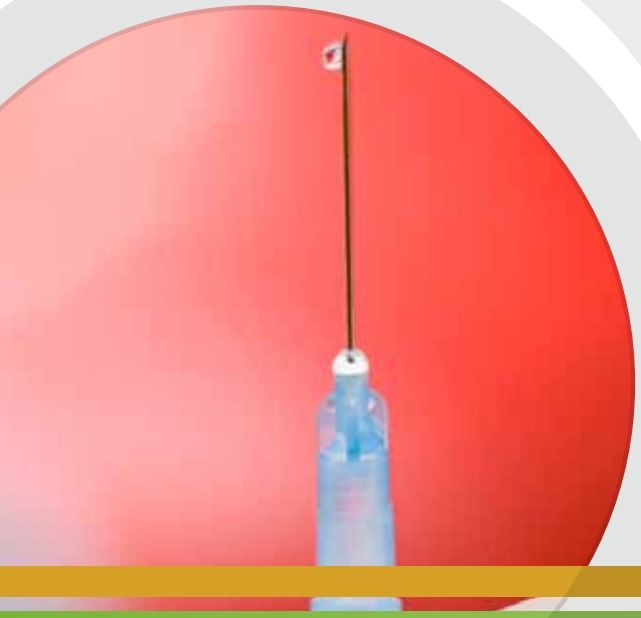


TECHNICAL DOCUMENT

Protocol for case-control studies to measure influenza vaccine effectiveness in the European Union and European Economic Area Member States

www.ecdc.europa.eu

ECDC TECHNICAL DOCUMENT

Protocol for case-control studies to measure pandemic and seasonal influenza vaccine effectiveness in the European Union and European Economic Area Member States

This report was commissioned by the European Centre for Disease Prevention and Control, coordinated by Bruno Ciancio (ECDC), and produced by

Esther Kissling, Alain Moren, and Marta Valenciano — EpiConcept, Paris, France.

Revised edition

Stockholm, November 2009

First published Stockholm, July 2009

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

þ (Tick/check mark indicates the sections that Member States should adapt and detail in their study annexes.)

1 Background

Influenza viruses constantly evolve, vaccines are reformulated every year. Therefore, vaccine effectiveness (VE) estimates from previous years cannot simply be carried over to subsequent years.

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 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, need for booster doses);
- better manage and respond to expected reports of vaccine failures (especially during a pandemic); and
- counterbalance the reports of adverse events following immunisation by providing elements for adequate risk management and cost-effectiveness analysis.

The recent occurrence of the A(H1N1)v pandemic augments the importance of obtaining reliable and early vaccine effectiveness estimates for the pandemic vaccine. It takes approximately four to six months for the first supplies of strain-specific vaccines to become available once a new strain of influenza virus is identified and isolated. Consequently, by the time the vaccines are first administered, the virus already circulates and VE results are needed rapidly. In addition, vaccine availability is likely to increase over time – depending on vaccine production output and the licensing of additional vaccines, which means that IVE measurements need to be repeated over time during the pandemic. VE studies may help determining if the seasonal influenza vaccination is effective or interacts with pandemic vaccine in protecting against the A(H1N1)v virus.

Currently, only observational studies can be used to provide real-time estimates of influenza VE in Europe.

In order to provide early estimates of influenza VE, it is necessary to define which methods can be adopted in the European Union and the European Economic Area (EU/EEA) Member States. These methods have to take into account the specific situation of each Member State in terms of resources and available data. It is assumed that expertise developed during seasonal influenza season can be applied when measuring influenza VE during a pandemic.

During the 2008-09 influenza season, the European Centre for Disease Prevention and Control (ECDC) funded five pilot case-control studies (Portugal, Spain, Denmark, Romania, Hungary) — coordinated by EpiConcept — to measure influenza vaccine effectiveness in the elderly (>= 60 or >= 65 years old). The five pilot studies were based on GP surveillance networks. Four of them used the influenza-like illness (ILI) EU case definition. All of them included a common set of variables to adjust for positive and negative confounding.

The five case-control study teams met in Madrid in March 2009 to discuss the preliminary results of the pilot studies. One of the main limitations identified was the small sample size in each of the studies. The expert group recommended that the possibility of conducting a pooled analysis should be studied by using individual data from the pilot studies conducted in the 2008-09 season. In addition, it was suggested that a common core protocol for the 2009-10 season should be developed, including a pooled analysis to facilitate more precise VE estimates at the European level. The core protocol would determine the study design and the standard methods to be used in each of the studies (case definitions, exposure, outcome, minimum covariates to be collected, and their definition and coding).

This publication presents the core European protocol, outlining the agreed methods for measuring pandemic and seasonal VE for each of the individual studies. The protocol includes a proposed plan for pooled analysis. For groups interested in adhering to the core protocol, the specificities of each study will be detailed in the study annexes. Country-specific annexes will be developed by each study team. The collected variables, their definition, and the plan of analysis were revised based on the results of the pilot studies 2008-09, the recommendations of a workshop entitled 'Monitoring vaccine effectiveness during seasonal and pandemic influenza in EU' (Lisbon, 15 to 17 June 2009), and the state of knowledge on the pandemic in August 2009. The protocol will be updated according to a) the final vaccination strategy in each participating Member State (target groups, vaccine delivery, number of doses); b) the time the vaccine will become available; c) the extent of virus circulation; d) the identification of new groups at risk; and e) the potential clinical evolution of the A (H1N1)v infection. The study will be conducted in accordance with Good Epidemiological Practice (GEP) guidelines (http://www.dundee.ac.uk/iea/GEP07.htm).

