 65+		N=xxxx		
	Sex	Male		
	Influenza vaccination	Yes		
	HCW	Yes		
	Travel in past 14 days before symptoms.	Yes		
	Etc.			

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[4]. 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 (when applicable) 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.

Annex 7: Summary of study on risk factors for hospitalisation and death among COVID-19 patients seen at primary care

This risk factor study also includes two pilot questions to help us understand if a study on risk factors for hospitalisation among COVID-19 patients seen at primary care level is feasible.

The objective of this study is to identify early risk and protective factors for poor outcomes to better understand

- who could benefit from early interventions
- which target groups need particular preventive measures.

This study is complementary to the I-MOVE-COVID-19 study on risk factors for severe outcomes among hospitalised patients.

The study on risk factors for hospitalisation and death among COVID-19 patients seen at primary care is a cohort study nested within a test-negative case control study. Among the cohort of laboratory-confirmed COVID-19 patients seen at primary care level, the severe outcomes of the illness episode will be sought: hospitalisation and/or death.

Among this cohort of confirmed COVID-19 patients, early risk and protective factors for hospitalisation and death will be measured using multivariable regression models.

Adding questions on follow-up of patients (were they hospitalised/did they die as part of this illness episode?) to the risk factor for COVID-19 at primary care level protocol (the main protocol of this document) will enable us to understand

- the feasibility of collecting this information
- the expected level of completeness and
- the study duration needed to achieve a high enough sample size.