IMI2 777363 – DRIVE

Development of Robust and Innovative Vaccine Effectiveness

WP4 – Framework for analysis and study report

D4.1 Framework for analysis of influenza vaccine effectiveness studies - Update

Lead contributor	Anke Stuurman (3 - P95)
Other contributors (original version)	Ulrike Baum, Ritva Syrjänen, Hanna Nohynek (6 - THL)
	Miriam Levi, Paolo Bonanni, Roberto Bonaiuti, Alfredo Vannacci (4 – UNIFI)
	Kaatje Bollaerts, Margarita Riera (3 – P95)
	Jos Nauta (13 – Abbott)
	Gaël Dos Santos, Silvia Damaso, Dominique Rosillon (15 – GSK)
	Hélène Bricout (12 Sanofi Pasteur)
	Josephine van Boxmeer, Mendel Haag (14 – Seqirus)
Other contributors (updated version)	Jos Nauta (13 – Abbott)
	Jorne Bicler (3 – P95)
	Miriam Levi (Azienda USL Toscana Centro)
	Gaël Dos Santos (15 – GSK)
	Bruno Lina (UCBL)

Document History

Version	Date	Description
V0.1	04 DEC 2017	First Draft
V0.2	22 JAN 2018	First Draft – compiled comments
V0.3	26 FEB 2018	Second Draft
V0.4	19 MAR 2018	Second Draft – compiled comments
V0.5	06 MAY 2018	Third Draft
V0.6	04 JUNE 2018	Fourth Draft for SC review
V1.0	26 JUNE 2018	Final Version
V1.1	10 DEC 2021	Update of all chapters
V1.2	10 JAN 2022	Finalization of chapters following WP4 review, review by medical editor, for SC review
V1.3	28 MAR 2022	Finalization of chapters following SC comments, with track changes
V2.0	28 MAR 2022	Final Version of Updated Deliverable

SC: steering committee

Publishable Summary

The aim of this document is to describe a standard set of analytical methods that can be applied to measure IVE in the European context of diverse vaccine products, distribution, recommendations, and administration as well as diverse influenza diagnostic and therapeutic approaches. This document served as guidance to revise the protocols within DRIVE and to document insights gained during the DRIVE project.

The different study designs that can be used to assess influenza vaccine effectiveness are presented, followed by aspects related to exposure and outcome definitions and collection of data. Sources of bias and confounders are explained and how they can be addressed through the design or the analysis. Available laboratory tests for the detection of influenza infections are described. Methods for rapid assessment of IVE are presented. From a data analysis perspective, methods for analysis of individual studies and pooled analyses (one-stage vs. two-stage pooling) are described.

Each chapter concludes with a set of recommendations, and where appropriate, a distinction between studies collecting primary data (test-negative design studies) and studies that make use of secondary data (such as cohort studies using healthcare databases) is made. These recommendations serve as guidance for optimal choices in the design/analysis of studies using the current existing methods for IVE estimation.

Table of Contents

List of abbreviations

AD-MA	Aggregated data meta-analysis
ADVANCE	Accelerated development of vaccine benefit-risk collaboration in Europe
ARI	Acute respiratory infection
CI	Confidence intervals
DFA	Direct fluorescent antibody
DRIVE	Development of robust and innovative vaccine effectiveness
ECDC	European Centre for Disease Prevention and Control
EHR	Electronic healthcare records
EIA	Enzyme immunoassay
ELISA	Enzyme linked immunoabsorbant assay
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
HAI	Hemagglutination inhibition assay
I-MOVE	Influenza - Monitoring Vaccine Effectiveness
IFA	Immunofluorescent antibody test
ILI	Influenza-like illness
IPD-MA	Individual participant data meta-analysis
IVE	Influenza vaccine effectiveness
MAH	Marketing Authorization Holders
NATs	Nucleic acid-based tests
OR	Odds ratio
PCR	Polymerase chain reaction
POC	Point-of-care
RIDT	Rapid diagnostic tests
RR	Relative risk
RSV	Respiratory Syncytial Virus
RT-PCR	Real-time polymerase chain reaction
SAP	Statistical Analysis Plan
SARI	Severe acute respiratory infection
SHR	Single radial hemolysis
SVC	Shell viral culture
TND	Test-negative design
US	United States
VE	Vaccine effectiveness
VN	Virus neutralization assay
VPD	Vaccine preventable disease
WHO	World Health Organization
WP	Work package

1 Background and objective

The main objective of DRIVE is to enable the collaboration of different public and private stakeholders to perform annual VE studies for the various influenza vaccines on the European market. Work package (WP) 4 aims to create a framework to analyze, present, interpret and report influenza vaccine effectiveness (IVE) study results in such a way that all stakeholders can endorse its conclusions. This document constitutes the deliverable of task WP4.1.