2 Objectives

2.1 Primary objectives

The primary objectives will be to measure in EU/EEA countries:

- pandemic influenza vaccine effectiveness in the target groups;
- seasonal influenza vaccine effectiveness among people aged 65 years and above in EU/EEA countries.

2.2 Secondary objectives

· Pandemic and seasonal vaccine:
  to estimate pandemic and seasonal VE:
  – in each of the participating countries;
  – by risk groups.

· Seasonal vaccine:
  – to estimate VE by influenza subtype;
  – to provide intra-seasonal VE estimates;
  – to monitor VE every year.

- Pandemic vaccine:
  – to provide early VE estimates;
  – to estimate VE for one and two doses;
  – to estimate VE by vaccine brand.

3 Methods

3.1 Study design

- Case-control study in each participating country, with various sets of controls.
- Multicentre case-control study in several countries, with various sets of controls.

3.2 Study population

Residents (permanent or visitors) of the participating GPs catchment area.

- Seasonal vaccine: the study population will be community-dwelling individuals aged 65 years and above with no contra-indication for influenza vaccination.
- For the pandemic vaccines, the study population will be the population targeted by the vaccine with no contra-indication for vaccination with the pandemic vaccine.

According to ECDC (http://www.ecdc.europa.eu/en/publications/Publications/Forms/ECDC_DispForm.aspx?ID=388; 25 August 2009), the target groups to be considered for the pandemic vaccine are:

- people aged less than 65 years with chronic underlying conditions, namely:
  - chronic respiratory diseases;
  - chronic cardiovascular diseases;
  - chronic metabolic disorders (notably diabetes);
  - chronic renal and hepatic diseases;
  - persons with deficient immunity (congenital or acquired);
  - chronic neurological or neuromuscular conditions;
  - any other condition that impairs a person's immunity or prejudices their respiratory function;
- young children (especially under the age of 2 years); and
- pregnant women.

The target groups may differ between countries. Other risk groups may be included during the pandemic.

þ Target groups for each study will be listed in the study annexes.

3.3 Study period

The study period will start when the influenza virus is circulating and the vaccine is available.

- Seasonal vaccine: the study period will start at the beginning of the seasonal influenza period and finish at the end of the influenza period.
  þ Each study will define 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).
- Pandemic vaccine: the study period will be defined depending on the gradual availability of vaccines and the pandemic incidence.
  þ Each study will define the beginning and end of the pandemic VE study period.

3.4 Outcome

The outcome of interest will be laboratory-confirmed influenza.

3.5 Cases

ILI definition

A case of influenza like illness (ILI) will be defined as an individual who consults a participating GP, presenting a 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; and
- shortness of breath.

Influenza case

An influenza case will be defined as an ILI case with a respiratory sample positive for influenza.

A pandemic influenza case will be defined as an ILI case with a respiratory sample positive for A(H1N1)v.

þ Indicators to define cases will be specified in the study annexes.

Laboratory confirmation

Specimens will be collected from ILI cases who consult their GP within seven days of symptom onset.

þ Mode of specimen collection, storage and transport for each study will be listed in the study annexes.

Influenza laboratory confirmation will be provided by using RT-PCR and/or culture.

þ RT-PCR characteristics for each study will be listed in the study annexes.

Isolates will undergo a molecular analysis for currently circulating influenza A viruses (subtypes H3 and H1) and (H1N1)v and influenza B. A systematic subset will undergo gene sequencing.

þ The selection of isolates for each study will be specified in the study annexes.

3.6 Case finding

Case identification

Cases will be identified among patients that present with ILI to a participating GP.

Following the procedures outlined by each study, all ILI cases (or a systematic sample of them) will be selected and asked to provide a nasal/throat swab specimen for influenza testing. Influenza-positive ILI cases will be considered as influenza cases.

þ Descriptions of the GPs participating in each of the studies (number, distribution, catchment population) will be available in the study annexes.

þ Descriptions of procedures on how to select ILI cases for swabbing in each of the targeted groups will be available in the study annexes.

Case inclusion criteria

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

þ 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.

Reasons for exclusion will be documented.

3.7 Controls

Various control groups can be selected depending on the available resources. All studies will include at least control group 1 (ILI influenza negative) as defined below.

- Seasonal vaccine: countries will assess whether other control groups could be used in addition to control group 1 (control group selection depends on feasibility).

þ The various control groups selected by each study will be detailed in the study annexes.

- Control group 1, ILI flu negative:
  - ILI cases that tested negative for influenza will be included in control group 1.
- Control group 2, non-ILI GP clients (density case-control design):
  - Controls will be selected among GP clients seen at the GP office and selected concurrently to cases (+/− one week from the date of consultation of the corresponding case). Controls will be selected by simple or systematic random sampling among clients who did not yet develop ILI during the influenza season.
    þ Procedures for selection in each of the studies will be detailed in the study annexes.
    If feasible, controls should be matched by age group (< 75 years; >= 75 years).
- Control group 3, GP clients (case-cohort design):
  - Controls will be selected randomly from a list of GP patients.
    þ Procedures for selection in each of the studies will be detailed in the study annexes.
- Control group 4, community controls (density case-control design):
  - Controls are randomly selected individuals living in the same geographical area (defined by postcode or other geographical/administrative division) as the corresponding case. Control individuals have not suffered from ILI during the current influenza season until the corresponding case became ill.
    þ Procedures for selection will be detailed in the study annexes.
    If feasible, controls should be matched by age group (< 75 years; >= 75 years).
- Control group 5, community controls (case cohort):
  - Controls will be randomly selected individuals living in the same geographical area of the corresponding case.
    þ Procedures for selection will be detailed in the study annexes.
    If feasible, controls may be matched by age group (< 75 years; >= 75 years).
- Control group 6, community controls (screening method):
  - For each of the target age groups, vaccine coverage of cases will be compared with the vaccine coverage of the population in the GP catchment area.
    þ Procedures for estimating VC will be detailed in the study annexes.

