Protocol for primary care test-negative design case-control studies to measure influenza vaccine effectiveness

in the European Union and

European Economic Area Member States

Marta Valenciano, EpiConcept

Esther Kissling, EpiConcept

Alain Moren, EpiConcept

July 2009, updated November 2019

Table of contents

Abbreviations

ECDC European Centre for Disease Prevention and Control
GP General Practitioner
ILI Influenza-like illness
MS Member States
OR Odds ratio
PCV Proportion of cases vaccinated
PPV Proportion of population vaccinated
VC Vaccination coverage
VE Vaccine effectiveness

☑ (Tick/check mark indicates the sections that Member States should adapt and provide details for in their study annexes.)

I) Background

Influenza viruses are the only vaccine preventable viruses that undergo frequent genetic and antigenic changes. As a consequence, the influenza vaccine is reformulated each year and annual revaccination is recommended. Observed vaccine effectiveness (VE) varies from year to year, between population subgroups (e.g. age- and risk-groups) and varies according to the measured outcome (laboratory confirmed influenza virus by (sub)type/clade or clinical outcome). VE may vary between vaccine types and products, by time since vaccination and according to previous influenza and vaccine history.

Conducting annual influenza VE estimates at the European level right at the beginning of a seasonal influenza epidemic/pandemic and monitoring VE along the course of the epidemic/pandemic is crucial in order to:

- decide on recommendations for the use of the current season vaccine;
- target complementary or alternative public health measures (e.g. antivirals) for segments of the population where the vaccine is less effective;
- allow more precise estimates of the impact of current vaccination strategies on the burden of disease to support vaccination campaigns;
- trigger further investigations on seasonal and pandemic vaccines (improve composition, use of adjuvants, different doses, need for booster doses);
- better manage and respond to expected reports of vaccine failures (especially during a pandemic);
- guide the selection of strains to be included in the next seasonal vaccine; and
- counterbalance the reports of adverse events following immunisation by providing elements for adequate risk management and cost-effectiveness analysis.

Currently, observational studies are the most frequently used method to provide estimates of influenza VE in Europe.

In case of an influenza pandemic, an established EU platform to rapidly measure influenza VE by vaccine type and product will allow the evaluation of any pandemic vaccine and the adaptation of preventive and control strategies.

I-MOVE (Influenza Monitoring Vaccine Effectiveness in Europe) was the first network to monitor influenza VE within and across the seasons in the EU and the European Economic Area (EEA). I-MOVE

was established in 2007 with European Centre for Disease Prevention and Control (ECDC) funding[1]. One component of the I-MOVE network is the multicentre case control study at primary care level.

Besides measuring VE by influenza (sub)type early and late in the season and by vaccine type, these studies can be used to help answer further research questions, such as what the effect is of previous influenza vaccination on current season influenza infection, to what extent may influenza VE decline within the season, what effect does first influenza infection have subsequent vaccine effectiveness and to measure VE against different influenza clades or genetic variants.

This document presents the core European protocol for this study, outlining the agreed methods for measuring influenza VE for individual study sites as well as a pooled analysis. The protocol is based on results from the I-MOVE pilot studies in 2008–09, on knowledge gained from the 2009 pandemic, and from I-MOVE studies carried out from 2008–09 to the present (at time of writing: 2018–19).

This protocol is written in a generic manner and country-specific details of each study will be outlined in the study annexes (Annex 8).

.

[1] Valenciano M, Ciancio BC, on behalf of the I-MOVE study team. I-MOVE a European network to measure the effectiveness of influenza vaccines. Euro Surveilance.2012;17(39):pii=20281.

2) Objectives

2.1 Primary objective

The primary objective will be to measure, in EU/EEA countries at primary care level, early and late in the season, the direct effect (effectiveness) of seasonal (or pandemic if applicable) influenza vaccines against laboratory confirmed (PCR positive) influenza by (sub)type using a pooled analysis.

2.2 Secondary objectives

- To estimate seasonal (or pandemic if applicable) VE:
  - in each of the participating countries;
  - by risk groups,
  - by age groups;
  - by influenza vaccination target group;
  - by influenza (sub)clades and genetic variants.

- To measure VE by vaccine types (e.g. adjuvanted vs. non-adjuvanted, high-dose vs. normal dose, trivalent vs. quadrivalent, live-attenuated vs. inactivated, egg-based vs. cell-based), groups of vaccines (split virion, subunit, etc.) and brand;
- To estimate VE for one or two doses of seasonal (or pandemic) vaccine, if applicable;
- To understand the factors affecting influenza VE: duration of protection, the role of repeated seasonal vaccinations; childhood exposure to influenza virus/antigenic clusters;
- To identify key influenza virus phenotypic or genotypic evolutions that could affect vaccine performance and to estimate VE against specific clades.

3) Methods

3.1 Study design

- Test negative design case-control study in each participating country.
- Multicentre test-negative case-control study using data from several countries.

3.2 Study population

The study population comprises community-dwelling individuals with no contra-indication for influenza vaccination who consult a participating physician with symptoms of an influenza like illness (ILI).

Target population for vaccination

The population for which the seasonal (or pandemic if applicable) influenza vaccine is recommended. This may include certain age groups, persons with chronic medical conditions or other risk conditions (such as pregnancy and obesity), professional groups, caregivers.

☑ Each study site defines the target group for vaccination in their study.

3.3 Study period

The study period starts when the influenza virus is circulating and the vaccine is available in each of the participating countries.

Seasonal vaccine: the study period starts at the beginning of the seasonal influenza period and >14 days after the start of the influenza vaccination campaign and the study period finishes at the end of the influenza period.

- Inclusion period: Cases and controls are included from the week of onset of the first influenza positive case presenting in the country-specific study.

  ☑ Each study defines the beginning, the peak and the end of the study period according to the information provided by the country influenza sentinel surveillance system (details available in the study annexes).

  ☑ Each study site specifies the date of the start of their vaccination campaign.

Pandemic vaccine: the study period is defined depending on the gradual availability of vaccines and the pandemic incidence.

☑ Each study site defines the beginning and end of the pandemic VE study period.

3.4 Outcome

The outcomes of interest are:

- subtype-specific laboratory-confirmed influenza A,
- laboratory-confirmed influenza B overall and, if available, by lineage (B Victoria/B Yamagata),
- laboratory-confirmed influenza by clade/genetic variant (where possible).

3.5 Cases

ILI definition

A case of influenza-like illness (ILI) is defined as an individual who consults a participating physician, presenting with sudden onset of symptoms AND at least **one** of the following four systemic symptoms:

- Fever or feverishness
- Malaise
- Headache
- Myalgia

AND at least **one** of the following three respiratory symptoms:

- Cough
- Sore throat
- Shortness of breath

For the pandemic vaccine, the ILI case definition may be revised during the course of the pandemic.

Influenza case

An influenza case is defined as an ILI case with a respiratory sample positive for influenza with at least (sub)type information.

☑ Indicators to define cases are specified in the study annexes.

Laboratory confirmation

Specimens are collected from ILI cases who consult their GP within 7 days of symptom onset.

☑ Mode of specimen collection, storage and transport for each study are listed in the study annexes.

Influenza laboratory confirmation is provided by using RT-PCR.

☑ RT-PCR characteristics for each study are listed in the study annexes.

Where possible, isolates undergo a genetic analysis for currently circulating influenza A viruses (subtypes H3 and H1) and influenza B.

Following the procedures outlined by each study, a systematic sample of viruses (or all viruses) undergo gene sequencing of at least the influenza virus hemagglutinin segment. The sampling procedure can include sequencing all viruses, where technically possible, or a systematic sample thereof. The systematic sample should be representative of cases (ideally a random selection) and be large enough to provide reasonable precision when calculating proportions of virus change over time. Please see annex 2 for data collection of genetic information and the separate laboratory protocol for more information on the analysis of genetic information.

☑ The selection of viruses for each study is specified in the study annexes.

☑ Each study site is to specify laboratory procedures for genetic and antigenic tests.

3.6 Case finding

Case identification

Cases are identified among patients presenting to a participating GP with ILI.

Following the procedures outlined by each study, all ILI cases (sampling all ILI cases is preferred; but if this is not possible, then a systematic sample can be taken, e.g. the first two ILI cases seen each week per GP) are selected and asked to provide a nasal/throat swab specimen for influenza testing. We recommend to select all individuals aged >59 or >64 years. Influenza-positive ILI cases are considered as influenza cases.

☑ Description of the GPs participating in each of the studies (number, distribution, catchment population) is available in the study annexes.

☑ Description of procedures to select ILI cases to swab is available in the study annexes

Case inclusion criteria

Cases are eligible if they meet the above case definition and consent to participate.