1.1 Purpose of this document

This document aims to describe a standard set of analytical methods that can be applied to measure IVE in the European context of diverse vaccine manufacturing, distribution, and administration, as well as diverse influenza diagnostic and therapeutic approaches. This document has served as guidance to revise the protocols within DRIVE. Please refer to D7.3 for novel or innovative methods; these will not be discussed in this document.

This set of methods builds upon existing guidance documents, such as IVE guidance from I-MOVE [1] and the World Health Organization (WHO) [2], and general vaccine effectiveness (VE) guidance from ADVANCE and ENCePP.

The European Medicine Agency (EMA) guidance released in 2016 requires marketing authorization holders (MAHs) to estimate product-specific IVE against laboratory-confirmed influenza on an annual basis [3, 4]. Challenges faced when performing IVE studies and pooling results from those in different countries include bias and confounding, differences in strain circulation between EU countries, and differences in coverage affecting potential herd immunity. Additional challenges to be overcome when estimating brand-specific IVE include timely influenza-vaccine brand identification at individual level, adequate sample size and lack of uniform administration of vaccine types across populations.

This framework was developed during the first year of DRIVE and updated in the final year based on new insights and experiences gained in DRIVE.

1.2 Contents of this document

Each chapter describes one aspect of IVE studies. At the end of each chapter, recommendations are provided for the “optimal” scenario; where appropriate, we distinguish between studies collecting primary data, such as test-negative design studies, and studies that make use of secondary data, such as data from healthcare databases. These

recommendations serve as guidance for the ideal study through existing methods. Importantly, however, following these recommendations are *not* prerequisites for data analyzed within the context of DRIVE, nor do they consider novel or innovative methods.

Chapter 2 *Study designs* reviews the different study designs that can be used for IVE studies and the pros and cons of each.
Chapters 3 *Exposure* and 4 *Outcome* describe data required for exposure and outcome ascertainment and issues of misclassification.
Chapter 5 *Potential biases and confounders expected across different settings* reviews potential biases and confounders, how they can be dealt with in the study design and what information is required to control them in the analyses.
Chapter 6 *Optimization of the value of microbiological and virological information* describes characteristics of different laboratory tests, including their advantages and limitations.
Chapter 7 *Methods for rapid assessment of IVE* deals with methods and considerations for carrying out real-time or rapid IVE assessment.
Chapter 8 *Data analysis for individual studies* describes data analysis methods and considerations by study design.
Chapter 9 *Data pooling* reviews the advantages and limitations of data pooling and describes one-stage and two-stage pooling.
Chapter 10 *Overall recommendations* summarizes and integrates the guidance provided in all the chapters.

2 Study designs

Observational IVE studies use a variety of designs, with the cohort design and the case-control design, specifically the test-negative control design, being the most commonly used designs. There are different variations to these two core designs, with different strengths and limitations.

2.1 Cohort studies

The cohort design is a commonly used and intuitive observational study design.

Cohort studies are intuitive, which makes it easy to communicate their results. Influenza incidence rates can be estimated using cohort studies, which makes these studies useful for disease burden studies. However, cohort studies have strict data requirements with data on exposure, outcome and covariates to be collected for the entire source population. Data collection can be secondary using existing health records or collecting new data specifically for the purpose of a study. For cohort studies using secondary data, typical data sources for cohort studies are linked (vaccination) registers and electronic healthcare records (EHR) from primary and secondary care. In these types of cohort studies, secondary data is used from pre-existing databases, and the study protocols cannot affect the way the data is primarily collected, but only which data should be used and how to analyze the available data. There may also be prospective cohort studies with primary data collection.