Control exclusion criteria

Controls will be 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.

Pandemic vaccine: Controls will be excluded if they have previously been diagnosed with lab-confirmed A(H1N1)v after the start of the pandemic.

Reasons for exclusion will be documented.

3.8 Exposure (vaccination)

Definition of vaccination status

- Seasonal vaccine: An individual is considered as vaccinated against influenza if the vaccination occurred more than 14 days before disease onset or more than 14 days before being selected as a control.

- Pandemic vaccine: An individual is considered:
  - *fully vaccinated* if he/she has received two doses of the vaccine more than 14 days before developing ILI symptoms or before being selected as a control;
  - *partially vaccinated* if he/she has received:
    a) only one dose of the vaccine more than 14 days before developing ILI symptoms (or more than 14 days before being selected as a control);
    b) a second dose less than 15 days before developing ILI symptoms (or more than 14 days before being selected as a control);
  - *unvaccinated* if he/she has not received any vaccine or if he/she has received one dose less than 15 days before developing symptoms.

The definition of the vaccination status may be revised and should be taken into account in the analysis if preliminary clinical data indicate that one dose is sufficient to confer protection, or that the immunisation delay is different from the currently assumed 14 days.

Vaccination status ascertainment

The exposure of interest in this study will be vaccination history with trivalent influenza vaccine (for seasonal vaccine) and vaccination history with the pandemic vaccine (once the vaccine becomes available). The vaccination history includes date of administration and brand names. Documenting the flu batch codes 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 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.

- Pandemic vaccine: The number of doses will be documented.
  - þ Seasonal and pandemic vaccines used will be detailed in the study annexes.
  - þ The precise mode of vaccine ascertainment for each study will be specified in the study annexes.

3.9 Risk groups

Seasonal vaccine

Individuals will be considered to belong to a risk group if this is indicated in the GP records or if the patient reports suffering from one of the underlying conditions included in the interview questionnaire (see below).

Pandemic vaccine

According to ECDC (http://www.ecdc.europa.eu/en/publications/Publications/Forms/ECDC_DispForm.aspx?ID=388; 25 August 2009), the risk groups to be considered for the pandemic vaccine are:

- people aged less than 65 years with chronic underlying conditions, namely:
  - chronic respiratory diseases;
  - chronic cardiovascular diseases;
  - chronic metabolic disorders (notably diabetes);
  - chronic renal and hepatic diseases;
  - persons with deficient immunity (congenital or acquired);
  - chronic neurological or neuromuscular conditions;
  - any other condition that impairs a person's immunity or prejudices their respiratory function;
- young children (especially under the age of 2 years); and
- pregnant women.

The target groups may differ between countries. Other risk groups may be added during the pandemic.

þ Risk groups for each study will be listed in the study annexes.

3.10 Confounding factors and effect modifiers

Chronic diseases

Seasonal vaccine:

The list of underlying conditions in the questionnaire should include at least those for which the vaccine is recommended in the study country and may include:

- 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 (asthma, chronic bronchitis, bronchopulmonary dysplasia, cystic fibrosis);
- chronic renal diseases (chronic renal failure);
- immunodeficiency (conditions that suppress the immune function due to underlying disease and/or therapy – e.g. people receiving chemotherapy, HIV infection); and
- anaemia.

þ The list of chronic diseases for which the seasonal vaccine is recommended will be part of the study annexes.

Pandemic vaccine:

The list of underlying conditions in the questionnaire should include all conditions that define the risk groups in each of the study countries (see 'risk groups' above).

þ The list of chronic diseases for which the pandemic vaccine is recommended will be part of the study annexes.

Severity

The severity of the underlying conditions will 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 will be collected and coded as follows: never-smoker, former smoker (stopped smoking at least one year before inclusion in the study), current smoker.

Previous vaccinations

Vaccination against seasonal influenza in the last two seasons (recording vaccination information for each influenza season).

Functional status

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

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 all GP visits in the 12 months before inclusion in the study will be recorded.

Antiviral administration

Use of antivirals will be documented: type and date of administration.

Source of information

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

Seasonal vaccine: For community controls, the procedures for data collection will be defined by each study coordinator.

þ The precise modes of data collection for community controls will be specified in the study annexes.

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.

3.11 Sample size

A minimum of 30 GPs will be required for each of the studies.

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 to include a big enough sample size to obtain exact estimates for each separate study.

þ The sample size calculation for each study will be detailed in the study annexes.

Table 1 illustrates the various sample sizes that would ensure an alpha error of 0.05, a power of 0.8 or 0.9, a detectable odds ratio ranging from 0.3 to 0.6, and a vaccine coverage among the source population (or among controls) ranging from 50 to 70 %.