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

Case exclusion criteria

Cases are excluded if they:

- refuse to participate in the study;

- are not eligible for influenza vaccination due to a condition listed in the summary of product characteristics;
- are institutionalised;
- are unable to give informed consent or follow an interview in their native language because of aphasia, reduced consciousness, or other reasons;
- are swabbed >7 days after symptom onset;
- have received antivirals ≤7 days prior to swabbing;
- were vaccinated <15 days before symptom onset;
- had tested positive before to any influenza virus in the current season.

Reasons for exclusion are documented.

3.7 Controls

Controls are ILI cases that tested negative for influenza.

Control exclusion criteria

Controls are excluded if they:

- refuse to participate in the study;
- are not eligible for influenza vaccination due to a condition listed in the summary of product characteristics;
- are institutionalised;
- are unable to give informed consent or follow an interview in their native language because of aphasia, reduced consciousness, or other reasons;
- are swabbed >7 days after symptom onset;
- have received antivirals >7 days prior to swabbing;
- were vaccinated <15 days before symptom onset;
- tested positive to any influenza virus in the current season.

Reasons for exclusion are documented.

3.8 Exposure (vaccination)

Definition of vaccination status

Current seasonal influenza vaccine:

- an individual is considered as vaccinated against influenza if the vaccination occurred more than 14 days before disease onset.
- an individual is considered as unvaccinated if they did not receive influenza vaccine in the current season or were vaccinated after inclusion in the study.

Brand-specific seasonal influenza vaccine:

An individual is considered as vaccinated against influenza with a brand-specific vaccine if he/she has received a vaccination with an influenza vaccine of a named brand more than 14 days before

disease onset.

An individual is considered as unvaccinated if they did not receive any influenza vaccine in the current season.

Pandemic vaccine:

- The definition of vaccinated, partially vaccinated and unvaccinated will be defined when it is known how many doses of vaccine are recommended. Once this is known, the protocols will be updated.

Vaccination status ascertainment

The exposure of interest in this study is a vaccination history with trivalent/quadrivalent influenza vaccine (for seasonal vaccine) and vaccination history with the pandemic vaccine (in case of a pandemic). The vaccination history includes date of administration and brand names. Documenting the flu batch codes (where this is feasible) will allow identifying the vaccine brand, the vaccine content (seasonal, pandemic) and the dose.

An individual is considered as vaccinated against influenza if:

- he or she reports having received an influenza vaccination during the current season;

  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 influenza vaccine/vaccination during the current influenza season.

  or

- influenza vaccination has been recorded this season in his/her vaccination card/vaccination booklet.

- Pandemic vaccine: if more than one dose is recommended, the number of doses is documented.

☑ Each study site to document:

- the seasonal and pandemic vaccines used;
- the precise mode of vaccine ascertainment for each study is specified in the study annexes;

- If no precise date of vaccination collected, the variable allowing a patient to be defined as vaccinated or unvaccinated.

3.9 Confounding factors and effect modifiers

Chronic diseases

If physicians are recruiting cases and controls using electronic medical records, the list of ICD codes or classification of health problems in primary care (ICHPPC-2) codes can be used to document a study participant's chronic diseases (see Table 1):

Table 1: ICD-9 and ICHPPC-2 codes for chronic diseases

Chronic diseases	ICD-9 code	ICHPPC-2 code
Enlarged spleen, anaemia	280–289, 759.0	B82
Cirrhosis	571	D97
Diabetes and endocrine disease	250, 251	T89, T90
Heart disease	093, 112.81, 130.3, 391, 393–398, 402, 404, 410–429, 745, 746, 747.1, 747.49, 759.82, 785.2, 785.3	K71, K74-77, K81-K84, K86-K87, K99
Hematologic cancer	200–208	B72, B74
Immunodeficiency and organ transplant	042, 079, 279, V08, V42	B99
Lung disease	011, 460, 462, 465, 466, 480–511, 512.8, 513–517, 518.3, 518.8, 519.9, 714.81	A70, R83, R79, R95, R96, R99
Nonhematologic cancer	140–198, 199.1	A79, D74-D78, F74, H75, K72, L71, N74, N76, R84, R85, S77, S79, T71, T73, U75-U77, U79, W72-W73, X75-X77, X81, Y77-Y
Nutritional deficiencies	254, 255, 259.2, 260–269	T05, T99
Renal disease	274.1, 408, 580–591, 593.71–593.73, 593.9	U99
Dementia, stroke	290–294, 331, 340, 341, 348, 438	P70, K90

Chronic diseases	ICD-9 code	ICHPPC-2 code
Rheumatologic diseases	446, 710, 714.0–714.4, 714.8, 714.89, 714.9	L88

☑ The exact codes used in each study are specified in the study annexes.

Each patient should be evaluated for the presence (currently) of any of the diseases/codes and is classified as 'high risk' if any of them are present.

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

Seasonal vaccine:

The list of underlying conditions in the questionnaire should include at least:

- diabetes, if treated for insulin-dependent or non-insulin-dependent diabetes;
- cardiovascular disease: myocardial infarction, angioplasty, coronary artery bypass surgery, stroke, transient ischemic attacks, treated hypercholesterolemia, treated hypertension;
- chronic pulmonary disease;
- immunodeficiency.

Pandemic vaccine:

The list of underlying conditions in the questionnaire should include all those defining the risk groups in each of the study countries.

☑ Each study site to specify the list of chronic conditions documented.

Severity

The severity of underlying conditions should be measured by the number of hospital admissions due to underlying conditions in the 12 months prior to inclusion in the study.

Smoking history

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

Previous influenza vaccinations

Vaccination against seasonal influenza in the last season (recording vaccination information for the previous influenza season). If information on influenza vaccination in other (earlier) seasons is available, this can be documented.

☑ Each study site to specify the number of seasons for which vaccination information is collected.

Pneumococcal vaccination

Where possible, information on pneumococcal vaccination will be collected, including type of pneumococcal vaccine (e.g. PPSV23 or PCV13) and date or year of receipt.

☑ Each study site to specify pneumococcal vaccine recommended

Functional status

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

☑ Each study site to specify how they define low functional status.

Number of GP consultations in the previous 12 months

In order to document and control for access to care in the various control groups, the number of GP visits in the past 12 months before inclusion in the study is recorded. The consultation for seasonal influenza vaccination should not be included in the count.

Antiviral administration

Use of antivirals should be documented: type, dosage (if possible) and date of administration (patients receiving antivirals prior to swabbing will be excluded from analysis).

Obesity

☑ Each study site to define how obesity is documented (e.g. Body Mass Index > specified value).

Statins

Statins, a class of drugs used to lower cholesterol, may have a confounding or effect modifying effect on influenza VE2,3. If available to collect, statin history can include date the patient started

[2] Vandermeer ML, Thomas AR, Kamimoto L, Reingold A, Gershman K, Meek J, et al. Association between use of statins and mortality among patients hospitalized with laboratory-confirmed influenza virus infections: a multistate study. J Infect Dis. 2012 Jan 1;205(1):13–9.

[3] Omer SB, Phadke VK, Bednarczyk RA, Chamberlain AT, Brosseau JL, Orenstein WA. Impact of Statins on Influenza Vaccine

on statins where known; else just the year, if the patient was known to have been on statins before the current season or if the precise date is unknown. If both of these are unknown, then a simple yes/no response to whether the patient was on statins at the start of October for that season will be used (e.g. on statins on 01 October 2018 for the 2018–19 season). In addition, statin history could include type of statin (synthetic vs natural) and brand name, and number/frequency of doses.

The sources of information for statin status may include:

- consultation of the patient's hospital record;
- interview with the patient's GP;
- interview with the patient's pharmacist;
- data from the patient's insurance company showing evidence of pharmacy delivery or re-imbursement for statins during the current influenza season;
- interview of the patient and/or his/her relatives.

☑ Each study site to describe how statin status ascertainment will be done.

☑ Each study site to define statin use based on data collected.

Source of information

Data is collected using a standardised questionnaire. For cases and controls selected at GP practices, data are collected face-to-face.

If GPs use electronic medical records, information on collected variables can be extracted from these records to validate the information collected through the standardised questionnaire.

3.10 Sample size

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

☑ The sample size calculation for each study is detailed in the study annexes.

In influenza 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 an influenza VE of 70%, a lower boundary confidence interval of 1% does not provide us with a very

Effectiveness Against Medically Attended Acute Respiratory Illness. J Infect Dis. 2016 Apr 15;213(8):1216–23.