Pros

- Intuitive design, easier to communicate results
- Can be conducted with secondary data (if of sufficient quality)

Cons

- Large operational sample (not necessarily a con when using secondary data)
- Data on exposure, outcomes and potential covariates needed for the entire source population
- May be subject to healthcare-seeking bias
- Secondary data from pre-existing databases may be incomplete (e.g., in the case of influenza, lack of systematic laboratory confirmation)

2.2 Case-control studies

These designs compare vaccine exposure in cases to vaccine exposure in separate but ‘comparable’ controls. The big advantage of case control studies over cohort methods is the

reduction in operational sample size, though choosing suitable controls can be tricky and may introduce a selection bias. In case-control designs, researchers identify subjects who experienced the outcome of interest (cases) and a comparison group of subjects who did not experience the outcome of interest (controls).

Pros

- Reduction in operational sample size compared to cohort studies with only minor loss of precision of estimates.

Cons

- Choice of adequate controls may be difficult, as the vaccine coverage in controls should be the same as in the population that gave rise to the cases
- Selection bias may occur in the absence of sampling protocols

2.3 Variants of the case-control design

There are several variants to the classical case-control design. These designs are mainly different with respect to the way controls are selected.

2.3.1 Test-negative method

The test-negative design is the most frequently used design to estimate IVE. The design controls for selection bias due to healthcare-seeking behavior by restricting the source population to patients who seek medical care for a respiratory illness. Participants are selected among individuals who seek care for disease syndrome likely associated with influenza, such as acute respiratory illness (ARI), influenza-like illness (ILI) or severe acute respiratory infection (SARI), who are subsequently subjected to confirmatory testing [5, 6]. The cases are then chosen among the test positives and the controls among the test negatives. The method may be useful to minimize selection and ascertainment biases due to differences in (parental) attitude when seeking medical care and to physician differences in making decisions regarding laboratory diagnosis.
Within Europe, the test-negative case-control design has been frequently used by the I-MOVE network and the ECDC [7, 8] and is also used for the assessment of COVID-19 VE.

Pros

- Reduction in healthcare-seeking bias

- Reduction in ascertainment bias when sampling protocols are used and also due to information collection on vaccination status prior to knowing the laboratory influenza test results
- Reduction in outcome misclassification
- Can be integrated in the existing influenza surveillance system
- Reduction in operational sample size to select and include the controls
- Can produce estimates comparable with those of case-control and cohort studies in the presence of a highly specific diagnostic test [9]

Cons

- Less suitable for secondary data unless sampling protocols are in place in routine care (TND nested in a clinical cohort)
- Inappropriate control groups or screening methods could lead to biased results
- Less intuitive to interpret the results

2.3.2 Nested case-control studies

Nested case-control studies are case-control studies undertaken within cohort studies. Each incident case is matched with a number of controls sampled from the risk set for that case. The risk set usually comprises individuals who have not experienced the outcome event at the time of occurrence of the case (incidence density sampling). Nested case-control studies potentially offer great reductions in costs and data collection and analysis efforts compared with the full cohort approach, with relatively minor loss in statistical efficiency [10].

2.3.3 Case-cohort studies

A case-cohort design is similar to a nested case-control design with cumulative sampling. Whereas in the nested case-control design, the controls are randomly sampled from the non-cases, that is, those who did not get infected during the surveillance period, in the case-cohort design, controls are randomly sampled from the whole cohort, regardless of their disease status (case-base sampling). This has the advantage that in the case-cohort design the ‘rare-disease assumption’ does not need to be made. Controls may include both cases and non-cases. The case-cohort design was proposed by Prentice [11] to reduce the burden of data collection on covariates. However, as a result of the possible overlap between cases and controls, the statistical analysis of case-cohort data is much more complicated than the analysis of nested case-control data.