Table 1: Sample size calculations

Power	Alpha	Controls/ case	Vaccine coverage in source population/controls	Detectable OR	Number of cases	Number of controls
0.90	0.05	1	0.5	0.6	345	345
0.80	0.05	1	0.5	0.6	262	262
0.90	0.05	1	0.5	0.5	194	194
0.80	0.05	1	0.5	0.5	148	148
0.90	0.05	1	0.5	0.4	116	116
0.80	0.05	1	0.5	0.4	89	89
0.90	0.05	1	0.5	0.3	72	72
0.80	0.05	1	0.5	0.3	56	56
0.90	0.05	1	0.6	0.6	341	341
0.80	0.05	1	0.6	0.6	259	259
0.90	0.05	1	0.6	0.5	188	188
0.80	0.05	1	0.6	0.5	144	144
0.90	0.05	1	0.6	0.4	110	110
0.80	0.05	1	0.6	0.4	85	85
0.90	0.05	1	0.6	0.3	67	67
0.80	0.05	1	0.6	0.3	52	52
0.90	0.05	1	0.7	0.6	370	370
0.80	0.05	1	0.7	0.6	281	281
0.90	0.05	1	0.7	0.5	200	200
0.80	0.05	1	0.7	0.5	153	153
0.90	0.05	1	0.7	0.4	115	115
0.80	0.05	1	0.7	0.4	88	88
0.90	0.05	1	0.7	0.3	67	67
0.80	0.05	1	0.7	0.3	52	52

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

3.12 Data

Data on cases and GP controls will be collected at GP office level. GPs 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).

þ Data collection methods for the community controls will be detailed in the study annexes.

EpiConcept may develop an electronic questionnaire and a web-based questionnaire for participating GPs.

Double data entry will be required unless medical electronic records are used.

þ Details on data collection methods, data entry and data transmission will be available in the study annexes.

Collected information

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

- study identification: country and GP;
- case/control demographics;
- signs, symptoms;
- date of onset of ILI;
- date of swabbing;

- laboratory results;
- selected underlying chronic conditions (including diabetes, heart disease, chronic obstructing pulmonary disorder, renal diseases, and immunodeficiencies);
- number of hospitalisations for chronic diseases in the previous 12 months;
- total number of GP visits in the previous 12 months;
- smoking history;
- current season influenza vaccination including date;
- pandemic vaccination including number of doses, date, brand;
- influenza vaccination in the previous two seasons;
- pregnancy status;
- functional status;
- antiviral administration; and
- information on laboratory confirmed A(H1N1)v influenza since the beginning of the pandemic.

Pandemic vaccine: Collected data will be revised as more information on the vaccine and the target groups becomes available.

Data 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 will be checked against the medical records and against the study database.

The agreement between patient vaccine records/vaccination status reported by study participant/vaccine registries will be measured.

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

3.13 Analysis

Seasonal vaccine: Analysis will be restricted to the groups in each of the participating countries for which seasonal vaccine is recommended.

Pandemic vaccine: Analysis will be restricted to the groups in each of the participating countries for which the pandemic vaccine is recommended.

Analysis will be carried out first for each individual study. In a second step, a pooled analysis is conducted.

All analyses will be done separately for seasonal and pandemic vaccine.

Analyses will be conducted for:

- all data – restricted to cases/controls from which swabs were taken < 4 days since the date of onset of symptoms;
- overall VE and strain-specific VE, sample size permitting; and
- the pooled analysis, separately for studies with different control groups (using an unmatched or a matched analysis as appropriate).

Pandemic vaccine: In addition, analysis is conducted for each specific brand.

Individual study analysis

Descriptive and univariable analysis

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

Study participants will be described by baseline characteristics. Baseline characteristics of cases and controls in unmatched studies will be compared using the chi-square test, Fisher's exact test, t-test or the Mann-Whitney test (depending on the nature of the variable and the sample size). In matched case-control studies, characteristics of cases and controls will be compared using McNemar's chi-square test, paired t-test, conditional logistic regression, or the Wilcoxon signed-rank test (depending on the nature of the variable and the number of controls).

The association between vaccination status and baseline characteristics will be measured for both case and control groups.

Measure of effect

Vaccine effectiveness will be computed as VE = 1 – OR. An exact 95 % confidence interval will be computed around the point estimate. (For studies using the screening method, refer to the ECDC generic screening protocol.)

Stratified analysis

Analysis will be stratified according to:

Seasonal vaccine:

- Age groups < 75 years and > 74 years.
- Presence or absence of high-risk conditions.
- Time: early influenza season, peak, late influenza season.
- Virus strain.

Pandemic vaccine:

- Vaccine brand.
- Time: different periods defined by pandemic vaccine availability.
- Target groups.

A sufficient sample size should be planned in order to ensure enough individuals in each stratum for a precise estimate. For example, assuming a vaccination coverage of 50% and a VE of 60%, studies should aim to have at least 90 cases and 90 controls in each of the strata. Effect modification will be assessed comparing the OR across the strata of the baseline characteristics. Confounding will be assessed by comparing crude and adjusted OR for each baseline characteristic.

Multivariable analysis

A multivariable (conditional if matched) logistic regression analysis will be conducted to control for negative and positive confounding. Odds ratios and standard errors will be obtained. Variables will be tested for multicollinearity. Interactions will be tested using the likelihood ratio test or Wald's test and will be included in the model if significant at the 5 % level. Factors other than statistical significance will also be used as criteria for inclusion of an interaction term.

Pooled analysis

See Annex 2 for data management of pooled data and Annex 3 for pooled data analysis.