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, which can be the case in particular for A(H3N2),[4] we would need a huge sample size to provide a VE estimate that does not include 0%. For example, if the true VE is 5–10%, then a study providing a lower boundary not including 0% 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 1:1. 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/ case	Detectable OR	Vaccine coverage in source population/controls	Number of cases	Number of controls	VE	CI
0.3	1	0.2	0.3	85	85	80	51-92
0.3	1	0.3	0.3	118	118	70	40-85
0.3	1	0.4	0.3	157	157	60	30-77
0.3	1	0.5	0.3	203	203	50	20-69
0.3	1	0.6	0.3	255	255	40	10-60
0.3	1	0.7	0.3	314	314	30	0-51
0.2	1	0.2	0.3	148	148	80	60-90
0.2	1	0.3	0.3	216	216	70	50-82
0.2	1	0.4	0.3	299	299	60	40-73
0.2	1	0.5	0.3	395	395	50	30-64
0.2	1	0.6	0.3	507	507	40	20-55
0.2	1	0.7	0.3	633	633	30	10-46
0.1	1	0.2	0.3	433	433	80	70-87
0.1	1	0.3	0.3	681	681	70	60-77
0.1	1	0.4	0.3	985	985	60	50-68
0.1	1	0.5	0.3	1346	1346	50	40-58

[4] Belongia et al., Lancet Infect Dis. 2016 Aug; 16(8):942-51

0.1	1	0.6	0.3	1764	1764	40	30-49
0.1	1	0.7	0.3	2240	2240	30	20-39
0.3	1	0.2	0.4	63	63	80	49-92
0.3	1	0.3	0.4	91	91	70	40-85
0.3	1	0.4	0.4	125	125	60	30-77
0.3	1	0.5	0.4	165	165	50	20-69
0.3	1	0.6	0.4	212	212	40	10-60
0.3	1	0.7	0.4	265	265	30	0-51
0.2	1	0.2	0.4	111	111	80	60-90
0.2	1	0.3	0.4	168	168	70	50-82
0.2	1	0.4	0.4	238	238	60	40-73
0.2	1	0.5	0.4	323	323	50	30-64
0.2	1	0.6	0.4	421	421	40	20-55
0.2	1	0.7	0.4	534	534	30	10-46
0.1	1	0.2	0.4	323	323	80	70-87
0.1	1	0.3	0.4	528	528	70	60-77
0.1	1	0.4	0.4	786	786	60	50-68
0.1	1	0.5	0.4	1098	1098	50	40-58
0.1	1	0.6	0.4	1466	1466	40	30-49
0.1	1	0.7	0.4	1891	1891	30	20-39
0.3	1	0.2	0.5	51	51	80	51-92
0.3	1	0.3	0.5	77	77	70	40-85
0.3	1	0.4	0.5	109	109	60	30-77
0.3	1	0.5	0.5	148	148	50	20-69
0.3	1	0.6	0.5	193	193	40	10-60
0.3	1	0.7	0.5	246	246	30	0-51
0.2	1	0.2	0.5	90	90	80	60-90
0.2	1	0.3	0.5	142	142	70	50-82
0.2	1	0.4	0.5	208	208	60	40-73
0.2	1	0.5	0.5	289	289	50	30-64
0.2	1	0.6	0.5	384	384	40	20-55
0.2	1	0.7	0.5	495	495	30	10-46
0.1	1	0.2	0.5	262	262	80	70-87
0.1	1	0.3	0.5	447	447	70	60-78
0.1	1	0.4	0.5	687	687	60	50-68
0.1	1	0.5	0.5	983	983	50	40-58
0.1	1	0.6	0.5	1337	1337	40	30-49
0.1	1	0.7	0.5	1751	1751	30	20-39

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

The sample size should also be respected for each population subgroup for which a sub (stratified)

analysis (e.g. VE among different age groups, VE by clade, etc.) is planned.

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

3.11 Data

Data on cases and controls are collected at GP office level. Physicians interview the patients using a standardised questionnaire. GPs using electronic medical records can extract some or all of the variables from these records (e.g. vaccination status, chronic diseases based on ICD codes).

Epiconcept has developed a secure electronic questionnaire and a web-based questionnaire for participating GPs. Epiconcept has an accreditation to host personal health data.

Laboratory information will be reported to the study site coordinator using the reporting procedures existing in each study site for influenza surveillance.

Double data entry is recommended unless medical electronic records are used.

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

☑ Details on data collection methods, data entry and data transmission are available in the study annexes.

Collected information

Collected information should include (see also Annex 1: List of variables, definition and coding):

- study identification: country and GP;
- case/control demographics;
- ILI signs, symptoms;
- date of onset of ILI;
- date of swabbing;
- laboratory results (including information antigenic and genetic analysis, where available);
- selected underlying chronic conditions (including diabetes, heart disease, chronic obstructing pulmonary disorder and immunodeficiencies);
- number of hospitalisations for the chronic diseases in the previous 12 months;
- number of GP visits in the previous 12 months;
- smoking history;
- current season influenza vaccination including date and product;
- pandemic vaccination including number of doses, date, product (if applicable);
- influenza vaccination in the previous season (or more seasons if available);
- pneumococcal vaccination status, type of vaccine and either date or year of vaccination
- obesity status;
- functional status;
- antiviral administration.
- statins (optional)

☑ Each study site to list the variables collected.

Pandemic vaccine data collected will be revised as more information on the vaccine and the target groups become available.

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 agreement between patient vaccine records/vaccination status reported by study participant/vaccine registries is measured.

☑ The specific validation procedures, including sample size calculation for questionnaire validation (if applicable) are specified in the study annexes.

3.12 Data management

Individual analysis

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

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

Data checking and cleaning

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 changes to the data are documented and stored separately from the crude database. Any recoding of data (e.g. age) is documented. A guide and/or an example Stata do-file for data cleaning is provided if so desired. A list of possible checks is outlined here:

- Checks for missing data in all variables
- Check for persons not participating
- Check if persons participating despite contraindications
- Age<0
- Age>99
- GP visits<0

- GP visits>20
- Pregnant males
- Pregnant <15 years (not impossible but worth checking)
- Pregnant over 45 (not impossible but worth checking)
- Swab date before onset date
- >7 days between onset of symptoms and swabbing
- Very early onset date (e.g. onset date before start of vaccination campaign)
- Very late onset date (e.g. onset date in the future)
- Very early swab date (e.g. swab date before start of vaccination campaign)
- Very late swab date (e.g. onset date in the future)
- Very early influenza vaccination date (e.g. before August prior to season begin)
- Very late influenza vaccination date (e.g. influenza vaccination date in the future)
- Missing vaccination date
- Not vaccinated, but vaccination brand reported
- Vaccinated, but vaccination brand not reported (missing data)
- Vaccination after or on day of onset of symptoms

- Not coded as vaccinated, but with vaccination date
- Vaccinated, but dates not available (missing data)
- Hospitalisations, but chronic disease coded as “no”
- Hospitalisations, but missing chronic disease
- Improbably high number of hospitalisations (e.g. >7)
- Duplicates in ID number
- Antiviral date but no antivirals taken
- Antivirals given prior to swabbing (these patients should have been excluded)
- Not coded as target group, but 65+ years
- Not coded as target group, but report having chronic disease
- Not coded as target group, but pregnant (depending on target groups by country)

Data management pooled analysis

EpiConcept conducts the pooled analysis. Individual data from each study is sent securely to EpiConcept’s study database. This can be done via the secure data entry system provided by EpiConcept or via EpiFiles (https://epifiles.voozanoo.net), which is a web platform which allows secure file exchanges between entities, or another secure method of data transfer. All personal identifier information such as names, addresses, and medical registration codes should be deleted before data transmission to EpiConcept, where all individual data are pooled. A country (or study) identifier is included in each record (e.g. ES for Spain, RO for Romania), a GP code is included

(e.g. a unique number), and each record is given a unique number. This number is also included in the study team's database and will be used by EpiConcept and the study teams during pooling, so that records can be traced back whilst maintaining anonymity, if there are any queries during the data checking process. Study databases can be sent to EpiConcept in any format. Data can be coded as outlined in Annex 1, or a codebook can be provided by the study teams to EpiConcept that includes the variable names, descriptions and coding. EpiConcept performs all necessary data checking and cleaning. EpiConcept documents and shares any further data cleaning and analysis procedures with all study coordinators to ensure they can be reproduced.

See annex 4 for detailed guidelines to the pooled analysis data management.

3.13 Analysis

Analyses are carried out first for each individual study. In the second step, a pooled analysis is conducted (see annex 4).

If sample size permits, analyses are conducted:

- on all data and separately with cases/controls restricted to an interval between date of onset of symptoms and swab taken of <4 days;
- for VE against (sub)type-specific influenza, influenza B by lineage and VE by clade and genetic variants;
- by time within the season (e.g. early, peak, late influenza season or by week or group of weeks in the season [VE for weeks 2-3, 4-5, 6-7, etc.] or by month);
- for the various types of vaccines (adjuvanted/non-adjuvanted; high-dose vs. normal dose; trivalent or quadrivalent, egg-based vs. cell-based), groups of vaccines (split virion, subunit, etc.), mode of injection (intradermal vs. intramuscular) and by vaccine brand.
- Target population

All analyses are done separately for seasonal and pandemic vaccine (if applicable).