Pros

- Rare-disease assumption not needed

Cons

- Requires complex statistical analysis

2.3.4 Case-coverage method with external coverage cohort (screening method)

The case-coverage or screening method uses data on the exposure prevalence in cases and in the coverage cohort, from which the cases originate [12]. The unadjusted VE is obtained as

$$\widehat{VE}_{SCREEN} = 1 - \widehat{OR}_{SCREEN} = 1 - \frac{\widehat{p_d}/(1-\widehat{p_d})}{\hat{X}/(1-\hat{X})},$$

with the odds ratio $\widehat{OR}_{SCREEN}$ derived from the vaccine exposure prevalence among the cases ($\widehat{p_d}$) and the, often externally derived, estimate of the vaccine coverage in the coverage cohort ($\hat{X}$).

Both estimates $\widehat{p_d}$ and $\hat{X}$ are often available from routine surveillance, making the case-coverage method an inexpensive and ready-to-use method that might be useful to provide early effectiveness estimates or to monitor changes in effectiveness over time. Control for confounding is possible using stratified analysis, provided that the confounders are similarly measured for the cases and the coverage cohort. The method does not allow for uncertainty in the expected odds of exposure in the coverage cohort. This is immaterial when the coverage cohort is large, the major issue being the possible bias if cases are drawn from a population with a different vaccination profile from that of the coverage cohort. The method has been used to monitor IVE among older people in Germany [13].

Pros

- Inexpensive and ready-to-use

Cons

- If multiple vaccines are used, it is not possible to determine IVE per brand/type, but only overall (unless brand-specific coverage data is available)
- Bias may be introduced if cases are drawn from a population with a different vaccination profile from that of the coverage cohort

- Difficult to correct for confounding (as this would require knowledge of vaccine coverage in subgroups with the confounding factors of interest)

2.4 Recommendations

For studies using primary data to monitor IVE, we suggest to use the test-negative case-control design, with an appropriate choice of control group and implementation of sampling protocols [14].

For studies using secondary data we suggest to use the cohort design.

3 Exposure

In general, the exposure of interest is vaccination against influenza in the season under study (index season). The level of detail of the exposure definition depends on the study design and objectives.

3.1 Influenza vaccines

The recommendation on the composition of the seasonal influenza vaccines, i.e. the selection of the strains, is reviewed twice a year by the WHO, one for the Northern Hemisphere and another formulation for the Southern Hemisphere [15], to adapt to changes in the virus' epidemiology. Influenza vaccines can be characterized by the influenza number of virus strains included, the vaccine type, the presence or absence of an adjuvant, the administration route, the dose and the production process [16] (Table 3.1).

Table 3.1. Characteristics of seasonal influenza vaccines available in Europe in 2021/22

Characteristic	Options seasonal influenza vaccines available in Europe
Valency	Trivalent (two influenza A strains and one influenza B strain) Quadrivalent (two influenza A strains and two influenza B strains)
Antigen preparation	Inactivated/split Inactivated/subunit Live attenuated
Adjuvant	Without adjuvant With adjuvant
Administration route	Intramuscular Intranasal (live attenuated vaccines only)
Antigen dose	Standard High
Production base	Egg-based Cell-based Recombinant

3.2 Influenza vaccination and its special features

An influenza vaccination is the event that indicates the administration of an influenza vaccine. The person receiving the vaccination is thereafter considered vaccinated. The WHO recommends that vaccine-naïve children younger than nine years of age receive two doses during the season when they first receive an influenza vaccine [2]. The full schedule consists of a single dose for all others. However, the immune response may need up to two weeks to fully develop; therefore, a person is typically considered immunized only at day 14 post-vaccination (after the last dose). The immune response does wane with time [2]. Thus, one

might want to take into account the time that has passed since vaccination when studying its effect.

Influenza vaccination is recommended on an annual basis, especially to those at high risk [17]. As a consequence, people can be repeatedly vaccinated against influenza over several seasons.

3.3 Data to be collected

Resulting from 3.2, the vaccine brand(s) or type(s) (for studies on brand- or type-specific IVE) and vaccination date(s) of all influenza vaccinations given during the index season should be collected to fully capture a person's vaccination history. Knowledge of the vaccine brand allows the deduction of the administered vaccine's characteristics (see Table 3.1) and subsequently the estimation of brand-specific effects. In a small number of cases, when there is only one brand per type, vaccine brand and characteristics can be inferred from the type. The exposure data should include the vaccination date because it is important to know the timing of the vaccination within the season and in relation to the outcome. Influenza vaccination status in previous seasons is highly collinear with influenza vaccination status in the present season and is therefore not recommended to be included as a confounder, although it can be considered as a potential effect modifier. However, depending on the available data sources and the study design, the data requirements can be adapted and reduced.