Pandemic vaccine

For the pooled data, interim analyses will be conducted in different periods according to the available sample size. Based on the available sample size, different outputs will be estimated in each of the interim analyses. For each output, a minimum of 90 cases and 90 controls will be needed (assuming a vaccination coverage in the source population of 50% and a detectable OR of 0.4; see Table 1 on 'sample size calculations').

Table 2: Sample size calculation

Proportion vaccinated	OR	Number of cases	Number of controls
0.5	0.4	89	89

Table 3: Sample size interim analysis

Outputs	Sample size needed in pooled data	Sample size needed for each individual site
Pooled crude VE for all vaccines and target groups	90 cases*	
Pooled adjusted VE for all vaccines and target groups	90 cases*	
Pooled VE stratified by specific variables (e.g. age group, target group)	90 cases per stratum*	
Crude VE by study site for all vaccines and target groups		90 cases per study site
Pooled adjusted VE for all vaccines and target groups		90 cases per study site
Stratum-specific, site-specific VE (e.g. age group, target group		90 cases per study site and stratum
Pooled brand VE	90 cases for each specific brand analysis*	
Pooled brand analysis stratified by specific variables	90 cases per stratum for each specific brand analysis*	
Site-specific, brand-specific, stratum-specific VE (e.g. age group, target group)		90 cases per study site and stratum for each specific brand analysis

** Using one-staged pooled analysis (see Annex 3 for further details)*

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 ILI incidence, the influenza incidence, the vaccination coverage, and the swabbing scheme used by GPs and the number of GPs (see Excel calculator, attached to the PDF version of this document).

The sample size suggested for brand-specific VE is an estimate and cannot be precisely defined at this stage. The sample size for brand VE will depend on the vaccination coverage for each brand and therefore on the market share.

3.14 Data management

Individual analysis

EpiConcept will provide 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 EpiConcept's web-based data collection methods will be not used, data can be coded as outlined in Annex 1. Alternatively, study teams can provide EpiConcept with a codebook that includes the variable names, variable descriptions, and the coding of variable values.

Data cleaning

Summary and frequency tables as well as visual representations of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of swabbing before date of onset of symptoms). These values will be checked against the questionnaires or queried with the GP. Any changes to the data will be 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 will be provided if so desired.

Pooled analysis

EpiConcept conducts the pooled analysis. Each individual study will be sent to EpiConcept's study database. All personal identifier information such as names, addresses, and medical registration codes will be deleted before data transmission to EpiConcept, where all individual data will be pooled. A country (or study) identifier will be included in each record (e.g. ESP for Spain, UK for the United Kingdom), a GP code will be included (e.g. a unique number), and each record will be given a unique number. This number will also be included in the study team's database and will be used by EpiConcept and the study teams during pooling, so that records can be traced back whilst maintaining anonymity, if there should be any further queries. Study databases can be sent to EpiConcept in any format. EpiConcept will perform all necessary data cleaning. EpiConcept will document and share any further data cleaning and analysis with all study coordinators to ensure it can be reproduced.

See Annex 3 for detailed guidelines on the pooled analysis.

Missing data

Any missing data will be documented.

If many 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, multiple imputation methods at study level will be used to replace missing values.

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

3.15 Potential biases

Negative confounding

Negative confounding refers to biases that reflect the fact that high risk groups will be more likely to be vaccinated and therefore reduce VE.

Positive confounding

Positive confounding refers to biases that reflect a 'healthy vaccine effect'. People with a healthy lifestyle will be more likely to accept/request vaccination, thus leading to an increase of measured VE.

Positive and negative confounding will be minimised through stratification and multivariable analysis (including presence of chronic diseases) as well as variables that will be collected in order to measure positive and negative

confounding. It will not be possible to rule out the presence of characteristics in the study population for which no information is collected in the study questionnaire and that therefore could lead to positive or negative confounding. Therefore, residual positive or negative confounding may be present. A sensitivity analysis will be conducted to assess the effect of a potential and unmeasured confounding factor.

Representativeness of cases

The study includes only cases consulting a GP for ILI. In the context of the pandemic, health-seeking behaviour may differ by country depending on the case management strategy (e.g. recommendation of not seeing a GP). In some countries, only severe cases will consult a GP. In other countries, severe cases will directly report to emergency rooms without consulting their GPs. The type of cases included in the study should be described for each study and it should be noted how its 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 will be low if there are few studies. 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 vaccine effectiveness.

3.16 Consent

According to country-specific regulations, informed (oral or written) consent will be required from each study participant.

þ Details will be available in the study annexes.

3.17 Dissemination of results

The enrolment of cases/controls will be regularly updated by each study coordinator on the 'I-MOVE in Europe' web page (http://sites.google.com/site/epiflu).

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

Pandemic vaccine: First estimates will be disseminated once the sample size allows for meaningful interpretation (see paragraph on pooled analysis and Excel calculator attached to the PDF version of this document).

Publications, scientific communication

Each study coordinator will decide where the results of the individual studies will be published and which scientific conferences will be 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. The list of authors will include one representative for each of the studies. Co-workers and contributors will be acknowledged. The actual authorship for the pooled article will be discussed with the study teams at the beginning of the study.

3.18 Training

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

4 Logistical aspects

4.1 Study leader

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

4.2 Human resources

In each country, a part-time investigator will be 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.

þ The specific human resources needed in each country will be detailed in the study annexes. EpiConcept will ensure the overall coordination of the various studies.