Individual study site analysis

Descriptive and univariable analysis

The proportion of eligible ILI cases and controls who consented to participate in the study is calculated (response rate).

Study participants are described by baseline characteristics. Baseline characteristics of cases and controls in unmatched studies are compared using the chi-square test, Fisher's exact test, t-test or the Mann-Whitney test (depending on the nature of the variable and the sample size). The association between vaccination status and baseline characteristics is measured for both case and control groups, an example layout of this is in table 3 below.

Table 3: Example of descriptive table for cases and controls

Variables	Number of influenza A(H3N2) cases /total n (%)	Number of test-negative controls /total n (%)
Median age	x	x
Missing	x	x
Age groups		
0-4	x/x (x)	x/x (x)
5-14	x/x (x)	x/x (x)
15-64	x/x (x)	x/x (x)
≥ 65	x/x (x)	x/x (x)
Missing	x	x
Sex		
Female	x/x (x)	x/x (x)
	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.		

Measure of effect

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

Stratified analysis

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

- age groups 0-14 years, 15-64 years, 65 years and older;
- presence of at least one chronic condition;
- time: early influenza season, peak, late influenza season.

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.

Multivariable analysis

A multivariable logistic regression analysis is conducted to control for negative and positive confounding. Odds ratios and standard errors are obtained. Variables are tested for multicollinearity. Interactions are tested using the likelihood ratio test or Wald's test and included in the model if deemed to be biologically plausible and at a reasonable statistical significance (e.g. 5%).

If possible, a variable for age and for onset time should always be included in the model. See also "Minimum sample size" later in this section.

Continuous variables

Continuous variables in the I-MOVE datasets include age, date of onset of symptoms and GP visits. 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. Models will be compared with the continuous variable coded as categories, a linear term, polynomial or a spline – the Akaike information criterion (AIC) can be used for comparison if the data in the models are the same (the lower the AIC, the better). Coefficients and standard errors of the parameters (e.g. the individual spline parameters) will be looked at and models where the standard errors exceed the coefficients will not be used. In addition, a balance will be sought between simplicity of the model (so a non-expert can understand what is going on), precision, and a model that estimates the VE with the least bias.

Output tables presenting VE estimates

In order to present the results in the most transparent manner and to enable the reader to best understand the data, tables similar to table 4 below can be used. Useful information includes numbers of cases and controls (overall and vaccinated) and presentation of results for different models.

Influenza (sub)type	**Analysis scenarios, population included**		**VE (%)**	**(95%CI)**
	All ages			

A(H1N1)pdm09		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, age (cubic spline),chronic condition		
	15-64 years			
		N (cases/ vaccinated; controls/ vaccinated)		
		Crude*		
		Adjusted for onset week, age (cubic spline), chronic condition, sex *		

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

Further analyses

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

- VE at different time points along the season (e.g. VE by week or group of weeks in the season [VE for weeks 2-3, 4-5, 6-7, etc.] or by month)
- VE by time since vaccination. Time since vaccination can be calculated by subtracting the date of vaccination from the date of onset. Time since vaccination can then be modelled as a categorical or continuous variable.
- VE of previous season influenza vaccination only, current season influenza only and combined season vaccination
- Using the systematic samples of the sequenced isolates, the proportion of virus changes will be calculated at different time points along the season (e.g. by month or week or group of weeks in the season: weeks 2-3, 4-5, 5-6, etc.). This will be compared to VE at different time points along the season.

Minimum sample size

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

- There are at least 10 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 additional parameter in the model)
- There are at least five records in each cell of the two-by-two table of case and vaccination status and between case status and any other covariate category in the model

With low sample size, sensitivity analyses can be carried out comparing VE with "standard" logistic regression and VE using Firth's method of penalised logistic regression. Where there is a discrepancy of 10% or more, both results should always be reported.

☑ Each study site to specify criteria for which to determine minimum sample size if desired.

Missing data

Any missing data will be documented.

If a lot of data are missing and/or there is evidence of bias in the missing data, and variables that are considered good predictors of the missing data are available, then multiple imputation methods at study level will be used to replace missing values.

A sensitivity analysis will be carried out comparing results from complete case analysis (where records with missing data are dropped) and full set analysis (with imputed data).

Pooled analysis

For the pooled analysis, please see Annex 4.

3.14 Potential biases

Negative confounding

Negative confounding refers to biases that reduce the VE estimate if not accounted for. For example, high risk groups are more likely to be vaccinated, and may be more likely to get flu (or more severe symptoms).

Positive confounding

Positive confounding refers to biases that increase the VE estimate if not accounted for. For example,

a 'healthy vaccine effect'. People with a healthy lifestyle may be more likely to accept/request vaccination and less likely to get flu (although they might actually be more likely to consult with ILI).

Positive and negative confounding are minimised through stratification and multivariable analysis and variables collected in order to measure positive and negative confounding. It is also reduced by the use of the study design properties of the TND as this takes into account factors associated with the propensity to consult. It is not possible to rule out the presence of characteristics in the study population for which no information is collected in the study questionnaire and that therefore could lead to positive or negative confounding. In this way, residual positive or negative confounding may be present.

Representativity of cases

The study includes only cases consulting a GP for ILI. Health-seeking behaviour may differ by country depending on the case management strategy (e.g. recommendation of not going to the GP). In some countries, only severe cases will go to the GP. In others, severe cases will directly go to an emergency room without consulting their GP. The types of cases included in the study should be described for each of the studies, and how representativeness may affect the VE estimates.

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 VE.

3.15 Dissemination of results

The enrolment of cases/controls is regularly updated by each study coordinator on the 'I-MOVE' web page.

Seasonal vaccine: initial VE estimates (intra-seasonal) are disseminated early during the influenza season; final estimates follow at the end of the season.

Pandemic vaccine: initial estimates will be disseminated once the sample size allows for meaningful interpretation.

3.16 Publications, scientific communication

Each study coordinator decides where the results of the individual studies are published and which scientific conferences are attended in order to present the results. An article presenting the results of the pooled analysis and estimates for the EU/EEA will be submitted to a peer-reviewed journal. Regarding the list of

authors, we will follow the recommendations of the International Committee of Medical Journal Editors http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html and each season, the list of authors will be defined in the Principles of Collaboration. Co-workers and contributors are acknowledged.

I-MOVE results will contribute to the report prepared by the GIVE (Global Influenza Vaccine Effectiveness) collaboration for the annual Northern and Southern Hemisphere WHO Meeting on the Composition of Influenza Virus Vaccines.

3.17 Training

Participating GPs are trained on the study protocol before the start of the study. They receive the protocol, questionnaires and laboratory swabbing procedures.

4. Logistical aspects

4.1 Study leader

In each country, a principal investigator coordinates the study at the country level and acts as focal point for the European study. Epiconcept 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 country, a part-time investigator is in charge of monitoring data collection at the GP office level. GPs collect information among cases and controls. GPs could be offered a payment or compensation for their participation in the study.

☑ The specific human resources needed in each country are detailed in the study annexes. EpiConcept *ensures the overall coordination of the various studies.*

4.3 Supervision

Site visits and joint workshops may be organised by Epiconcept/Member States consortium in order to carry out an appraisal of the ongoing studies in the various countries involved. The appraisal team should be composed of two persons from the various project partners.

4.4 Questionnaires

Standardised questionnaires should be developed for the study. The variables used at the European level are collected in the same way for each of the studies (see Annex 1: List of variables, definition and coding).

4.5 Computer support

Data collection and entry are conducted at the country level. For countries willing to submit data electronically, Epiconcept provides an online questionnaire.

4.6 Consent

Each study should comply with national ethics committee requirements. Informed consent will be required from all participants. The national ethics committees will specify whether oral or written consent is required.

- ☑ Details are available in the study annexes.

4.7 Data confidentiality and data protection

Each study and the multicentre study is compliant with the requirements of data protection, specifically the European Union's General Data Protection Regulation (GDPR) and the any country-specific data protection acts.

- ☑ Study sites provide details in the study annexes.