3.4 Data sources

3.4.1 Administrative sources

Computerized or paper-based vaccination registers (see also D2.4 [18]) originating from routinely recorded medical records might be the gold standard among the administrative sources providing individual-level data including the administered product and the vaccination date for all people covered by the register.

Aggregated data, e.g. crude numbers of vaccine doses delivered in the target population, might also suffice for vaccination coverage estimation in a reference group of studies that utilize the screening method. However, unless vaccination coverage is brand-specific, brand-specific IVE cannot be calculated using the screening method in settings where multiple brands are used.

Only in rare cases can the structure of the market help to retrieve the vaccination brand of all those who received influenza vaccination in a specific area, such as countries or regions where only one brand is procured through tenders (see also D3.1 [19], D3.3 [20] and Stuurman et al. [21]).

3.4.2 Self-reporting and contacting healthcare providers

The initial information on vaccination can also be provided by the vaccinee. Self-reporting alone leads most likely to a dichotomous vaccination status without any details concerning the administered product or the exact vaccination date. Such data might suffice in studies that do not focus on brand-specific effects and only require the chronological order of exposure and outcome to be known.

Vaccination cards held by the vaccinee might improve the information available to be collected through surveys and self-reporting, as could contacting the pharmacy or physician that provided the vaccine [2, 22].

3.5 Common exposure definitions

Common exposure definitions frequently used by the ECDC [23, 24], the I-MOVE/I-MOVE+ network [25, 26], and elsewhere are summarized hereafter.

The above protocols agree that an individual is considered ‘vaccinated’ starting from >14 days after the last vaccination [23-26]. The first 14 days from vaccination are either considered ‘unvaccinated’ (leading to two distinct exposure levels) [23-25], or they are considered ‘partially vaccinated’ (leading to three distinct exposure levels). In DRIVE, the latter was used. In the first two seasons, ‘partially vaccinated’ individuals were considered in sensitivity analyses; from 2019/20 onwards, they were excluded.

Depending on the study design, an individual’s exposure status might vary over time in follow-up and, as such, may be considered as a time-dependent variable (cohort studies) [23] or assessed only once for the time point of the outcome’s occurrence in case-control studies and its variants [24].

Depending on whether or not they received influenza vaccination in a prior season, young children may require two doses of influenza vaccine. To operationalize this in the absence of a lifelong influenza vaccination history, in DRIVE (from 2020/21 onwards), children <9 years for whom the record showed that they received two doses in the season were only considered vaccinated 14 days after the last dose. In case the records showed that a child <9 years only received one dose, it was assumed that the child was vaccinated.

3.6 Exposure misclassification

Whichever data source is available or has been chosen, misclassification can occur. The consequences of exposure misclassification in terms of information bias are discussed in Chapter 5.

3.6.1 Adequacy of data source

When using administrative sources that record the presence of vaccination but not a potential absence, it must be ensured that they are complete and cover the entire population of interest (i.e. the population from which cases and non-cases are retrieved). Otherwise, the part of the population not or incompletely reflected in the data is by default assumed to be not vaccinated. In case of aggregated vaccination information (when applying the screening method), it must be further guaranteed that this information originates only from the population of interest to avoid misclassification towards a wrongly increased number of vaccinated subjects.

3.6.2 Data entry errors

As with any records, vaccination registers based on routinely entered medical records may contain data entry errors and consequently misclassification. Inaccurate records concerning the vaccinated subject, the administered product, or the vaccination date can lead to wrongly classifying study subjects as either not vaccinated or vaccinated with a certain vaccine during the study period.

Incomplete or inaccurate data entry may result in misclassification. As the vaccination status in computerized vaccination registers is assumed to be “not vaccinated” and a single typo can break the automated process of correctly identifying the vaccination event (vaccinated subject, administered product, vaccination date) and cause misclassification, it seems more likely to classify the vaccinated as not vaccinated (rather than equally classifying the vaccinated as non-vaccinated and vice versa)[27].