4.3 Supervision

Site visits and joint workshops will 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 will be composed of two persons from the various project partners.

4.4 Questionnaires

Standardised questionnaires will be developed for the study. The variables used at the European level will be 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 will be conducted at the country level. EpiConcept will provide a structured data entry form. For countries willing to submit data electronically, EpiConcept will provide an online questionnaire.

4.6 Consent

Each study will 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 will be required.

4.7 Further studies

Further potential studies could focus on:

- comparing results obtained from all six control groups (ILI flu negative cases; random or systematic sample of GP patients seen at GP office; random sample of people from the catchment area of GP-selected cases);
- comparing results obtained from various swabbing deadlines, e.g. 2, 3, or 4 days after onset of ILI;
- comparing vaccine status and severity of influenza cases selected through systematic or GP ad hoc selection of ILI cases;
- comparing acceptability and feasibility of various sizes of questionnaires to assess negative confounding;
- identify variables to assess the presence of positive confounding;
- acceptability of the study among GPs;
- comparing GP accessibility for different groups within the local population;
- any other topic that study teams consider helpful in interpreting the study results.

þ Further potential studies will be detailed in the study annexes.

5 Budget

Funding can be used for:

- payment of study supervisor;
- payment of GPs;
- training of GPs; and
- financial support for laboratory tests.

þ Each study group will provide a detailed budget in the study annexes.

6 References

Belongia EA, Kieke BA, Donahue JG, Greenlee RT, Balish A, Foust A, et al; Marshfield Influenza Study Group. Effectiveness of inactivated influenza vaccines varied substantially with antigenic match from the 2004-2005 season to the 2006-2007 season. J Infect Dis. 2009 Jan 15;199(2):159-67.

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).

Carrat F, Tachet A, Rouzioux C, Housset B, Valleron AJ. Field investigation of influenza vaccine effectiveness on morbidity. Vaccine 1998 May;16(9-10):893-8.

Centers for Disease Prevention and Control. Interim within-season estimate of the effectiveness of trivalent inactivated influenza vaccine — Marshfield, Wisconsin, 2007-08 influenza season. Morb Mortal Wkly Rep. 2008 Apr 18;57(15):393-8.

DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. Control Clin Trials 7, 177-88(1986).

Farrington CP. Estimation of vaccine effectiveness using the screening method. Int J Epidemiol 1993 Aug;22(4):742-6.

Groenwold RH, Hoes AW, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: the example of influenza vaccination. Int J Epidemiol. 2008 Dec;37(6):1422-9.

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.

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).

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.

Legrand J, Vergu E, Flahault A. Real-time monitoring of the influenza vaccine field effectiveness. Vaccine 2006 Nov 10;24(44-46):6605-11.

Mazick A, Christiansen AH, Samuelsson S, Molbak K. Using sentinel surveillance to monitor effectiveness of influenza vaccine is feasible: a pilot study in Denmark. Euro Surveill 2006;11(10):254-6.

National Research Council Combining information – Statistical Issues and Opportunites for Research (1992).

Nichol KL, Nordin JD, Nelson DB, Mullooly JP, Hak E. Effectiveness of influenza vaccine in the community-dwelling elderly. N Engl J Med. 2007 Oct 4;357(14):1373-81.

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).

Skowronski DM, Masaro C, Kwindt TL, Mak A, Petric M, Li Y, et al. Estimating vaccine effectiveness against laboratory-confirmed influenza using a sentinel physician network: results from the 2005-2006 season of dual A and B vaccine mismatch in Canada. Vaccine 2007 Apr 12;25(15):2842-51.

Skowronski DM, De Serres G, Dickinson J, Petric M, Mak A, Fonseca K, et al. Component-specific effectiveness of trivalent influenza vaccine as monitored through a sentinel surveillance network in Canada, 2006-2007. J Infect Dis. 2009 Jan 15;199(2):168-79.

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).

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.

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.

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)
casenr	Numeric (continuous)	Integer	Identifies which controls are connected to which case
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
lab_res	Numeric (categorical)	0 = Negative 1 = Positive 8 = Do not know	Laboratory result (positive/negative)
lab_virus	Text		Laboratory result: virus type
lab_subtype	Text		Laboratory result: virus subtype
seasvaccany	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received seasonal influenza vaccine 2009-10
seasvaccdate	Date	dd/mm/yyyy	Vaccination (seasonal vaccine) date

Variable name	Type	Values and coding	Definition
seasvacctype	Text		Type of seasonal vaccine (brand name)
panvaccany	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Received pandemic influenza vaccine 2009-10
panvaccdate1	Date	dd/mm/yyyy	Vaccination with pandemic vaccine date first dose
panvaccdate2	Date	dd/mm/yyyy	Vaccination with pandemic vaccine date second dose
panvacctype	Text		Type of pandemic vaccine (brand name)
panvaccdose	Numeric	0, 1, 2	Number of pandemic doses received
vacc_08	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Previous seasonal influenza vaccination 2008-09
vacc_07	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Previous seasonal influenza vaccination 2007-08
anemia_spleen	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Enlarged spleen, anaemia
cirrhosis	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Cirrhosis
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
hema_cancer	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Hematologic cancer
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
nonhem_cancer	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Nonhematologic cancer
pregn	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Pregnancy
ren_disease	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Renal disease
dem_stroke	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Dementia, stroke
rheumat	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Rheumatologic diseases
severity	Numeric (count)	integer	Number of hospitalisations previous year for the chronic disease
gpvisit	Numeric (count)	integer	Number of GP consultations previous year
fs_bath	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Requires assistance to bathe