4.8 Report

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

Bibliography

Test-negative design

- Orenstein EW, De Serres G, Haber MJ, Shay DK, Bridges CB, Gargiullo P, et al. Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. Int J Epidemiol 2007 Jun;36(3):623-31.
- Orenstein WA, Bernier RH, Dondero TJ, Hinman AR, Marks JS, Bart KJ, et al. Field evaluation of vaccine efficacy. Bull World Health Organ 1985;63(6):1055-68.
- Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990 Mar;19(1):205-13.
- Rothman, K., Greenland, S., Lash, T. Modern Epidemiology (ed.) (Philadephia: Lippincott-Williams-Wilkins, 2008).
- Valenciano M, Ciancio B, Moren A; Influenza Vaccine Effectiveness Working Group. First steps in the design of a system to monitor vaccine effectiveness during seasonal and pandemic influenza in EU/EEA Member States. Euro Surveill. 2008 Oct 23;13(43). pii: 19015.

Pooling

- Blettner, M., Sauerbrei, W., Schlehofer, B., Scheuchenpflug, T. & Friedenreich, C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. Int J Epidemiol 28, 1-9(1999).
- DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. Control Clin Trials 7, 177-88(1986).
- Darvishian M, Bijlsma MJ, Hak E, van den Heuvel ER. Effectiveness of seasonal influenza vaccine in community-dwelling elderly people: a meta-analysis of test-negative design case-control studies. Lancet Infect Dis. 2014 Dec;14(12):1228-39.
- Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index?. Psychol Methods 11, 193-206(2006).
- National Research Council Combining information - Statistical Issues and Opportunites for Research. ,(1992).
- Stukel TA, Demidenko E, Dykes J, Karagas MR. Two-stage methods for the analysis of pooled data. Stat Med. 2001 Jul 30;20(14):2115-30.
- Smith-Warner, S.A., Spiegelman, D., Ritz, J., Albanes, D., Beeson, W.L., Bernstein, L. et al. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. Am. J. Epidemiol. 163, 1053-64(2006).

Confounding and bias

- Hak E, Verheij TJM, Grobbee DE, Nichol KL, Hoes AW. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. J Epidemiol Community Health 2002;56:951–55.
- Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol doi:10.1093/ije/dyi274.

- Jackson LA, Nelson JC, Benson P, Neuzil KM, Reid RJ, Psaty BM et al. Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors. Int J Epidemiol doi:10.1093/ije/dyi275.

Statistical methods

- Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol. 1995 Dec;48(12):1495–501.

- Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995 Dec;48(12):1503–10.

Previous vaccination

- Belongia EA, Skowronski DM, McLean HQ, Chambers C, Sundaram ME, De Serres G. Repeated annual influenza vaccination and vaccine effectiveness: review of evidence. Expert Rev Vaccines. 2017;16(7):1–14.

- Skowronski DM, Chambers C, De Serres G, Sabaiduc S, Winter A-L, Dickinson JA, et al. Serial Vaccination and the Antigenic Distance Hypothesis: Effects on Influenza Vaccine Effectiveness During A(H3N2) Epidemics in Canada, 2010-2011 to 2014-2015. J Infect Dis. 2017 Apr 1;215(7):1059–99.

- Valenciano M, Kissling E, Larrauri A, Nunes B, Pitigoi D, O'Donnell J, et al. Exploring the effect of previous inactivated influenza vaccination on seasonal influenza vaccine effectiveness against medically attended influenza: Results of the European I-MOVE multicentre test-negative case-control study, 2011/2012-2016/2017. Influenza Other Respir Viruses. 2018 Apr 16;

Within-season waning of VE

- Ferdinands JM, Fry AM, Reynolds S, et al. Intraseason waning of influenza vaccine protection: Evidence from the US Influenza Vaccine Effectiveness Network, 2011-12 through 2014-15. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 2016;

- Kissling E, Nunes B, Robertson C, Valenciano M, Reuss A, Larrauri A, et al. I-MOVE multicentre case-control study 2010/11 to 2014/15: Is there within-season waning of influenza type/subtype vaccine effectiveness with increasing time since vaccination? Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2016 Apr 21;21(16).

Selected I-MOVE publications

- Valenciano M, Kissling E, Cohen J-M, Oroszi B, Barret A-S, Rizzo C, et al. Estimates of pandemic influenza vaccine effectiveness in Europe, 2009-2010: results of Influenza Monitoring Vaccine Effectiveness in Europe (I-MOVE) multicentre case-control study. PLoS Med. 2011 Jan;8(1):e1000388.

- 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).

- Kissling E, Valenciano M, Buchholz U, Larrauri A, Cohen JM, Nunes B, et al. Influenza vaccine effectiveness estimates in Europe in a season with three influenza type/subtypes circulating: the I-MOVE multicentre case-control study, influenza season 2012/13. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2014 Feb 13;19(6).

- Kissling E, Valenciano M, Pozo F, Vilcu A-M, Reuss A, Rizzo C, et al. 2015/16 I-MOVE/I-MOVE+ multicentre case-control study in Europe: Moderate vaccine effectiveness estimates against influenza A(H1N1)pdm09 and low estimates against lineage-mismatched influenza B among children. Influenza Other Respir Viruses. 2018 Jul;12(4):423–37.

- Rondy M, Kissling E, Emborg H-D, Gherasim A, Pebody R, Trebbien R, et al. Interim 2017/18 influenza seasonal vaccine effectiveness: combined results from five European studies. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2018;23(9).

- Kissling E, Pozo F., Buda S., Vilcu A., Rizzo C. et al., 'Effectiveness of influenza vaccine against influenza A in Europe in seasons of different A(H1N1)pdm09 and the same A(H3N2) vaccine components (2016–17 and 2017–18)', *Vaccine X*, p. 100042, Sep. 2019.

- Kissling E, Rose A, Emborg H., Gherasim A., Pebody R., et al., 'Interim 2018/19 influenza vaccine effectiveness: six European studies, October 2018 to January 2019', *Eurosurveillance*, vol. 24, no. 8, Feb. 2019.

Annexes

Annex 1: List of variables, definitions and coding

Variable name	Type	Values and coding	Definition
idcountry	Numeric	Coded according to international country codes	Identifier uniquely identifying the country
participate	Numeric (binary)	0 = No 1 = Yes	Agrees to participate
refuse	Text		Reasons for refusal to participate
id	Numeric (continuous)	Unique integer	Unique number for each record
case	Numeric (binary)	0 = control 1 = case	Identifies cases and controls
gpcode	Numeric (continuous)	Unique integer	Unique number for each GP (preventing identification of GP)
dob	Date	dd/mm/yyyy	Date of birth of study participant
age	Numeric (continuous)	Integer	Age of each participant in years
sex	Numeric (binary)	0 = female 1 = male	Sex of study participant
onsetdate	Date	dd/mm/yyyy	Date of onset of symptoms
swabdate	Date	dd/mm/yyyy	Swabbing date
fever	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Fever
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
shortness of breath	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Weakness

Variable name	Type	Values and coding	Definition
lab_res	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result (positive/negative)
lab_virusa	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result: virus type A
lab_virusb	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result: virus type B
lab_h1n1	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result: virus subtype AH1N1
lab_h3n2	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result: virus subtype AH3N2
byamagata	Numeric (categorical)	0 = Yamagata 1 = Victoria 8 = Do not know	Laboratory result: B Yamagata lineage
bvictoria	Numeric (categorical)	0 = Yamagata 1 = Victoria 8 = Do not know	Laboratory result: B Victoria lineage
genetic_group	Text		Laboratory result: genetic group
subclades	Text		Information on further mutations
antigenic_analysis	Text		Laboratory result: antigenic group
seasvaccany	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received flu vaccination in current season
seasvaccdate	Date	dd/mm/yyyy	Vaccination date
seasvacctype	Text		Type of vaccine (brand name)
pneumovacc	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received pneumococcal vaccination
pneumovacctype	Numeric (categorical)	1 = PPSV23 2 = PCV13 3 = Other (pls specify) 8 = Do not know	Type of pneumococcal vaccine
pneumovacctype_other	Text		Other type of pneumococcal vaccine if not PPSV23 or PCV13
pneumoyear	Number		Year of receipt of pneumococcal vaccination
vacc_n1	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Influenza vaccination in the previous season (n-1)
diabetes	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Diabetes and endocrine
heart_dis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Heart disease

Variable name	Type	Values and coding	Definition
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
obese	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Clinically obese
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
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
antivir	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Administration of antivirals
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	Exclusion criteria: living in a residential home
contra	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Exclusion criteria: contraindication for influenza vaccination
prev_flu	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Lab-confirmed previous influenza in the season
statin	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Patient was under statin treatment before symptom onset
stat_brand	Text		Name of statin product used
stat_type	Numeric (categorical)	1 = Synthetic 2 = Natural 8 = Do not know	Synthetic vs natural statin
stat_dose_mg	Numeric (in mg)		Statin dose in atorvastatin equivalents (in mg)
stat_dose_fr1	Numeric	0=per day 1= per week 8 = do not know	Frequency of statin dose (per day vs per week)
stat_dose_fr2	Numeric	Integer	# times dose given per day or week

Variable name	Type	Values and coding	Definition
stat_onsetd	Date	dd/mm/yyyy	Date patient started statin treatment if this season
stat_onsety	Numeric	yyyy	Year; if patient started statins before this season or precise date (stat_onsetd) is NK

This table represents a selection of confounders. Variables can be included or excluded as necessary.