3.6.3 Recall bias

Last but not least, self-reporting carries the risk of misclassification. The accuracy or completeness of recalling a subject’s vaccination history can vary e.g. depending on the study subject’s age and time since vaccination and might be further influenced by the outcome status, especially if the subject or researchers conducting the interviews are not blinded, or if the outcome is related to compensations or other benefits. In the context of brand information,

it may also be not known by the vaccinee. Complementing or verifying self-reported information through vaccination cards, contacting healthcare providers, or administrative sources reduces this risk.

3.7 Recommendations

The following data on exposure should be collected, both for studies collecting primary data and those using secondary data:

- Vaccine brand(s) or type (s) of all influenza vaccinations given during the index season (i.e. the season for which IVE is being estimated); for studies on brand- or type-specific IVE
- Vaccination date(s) of all influenza vaccinations given during the index season, or if not available, the sequential order and relative timings of exposure and outcome
- How the vaccination status was ascertained and whether it was confirmed e.g. through medical records

4 Outcome

IVE usually refers to effectiveness in preventing influenza infection. Hence, in IVE studies, the primary outcome of interest is infection with an influenza virus, which can only be accurately confirmed by a laboratory test. Occasionally other outcomes than laboratory-confirmed influenza infection are used. Other outcomes are employed for two reasons: the outcome may be life-threatening and therefore worth preventing, or the outcome is used as a surrogate endpoint for influenza infection.

4.1 Common outcome definitions

Outcome definitions can be specific or non-specific.

The specific outcomes in IVE studies require laboratory-confirmed influenza infection [2, 28]. According to the EU influenza case definition, a laboratory confirmation requires the 'isolation of influenza virus from a clinical specimen', the 'detection of influenza virus nucleic acid in a clinical specimen', the 'identification of influenza virus antigen by DFA test in a clinical specimen', or an 'influenza-specific antibody response' [29]. The currently available laboratory tests used to determine whether or not influenza infection has occurred are described in detail in Chapter 6.

An isolated strain can match with one of the strains included in the vaccine or not. If only infections with matching strains are considered, the IVE being estimated is that against matching strains; if all infections are considered, matching and non-matching, the IVE being estimated is that against any circulating strain. IVE frequently varies between strains; therefore, IVE estimates are often reported by influenza A subtype or by influenza B lineage [30].

Laboratory-confirmed influenza infections in different settings, reflecting increasing levels of severity, may be considered, such as influenza in primary care, influenza requiring hospitalization, and admission to an intensive care unit (ICU) following an influenza infection. Testing for influenza is usually only done when a subject shows signs and symptoms of influenza, frequently used clinical case definitions are listed in Table 4.1. This means that non-symptomatic cases go undetected.

Another non-specific outcome in observational studies has been death or all-cause mortality [2]. The problem with non-specific outcomes is the interpretation of the IVE. With a specific outcome, the maximum IVE is 100%, whereas with a non-specific outcome, the maximal IVE is less than 100% and usually unknown [28]. Nevertheless, even a low IVE can have a large

public health impact if the outcome of interest is serious or very frequent. Studies of IVE against non-specific outcomes must meet several demanding quality criteria, including a large sample size to achieve a precise estimate that can be clearly distinguished from the null and elaborate adjustment for confounders, which has been demonstrated to be an issue in several studies [2].

Table 4.1. Clinical case definitions of ILI and SARI by WHO and EU

Outcome	Institution	Case definition / clinical criteria
ILI	WHO	• 'an acute respiratory illness with a measured temperature of ≥38°C and cough, with onset within the past 10 days' [31]
	EU	• a 'sudden onset of symptoms' and • 'at least one of the following four systemic symptoms: fever or feverishness, malaise, headache, or myalgia' and • 'at least one of the following three respiratory symptoms: cough, sore throat, or shortness of breath' [29]
SARI	WHO	• 'an acute respiratory illness with a history of fever or measured fever of ≥38°C and cough, with onset within the past 10 days, requiring hospitalization' [31].
	EU	• a 'sudden onset of symptoms' and • 'at least one of the following four respiratory symptoms: cough, sore throat, shortness of breath, or coryza' and • 'a clinician's judgement that the illness is due to an infection' [29].