Variable name	Type	Values and coding	Definition
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)
prev_h1n1	Numeric (categorical)	0 = No 1 = Yes 8 = Do not know	Lab-confirmed A(H1N1)v prior to study.
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
dens_cc	Numeric (categorical)	0 = No 1 = Yes	Exclusion criteria for density case-control: ILI symptoms in the season

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

Annex 2: 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 4). All personal identifier information such as names, addresses, medical registration codes will be deleted. A country (or study) identifier will be included in each record (e.g. ESP for Spain, UK for the United Kingdom), a GP code will be included (e.g. a unique number) and each record will be given a unique number. This number 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 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 conditions;
- number of hospitalisations for the chronic diseases in the previous year;
- number of GP visits in the previous 12 months;
- smoking history;
- current season influenza vaccination including date;
- pandemic vaccination including number of doses, date, brand;
- influenza vaccination in the previous two seasons;
- pregnancy status;
- functional status;
- antiviral administration; and
- information on laboratory-confirmed A(H1N1)v influenza since the beginning of the pandemic (or other date according to the duration of the pandemic).

For pandemic IVE estimates, data collected will be revised as more information on the vaccine and the target groups become available.

Data transfer

Study databases can be sent to EpiConcept in any format. The minimum dataset can be coded as described in Annex 1. Alternatively, a codebook can be provided to EpiConcept with the variable names, variable descriptions, and the coding of variable values.

Data cleaning

EpiConcept will carry out data cleaning again, once the data are received. Summary and frequency tables and graphic displays of appropriate variables will be used to find illegal, implausible or missing values within the dataset. Checks for inconsistencies will be carried out (e.g. date of swabbing before date of onset of symptoms). Any improbable, illegal or missing values will be reported to the country in questions.

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 recoding

Variables will be recoded and new variables generated according to Annex 5. The recoded data will be stored separately from the crude data, and recoding will be documented.

Missing data

Great care will be taken to avoid missing data.

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

If many 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 will be dropped) and full set analysis (with imputed data).

Data appending and data flow

After data cleaning, the data will be appended, and a unique identifier for each GP will be created by concatenating the study code and the GP code. An example data flow chart is presented in Annex 4.

Annex 3: Pooled data analysis

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

- number of GPs participating and catchment population;
- beginning of the pilot study;
- beginning of influenza period, peak, end;
- beginning of vaccination campaigns for seasonal and pandemic vaccine;
- target groups for the seasonal and pandemic vaccination;
- vaccine brands available;
- proportion of ILI flu positive among all ILI cases; and sample size.
- Analyses will be carried out first for each individual study and then, in a second step, a pooled analysis will be conducted.

Analysis will be done:

- 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 overall VE and influenza type-specific VE if the sample size permits; and
- for the various types of pandemic vaccines, by age group targeted for vaccination.

All analyses will be carried out with Stata v10.1 (Stata Corp LP, College Station, TX, USA).

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 will minimise heterogeneity between studies. However, adherence to the protocol will be checked, and study design and study quality characteristics will also be used to take a qualitative decision whether one or more studies are substantially different from the others and thus should be excluded from the pooled analysis.

Statistical heterogeneity between studies will be tested using Q-test and the I^2 index (see boxes for formulae below).The Q-statistic follows a chi-squared 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.

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

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

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

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.

Two-stage pooling approach

If adequate sample size by study is achieved to obtain an adjusted OR, then a two-stage approach to pooled analysis will be taken. Studies considered to be heterogeneous will not be included in the two-stage model.

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) will then be 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 6 for an example of Stata syntax.

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

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

The study-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 two-stage analysis with fixed study effects, to assess the effects of model assumptions.

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

Stratified analysis

The same two-stage analysis will be carried out for the following strata:

- Age groups:
  – seasonal vaccine < 75 years and > 74 years.
  – pandemic vaccine age groups targeted for pandemic vaccination.
- Presence or absence of high risk conditions.
- Time (early influenza season, peak, late influenza season).
- Virus type.
- Vaccine brand.
- Target groups for pandemic influenza vaccine.

Controlling for GP effect

Primary analysis will be carried out using simple logistic regression to obtain the individual study estimates. However, there could be a GP effect 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 will be carried out for each individual study with GP as a random effect. Then the two-stage model as outlined above will be used to obtain a summary VE measure, using these estimates. The same applies to stratified analyses. Point estimates and confidence intervals of the respective summary VE measures will be compared in a sensitivity analysis.