In a pandemic, these further variables may be required:

Variable name	Type	Values and coding	Definition
panvaccany	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received pandemic flu vaccination in current year
panvaccdate1	Date	dd/mm/yyyy	Vaccination date first dose
panvaccdate2	Date	dd/mm/yyyy	Vaccination date second dose
panvacctype	Text		Type of vaccine (brand name)
panvaccdose	Numeric	0, 1, 2	Number of doses received

The variables above can be recoded to obtain variables useful for analysis. Possible ways to recode variables are outlined in the pooled data analysis section in annex 4 and 6.

Annex 2: Genetic and antigenic analysis data (examples)

Information on genetic and antigenic analyses can be collected with a form similar to the table 2.1 below. A simplified version with just country, ID number and GISAID number can be used as well.

A(H1N1)	Country	Region/City	ID number I-MOVE+ case-control study	Date sample	Strain	Amino acid mutation with position and changes	Amino acid mutation with position and changes	Amino acid mutation with position and changes	Amino acid mutation with position and changes	Amino acid mutation with position and changes	Type of sample (primary specimen or isolate)	GISAID number	Antigenic analysis (IHA)	Genetic analysis (HA1)	Genetic group	Selected for characterisation?	Reasons for not characterising?	CT value
Row for 2018/19 vaccine reference strain																		
Row for strain with AA substitutions compared with vaccine reference strain																		
Row for strain with AA substitutions compared with vaccine reference strain																		
A(H3N2)																		

Row for strain with AA substitutions compared with vaccine reference strain																	
B/Victoria																	
Row for strain with AA substitutions compared with vaccine reference strain																	
B/Yamagata																	
Row for strain with AA substitutions compared with vaccine reference strain																	

Table 2.1: Example of data collection form for genetic and antigenic data.

	Country	ID number I-MOVE+ case-control study	GISAID number	Selected for characterisation?	Reasons for not characterising?	CT value	Type of sample (primary specimen or isolate)

Row for 2018/19 vaccine reference strain							
Row for strain with AA substitutions compared with vaccine reference strain							
Row for strain with AA substitutions compared with vaccine reference strain							

Table 2.2: Example of simplified data collection form for genetic and antigenic data.

In order to better understand how viruses were selected for sequencing over time, an additional sampling fraction document, as outlined in table 2.3 can be used.

Time period A(H3N2)	**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					
Example1	*01/10/2016*	*31/12/2016*	*1*	*Date of onset*	*(this is only an example; all specimens were characterised)*
Example2	*01/01/2017*	*15/02/2017*	*0.2*	*Date of onset*	*(this is only an example; 20% of all specimens were characterised)*
A(H1N1)					

1
2
B/Victoria
1
2
B/Yamagata
1
2

Table 2.3: Example of documenting outlining how viruses were selected for sequencing over time

Annex 3: Pooled data management

Data preparation and transfer at study-level

Data validation, cleaning and verification will be carried out at study-level (for flow chart, see Annex 5). All personal identifier information such as names, addresses, medical registration codes will be deleted. A country (or study) identifier (e.g. ES for Spain, RO for Romania) and a unique code to identify a practitioner will be included in each record and each record will be given a unique number. This unique identifier is also included in the study team database and will be used by EpiConcept and the study teams during pooling, so that records can be traced back to study site level whilst maintaining anonymity, if there are any further queries.

Minimum dataset

The minimum dataset will be transmitted to EpiConcept where individual data will be pooled, and includes:

- study identification: country and GP
- case / control demographics
- signs, symptoms
- date of onset of ILI
- date of swabbing
- laboratory results
- selected underlying chronic conditions relating to country specific influenza vaccine recommendations
- number of hospitalisations for the chronic diseases in the previous 12 months;
- obesity status;
- current season influenza vaccination status including date received and brand; and
- previous season influenza vaccination status.

Further recommended variables

- Number of GP visits in the previous 12 months
- Functional status (ability to bathe alone and walk unaided)
- Smoking history
- Pneumococcal vaccine status, pneumococcal vaccine type and year of vaccination
- Antiviral administration (including dates of administration)
- Statin use
- Previous influenza infection during the season

Data transfer

Study databases can be sent to EpiConcept in any format (e.g. Stata, CSV, EpiData, etc.). The minimum dataset can be coded as in Annex 1, or a codebook can be provided to EpiConcept with the

variable names and descriptions and the coding of variable values. Data must be transferred securely either via the secure data entry system provided by EpiConcept or via EpiFiles (https://epifiles.voozanoo.net), which is a web platform which allows secure file exchanges between entities, or another secure method of data transfer.

Data checking and cleaning

EpiConcept will conduct data checking and cleaning again, once the data are received. Summary and frequency tables, logical error checks and graphic displays of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of swabbing before date of onset of symptoms). Any improbable, illegal or missing values will be reported to the country in question. See list of example data checks in section 3.13 under “Data checking and cleaning”.

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

Data exclusion and restriction

- Patients that fulfil the exclusion criteria (contra-indications for vaccination, institutionalised patients, unable to give informed consent) will be excluded from the analysis.
- The data will be further restricted, with patients included only if the following criteria apply:
    - no antivirals administered prior to swabbing
    - day of onset of symptoms more than 14 days after begin of national vaccination campaign
    - symptoms correspond to the EU ILI case definition
    - patients swabbed within 7 days of symptom onset
    - week of onset of symptoms not before the week of onset of the first influenza positive case in the study
    - week of onset of symptoms not after the week of onset of the last influenza positive case in the study, or the week of onset of an influenza positive case after the peak after which there were 2 consecutive weeks of no cases
    - unvaccinated or vaccinated >14 days before symptom onset
    - laboratory results are available

A study-site specific flowchart of exclusions and restrictions will be shared with each of the study sites.

Data recoding

Variables will be recoded and new variables generated according to Annex 6. The recoded data will be stored separately from the crude data and recoding will be documented.

Missing data

In the first instance (at country level), great care will be taken to avoid missing data.

Missing data will be described in terms of number and frequency of missing values for each variable of interest and in terms of number and frequency of records with 0, 1, 2, etc. missing values.

Baseline characteristics of records with missing data will be compared to records without missing data.

Each variable with missing data will be qualitatively assessed to determine the mechanism of missing data.

Reasons for missing data for each variable will be discussed with all study partners. The mechanism for missing data will be determined: missing completely at random, missing at random and not missing at random. If data are determined to be missing at random, then an imputation could be carried out.

If there is a reasonable amount of missing data, associations of the missing data with the other covariates in the study will be described. If there is at least one variable with missing data associated with the outcome and with an exposure/risk factor then a complete case analysis might be biased.

If this is the case and if variables that are considered good predictors of the missing data are available, then multiple imputation methods at study level will be used to replace missing values.

If there are very few missing data, an imputation will not be carried out.

A complete case analysis will be carried out and presented even if multiple imputation methods are applied on the dataset.

Multiple imputation

Multiple imputation is a procedure where missing data are estimated from the observed and the uncertainty of the missing data is taken into account in the final estimate.

Here, multiple imputation using chained equations will be carried out, using the Stata “mi impute chained” procedure. The imputation is done by creating a number of possible databases and using a pooled analysis (taking into account within- and between-database variance) to obtain final estimates. This takes the uncertainty of the imputation of missing data into account. Variables to include in the imputation model will include:

- Both variables with and without missing data
- The outcome variable
- If there is a measure of time in the dataset, it will be included
- All variables that are part of the multivariable model

- Other variables that are predictive of missing data and of a value to be missing
- Variables predictive of missing values

Different models for imputation, e.g. including different sets of predictors, different numbers of iterations and databases created will be compared and the robustness of the imputation assessed.

After imputation, the imputed data will be assessed in the following way

- Checking for any incomplete data in the imputed variables
- Imputing percentages for each variable compared to observed percentages (to check for very large differences, implausible values, etc.)

After carrying out a multiple imputation, the resulting dataset will be analysed with specific multiple imputation commands.

The results from the complete case analysis will be compared to the analysis using the imputed dataset.

Data appending and data flow

After data cleaning the data will be appended, and a unique identifier for each GP per country will be created by concatenating the study code and the GP code. An example data flow chart is presented in Annex 5.

Annex 4: Pooled data analysis

Descriptive analysis

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

- Number of GPs participating and catchment population
- Beginning of the study
- Beginning of influenza period, peak, end
- Beginning of vaccination campaigns for seasonal and pandemic vaccine (if applicable)
- Proportion of ILI flu positive among all ILI cases
- Proportion of persons belonging to target group for vaccination
- Sample size, including cases and controls by vaccination status

Individual level analysis

Analyses will be carried out first for each individual study, shared with the study site team for validation, and then, in a second step, a pooled analysis will be conducted.

Analysis will be done if sample size permits

- on all data and separately with cases/control restricted to an interval between date of onset of symptoms and swab taken of <4 days;
- for VE against (sub)type-specific influenza, influenza B by lineage and VE by strain-specific genetic groups
- for the various types of vaccines (adjuvanted/non-adjuvanted; high-dose vs. normal dose, trivalent vs. quadrivalent, live-attenuated vs. inactivated, egg-based vs. cell-based), groups of vaccines (split virion, subunit, etc.), mode of injection (intradermal vs. intramuscular) and by vaccine brand.
- By age group 0-14 years, 15-64 years, 65 years and over)

All analyses will be carried out with either Stata v15.1 (Stata Corp LP, College Station, TX, USA) or R (R Core Team (2018)).

For methods on individual level analysis, see main section.

Testing for heterogeneity

Study-specific crude and adjusted ORs and their Cis will be plotted in separate forest plots. Following the core protocol minimises heterogeneity between studies. However, adherence to the protocol and study design and study quality characteristics will be checked again. Other study site characteristics will be assessed where feasible, such as types of circulating virus, information on health care use, organisation of the vaccination campaign. Then a qualitative decision will be taken if one or more

studies are substantially different from the other. Statistical heterogeneity between studies will be tested using Q-test and the I^2 index (see boxes for formulae below). The Q statistic follows a Chi^2 distribution (with k-1 degrees of freedom).The Q-test reports presence or absence of heterogeneity, while the I^2 index (based on the Q-statistic) quantifies the extent of the heterogeneity.

According to the Higgens and Thompson classification, an I^2 index of around 25% indicates low, 50% indicates medium and 75% indicated high heterogeneity between studies.

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

Where:

$$w_i = 1/v_i$$

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

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

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

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

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

If heterogeneity is detected and there are adequate numbers of study sites, a metaregression could be carried out to explore which study characteristics may contribute to the heterogeneity.

1-stage pooled analysis approach

A 1-stage pooled approach may also be used for analysis, particularly if sample sizes are too small to measure vaccine effectiveness controlling for all potential confounders for each individual study site. Individual study data will be pooled into one dataset and analysed as a 1-stage model with study as a fixed effect (see Annex 7 for Stata syntax). This could provide a large enough sample size to obtain (for example) an estimate of VE early in the season with reasonable precision. The results of this analysis should be interpreted with caution, though, as it assumes that the underlying true exposure effect is the same in all studies. If the association of covariates with the outcome differs between studies, then interactions between study sites and covariates need to be introduced.

Formal tests of interaction between study site and covariates will be carried out to determine if the effect of each covariate differs across studies, to test the assumptions of the 1-stage pooled fixed effect analysis. Of course, the significance of interaction terms are themselves influenced by sample size and should be interpreted also with caution. Particular care needs to be taken if heterogeneity is found between study sites when using a 2-stage random effects approach (see above section). Reasons for heterogeneity need to be thoroughly investigated and the assumptions underlying the 1-stage pooling approach need to be revisited.

In a sensitivity analysis, a 1-stage pooled approach with study site as random intercept and vaccination as random slope can be compared to the analysis with study site as fixed effect.

Two-stage pooling approach

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

Study-specific adjusted ORs and standard errors for the effect of current influenza vaccination obtained from the individual studies, will be combined in a model that incorporates random effects of the studies, to account for unmeasured country- and study-specific factors that differ between studies.

The study-specific exposure-disease effects (ORs) are then weighted by the inverse of their marginal variances. The marginal variance is the sum of the individual study-specific variances and the variance of the random study effects (τ^2). This will give the pooled odds ratio and standard error. See Annex 7 for an example of Stata syntax.

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

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

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

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

Pooled analysis

The following analyses (1-stage or 2-stage) will be carried out if sample size permits:

- on all data and separately with cases/control restricted to an interval between date of onset of symptoms and swab taken of <4 days;
- for VE against (sub)type-specific influenza, influenza B by lineage and VE for strain-specific genetic groups
- for the various types of vaccines (adjuvanted/non-adjuvanted; high-dose vs. normal dose, trivalent vs. quadrivalent, live-attenuated vs. inactivated, egg-based vs. cell-based), groups of vaccines (split virion, subunit, etc.), mode of injection (intradermal vs. intramuscular) and by vaccine brand.

All analyses are done separately for seasonal and pandemic vaccine (if applicable).

Analyses by vaccine brand will include only countries (or regions if the information is available) where the brand is available. Countries (or regions if information is available) where the vaccine product is not available will be excluded from the analysis.

Stratified analysis

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

- age groups (0-14, 15-64 and 65+ years);
- presence of at least one chronic condition;
- time: early influenza season, peak, late influenza season.

Further pooled analyses

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

- VE at different time points along the season (e.g. VE by week or group of weeks in the season [VE for weeks 2-3, 4-5, 6-7, etc.] or by month)
- VE by time since vaccination. Time since vaccination can be calculated by subtracting the date of vaccination from the date of onset. Time since vaccination can then be modelled as a categorical or continuous variable.
- VE of previous season influenza vaccination only, current season influenza only and combined season vaccination. Where possible, influenza vaccination information from seasons more than one year before the current season will be taken into account.
- Using the systematic samples of the sequenced viruses, the proportion of virus changes will be calculated at different time points along the season (e.g. by week or group of weeks in the season: weeks 2-3, 4-5, 5-6, etc.). This will be compared to VE at different time points along the season.