Small sample sizes

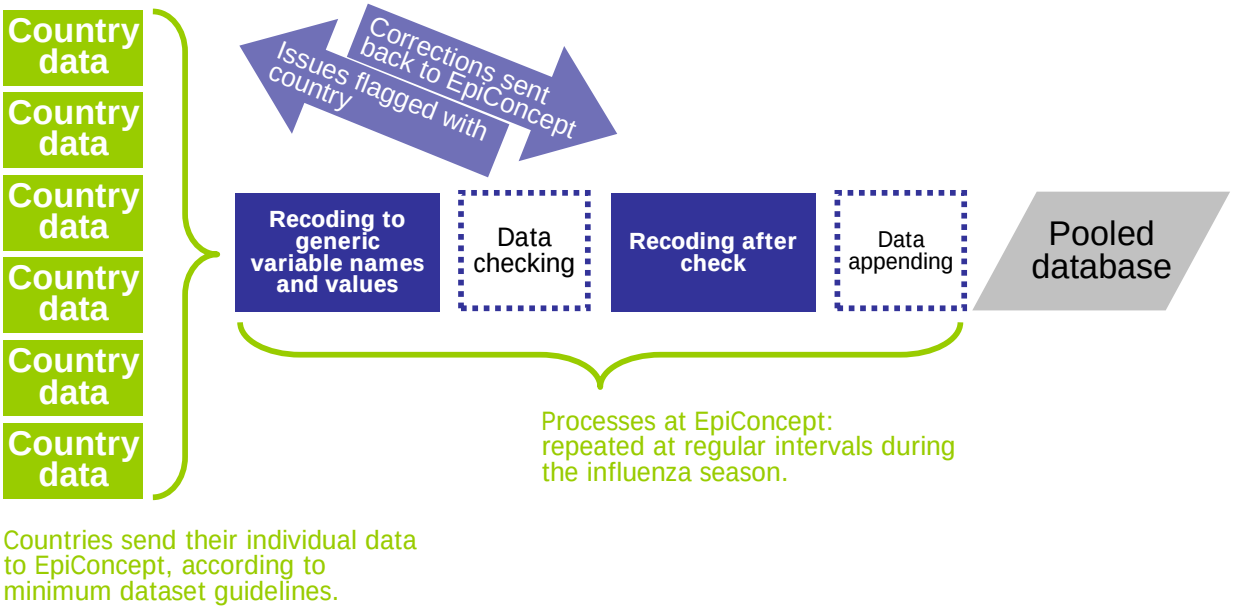
While great effort will be taken to ensure adequate sample size at individual study level, both for overall and stratified estimates, it will still be possible that small sample sizes occur, particularly when estimating early or intra-seasonal VE measures.

When sample sizes at the individual study level are too small to adequately control for confounders, individual study data can be pooled into one dataset and analysed as a one-stage model with study as a fixed effect (see Annex 6 for Stata syntax). This could provide a large enough sample size to obtain, for example, an estimate of VE with reasonable precision early in the season. The results of this analysis should be interpreted with caution, though, as it assumes not only that the underlying true exposure effect is the same in all studies, but also that the association of all covariates with the outcome is the same in all studies.

If a pooled dataset with large sample size is available (e.g. later in the season) then formal tests of interaction between study and covariate will be carried out to determine if the effect of each covariate differs across studies. This way, the validity of using a one-stage model analysis (earlier in the season) can be tested.

Annex 4: Data flow for pooled database

Figure: Data flow for pooled database



Annex 5: Generated/recoded variables

Table: Generated/recoded variables

Variable name	Type	Values and coding	Definition
caseflu	Numeric (binary)	0 = No 1 = Yes	Indicates ILI case that is lab-confirmed for influenza.
ili	Numeric (binary)	0 = No 1 = Yes	Variable that corresponds to EU ILI case definition (coded using the symptoms in dataset).
panvaccdelay	Numeric (continuous)	Integer	Number of days between pandemic influenza vaccination date and onset date of symptoms (needs to be modified if 2 doses are required).
panvacc	Numeric (binary)	0 = No 1 = Yes	Coded as yes if >14 days between pandemic vaccination and onset of symptoms (needs to be modified if 2 doses are required).
seasvaccdelay	Numeric (continuous)	Integer	Number of days between seasonal influenza vaccination date and onset date of symptoms.
seasvacc	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.
anychron	Numeric (categorical)	0 = No 1 = Yes 8 = Don't know	0 indicates no chronic disease. 1 indicates at least 1 chronic disease. 8 indicates no chronic disease reported as 'yes' and at least one chronic disease reported as 'unknown'.
smokcurr	Numeric (binary)	0 = No 1 = Yes	Current smoker (1) vs. former smoker or never-smoker (0).
hosp	Numeric (binary)	0 = No 1 = Yes	Not hospitalised for chronic disease in past 12 months (0). Hospitalized for chronic disease in past 12 months (1).
previousflu	Numeric (binary)	0 = No 1 = Yes	Vaccinated with seasonal flu in any of the two years prior to current season.
weekepi	Continuous	Integer	Week of first case coded as 1, subsequent weeks incremental.

Annex 6: Stata syntax

Syntax for two-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 two-stage pooled analyses.

local i=1
foreach country in country1 country2 country3 country4 {

// replace "countryn" with country/study abbreviation
logistic caseflu seasvacc age75 sex chronic smokcurr functional previousflu hosp gpvisit i.weekepi 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 one-stage pooling model:

```
// using pooled dataset with a variable for study
xi: logistic caseflu seasvacc age75 sex chronic smokcurr functional previousflu hosp gpvisit i.weekepi i.idcountry
```

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

Study-specific annexes

Study specifications for each country will be 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;
- criteria for selection of control groups and definition of control groups to be used;
  (þ Seasonal and pandemic vaccines used.)
- vaccine ascertainment method;
- list of underlying conditions considered for risk groups for seasonal vaccine;
  (þ List of underlying conditions considered for risk groups for seasonal vaccine)
- 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 vaccine strain characterisation);
- consent, ethical procedures (oral/written consent; submission to ethics committee, if applicable);
- human resources needed;
- provisions to train GPs;
- detailed budget;
- estimate of GPs acceptance of EpiConcept's web-based questionnaire; and
- outline of additional studies (if applicable).