Controlling for GP effect

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

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

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

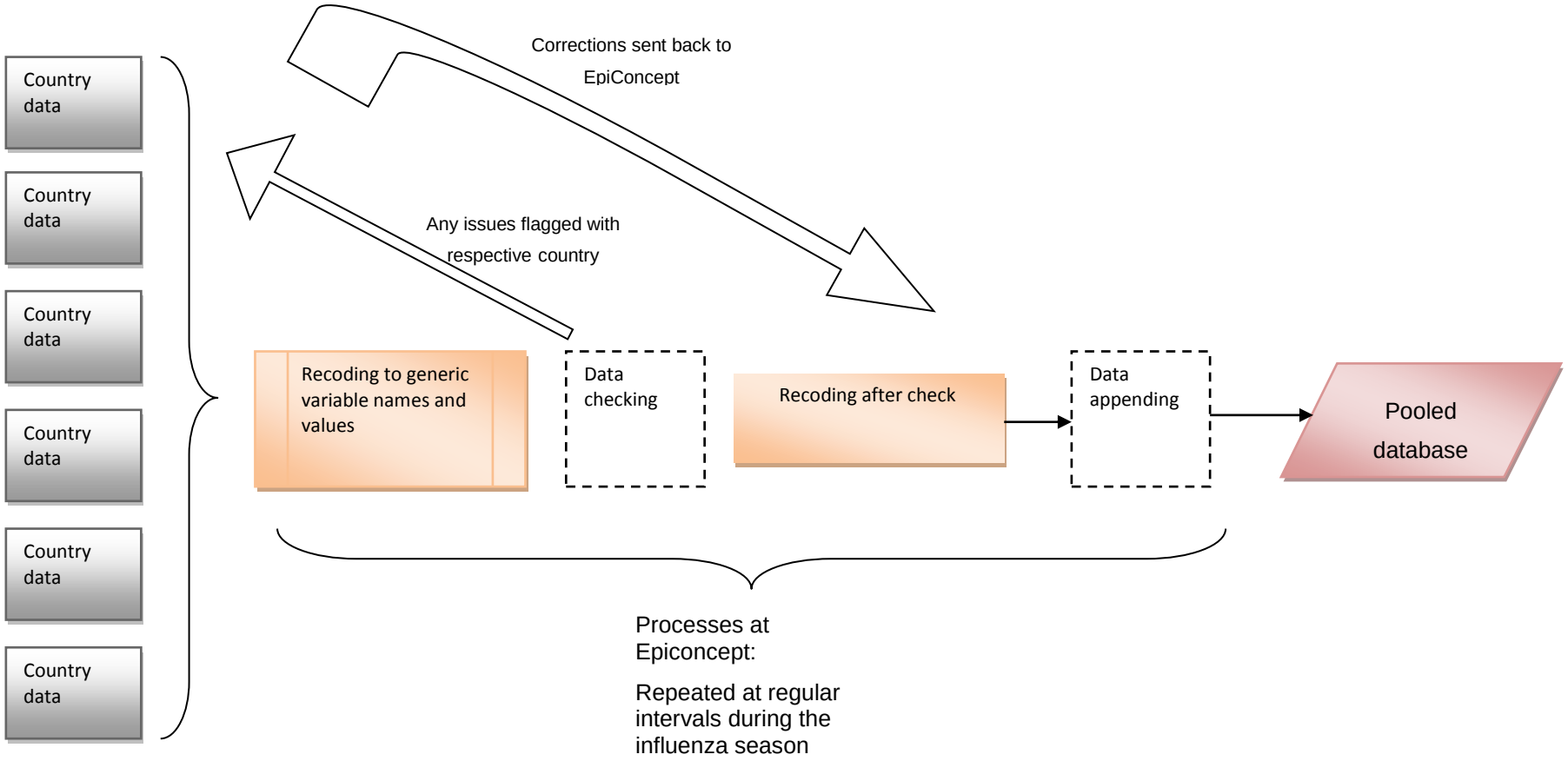
Continuous variables

Continuous variables in the I-MOVE datasets include age, time of onset of symptoms and GP visits. 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. Models will be compared with the continuous variable coded as categories, a linear term, polynomial or a spline – the Akaike information criterion (AIC) can be used for comparison if the data in the models are the same (the lower the AIC, the better). Coefficients and standard errors of the parameters (e.g. the individual spline parameters) will be looked at and models where the standard errors exceed the coefficients will not be used. In addition, a balance will be sought between simplicity of a model (so a non-expert can understand what is going on) and a model with the least bias.

GP level data

If available, study sites will provide data by number of ILI seen by GP by age group and numbers swabbed, in order to assess compliance to the protocol. In addition, if not all elderly meeting the ILI case definition are swabbed by the GPs, then we can calculate a sampling fraction by GP to correct for this.

Annex 5: Data flow for pooled database



Countries send their individual data to Epiconcept according to minimum dataset

Annex 6: Generated/recoded variables

Variable name	Type	Values and coding	Definition
cases	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for any influenza type.
casea	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for any influenza A type.
caseh1	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for influenza type A(H1N1).
caseh3	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for influenza type. A(H3N2)
caseb	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for any influenza B type (regardless of lineage).
caseby	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed influenza B Yamagata lineage.
casebv	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for influenza B Victoria lineage.
ili	Numeric (binary)	0 = No 1 = Yes	Variable that corresponds to EU ILI case definition (coded using the symptoms in dataset)
svaccdelay	Numeric (continuous)	Integer	Number of days between seasonal flu vaccination date and onset date of symptoms (needs to be modified if 2 doses are required)
svacc	Numeric (binary)	0 = No 1 = Yes	Coded as yes if >14 days between seasonal vaccination and onset of symptoms
swabdelay	Numeric (continuous)	Unique integer	Number of days between onset date of symptoms and swab date
swabless4	Numeric (binary)	0 = No 1 = Yes	1 indicates less than 4 days between symptom onset and swab date. 0 indicates more than 3 days.

Variable name	Type	Values and coding	Definition
anychron	Numeric (binary)	0 = No 1 = Yes	0 indicates no chronic disease for which flu vaccination is recommended 1 indicates at least 1 chronic disease for which flu vaccination is recommended
numchron	Numeric (continuous)	Unique integer	Number of chronic diseases reported for the patient.
twochron	Numeric (binary)	0 = No 1 = Yes	0 indicates no or only one chronic disease for which flu vaccination is recommended 1 indicates at least 2 chronic diseases for which flu vaccination is recommended
smokcurr	Numeric (binary)	0 = No 1 = Yes	Current smoker (1) vs. former or never smoker (0).
hosp_bin	Numeric (binary)	0 = No 1 = Yes	Not hospitalized for chronic disease in past 12 months (0), hospitalized for chronic disease in past 12 months (1)
gpvisitgp	Numeric (categorical)	0 = 0-1 visit 1 = 2-4 visits 2 = 5+ visits	The continuous variable GP visit is grouped into categories.
agegp10	Numeric (categorical)	0 = 65-74 years 1 = 75-84 years 2 = 85+ years	The continuous variable age is grouped into 10 year age groups, (although often splines are used for analysis of this continuous variable)
agegroup	Numeric (categorical)	0 = 65-74 years 1 = 75-max years	The continuous variable age visit is grouped into 2 age groups, used for stratification.
onsetweek1	Continuous	Integer	Week of onset of ILI symptoms, coded according to ISO weeks
adj	Numeric (categorical)	0 = Not vaccinated 1 = Non-adjuvanted 2 = Adjuvanted 8 = Vaccinated, brand unknown, 9 = Vaccination status unknown	Persons with adjuvanted vaccine received >14 days before symptom onset are coded as 1, those who received non-adjuvated vaccine >14 days before symptom onset are coded as 2 and those unvaccinated or vaccinated <15 days before symptom onset are coded as 0.

Variable name	Type	Values and coding	Definition
vaccgroup	Numeric (categorical)	0 = Not vaccinated 1 = Inactivated trivalent subunit (egg propagated) 2 = Inactivated trivalent split virion (egg propagated) 3 = Adjuvanted 4 = Inactivated trivalent subunit (egg propagated) 4 = Inactivated trivalent subunit (cell propagated) 5 = Inactivated quadrivalent subunit (egg propagated) 6 = Inactivated quadrivalent subunit (cell propagated) 7 = LAIV 8 = Vaccinated, unknown brand 9 = Vaccination status unknown	Classification of the different vaccine groups (this may change over time)
vaccval	Numeric (categorical)	0 = Not vaccinated 1 = Vaccinated with trivalent vaccine 2 = Vaccinated with quadrivalent vaccine 8 = Vaccinated, product unknown, 9 = Vaccination status unknown	Persons with trivalent vaccine received >14 days before symptom onset are coded as 1, those who received quadrivalent vaccine >14 days before symptom onset are coded as 2 and those unvaccinated or vaccinated <15 days before symptom onset are coded as 0.
vaccmode	Numeric (categorical)	0 = Not vaccinated 1 = Vaccinated intramuscularly 2 = Vaccinated intradermally 8 = Vaccinated, vaccination mode unknown 9 = Vaccination status unknown	Mode of vaccination
vaccproc	Numeric (categorical)	0 = Not vaccinated 1 = Egg-based manufacturing 2 = Cell-based manufacturing 8 = Vaccinated, manufacturing	Process of vaccine manufacturing

Variable name	Type	Values and coding	Definition
		process unknown 9 = Vaccination status unknown	

Annex 7: Stata syntax

Syntax for 2-stage pooling model:

```
// using pooled dataset with a variable for study
gen study=""
gen logor=.
gen or=.
gen logse=.

// With the loop below we are calculating the OR, the log OR and the log standard error for each study. Only these data will be used for the 2-stage pooled analyses.

local i=1
foreach country in country1 country2 country3 country4 {          // replace "country n" with country/study abbreviation
logistic cases svacc i.agegroup sex anychron smokcurr hosp_bin gpvisit i.onsetweek1 if idcountry=="`country'"
matrix b = e(b)
matrix se = e(V)
replace study="`country'" in `i'                                   // here we are creating a summary dataset with 1 row per study
replace logor= b[1,1] in `i'
replace logse=sqrt(se[1,1]) in `i'
replace or=exp(b[1,1]) in `i'
```

```
local ++i
}

// Dropping data, so only the variables interesting for the 2-level model remain:
keep if study!=""                                   // now our dataset only has 1 line per study
save twostage.dta, replace
metan logor logse, effect(Odds ratio) eform xlabel(0.25, 0.5, 1, 1.25, 1.5) textsize(250)  label(namevar=study) randomi
// Above is the meta-analysis command that uses the log OR and log SE to carry out a 2-stage random effects pooled analysis
// Outputs are the individual and pooled OR estimates and confidence intervals as well as a forest plot
```

Syntax for 1-stage pooling model:

```
// using pooled dataset with a variable for study
xi: logistic cases svacc i.agegroup sex anychron smokcurr hosp_bin gpvisit i.onsetweek i.idcountry
```

Stata syntax serves as guidance only and syntax should be adapted to the given situation

Annex 8: Study-specific annexes

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

- description of the GPs participating in the study (number, distribution, catchment population, mode of recruitment);
- definition of beginning, peak, end of influenza season;
- ILI cases: specify if all ILI cases are recruited or a simple random or systematic sample is taken;
- seasonal and pandemic (if applicable) vaccines used;
- vaccine ascertainment method;
- information on application of ICD or ICHPPC-2 codes;
- sample size calculation;
- 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, if applicable);
- procedures for GDPR compliance
- human resources needed;
- provisions to train GPs.