The influenza disease outcome has also been defined for surveillance purposes using diagnostic codes such as ICD-9, ICD-10, ICPC-2, or other codes such as Read Codes in the UK, which are utilized in many patient information systems to classify a patient's diseases, disorders, injuries, and other health-related conditions. Accordingly, a hospitalization or a medical encounter would be considered to be due to influenza e.g., in presence of the ICD-10 codes J09, J10, J11, and all their subclasses [32].

For case-control and test-negative design (TND) studies, EMA guidance states laboratory-confirmed, medically attended influenza should be the primary outcome; based on the study setting (general population or hospital), secondary outcomes of interest may include pneumonia, influenza-related hospitalizations (influenza related or associated with respiratory or cardiac disease), or death [3]. For cohort studies, EMA guidance states outcome of interests may include medically-attended respiratory infection (MAARI), medically-attended ILI, all cause deaths, respiratory deaths, hospitalizations for pneumonia and influenza, hospitalizations for all respiratory conditions, laboratory-confirmed cases of MAARI/hospitalized pneumonia, and influenza and ICU admissions [3].

Non-specific outcomes should be employed when the outcome is serious or life-threatening and worth preventing. The endpoint should always be clearly described and non-specific outcomes should not be used as a surrogate endpoint for influenza because in that case the IVE against infection will be underestimated [28].

4.2 Data to be collected and data sources

4.2.1 Challenges

In practice, it is challenging to find a specific outcome definition that covers the whole disease burden due to influenza. The severity of the symptoms can vary a lot, ranging from mild illness treated solely at home to very severe illness treated in hospitals' intensive care units and possibly leading to death. Further, influenza can make subjects more susceptible to secondary bacterial pneumonia and acute myocardial infarction or exacerbate chronic diseases. Then again, patients infected with other respiratory pathogens also circulating during the influenza season might present similar clinical syndromes [2].

4.2.2 Primary data to be collected

Consequently, aligning a non-specific clinical outcome definition like ILI or SARI with a laboratory test result is strongly recommended. The following data should be collected to optimally describe the outcome test-positive influenza, medically attended ILI or SARI across all study designs:

- the symptoms forming the clinical syndrome of ILI or SARI, including the information whether hospitalization or intensive care treatment was required,
- the date of symptom onset,
- the date the respiratory specimen was taken,
- and the detected influenza type and subtype (if possible).

Such data could either arise from an active collection as part of a study designed to estimate IVE (e.g. TND or prospective cohort study); data sources for this could include medical records kept in hospitals or by general practitioners (GPs) or from surveillance/sentinel systems.

4.2.3 Secondary data collection

Nevertheless, this recommendation does not intend to generally exclude all other data collections from IVE estimations, especially from cohort studies through which less-specific

outcomes may be explored, although they might not provide as sensitive or specific information as indicated above. On the one hand, statutory infectious diseases registers covering clinically- and laboratory-confirmed influenza cases might provide outcome data that is less sensitive than that obtained through primary data collection (as not all potential cases will be routinely tested) but still highly specific. On the other hand, diagnostic codes or data on antiviral drug prescriptions can be considered for deriving non-specific proxies for the outcome [22, 32]. Accordingly, information on the diagnostic code or antiviral drug describing the disease entity and the date of diagnosis or prescription approximating the disease onset should be collected. However, the resulting IVE estimates must be interpreted carefully (see 4.3 and 5.2.17). Additionally, outcomes that are not a proxy for laboratory-confirmed influenza (e.g. all-cause pneumonia) may be considered.

4.2.4 Recurrent infections

A second influenza infection in the same subject in the same season is rare but might happen. In IVE studies, only the first infection is counted. The broader the outcome, the more frequently subjects might qualify again as cases in the same study. Especially when diagnostic codes or prescriptions are used to determine the outcome, an additional definition for a disease episode is required to distinguish frequently repeated events that are likely to belong to the same infection and sparse events that represent separate infections [6, 7]. The data collection should comprise all incident cases. While a physician involved in the collection of primary data might be able to differentiate between an incident and a prevalent case, the timely distinction between these two is difficult when using secondary data, as it strongly depends on the underlying coding and recording practices.

4.3 Outcome misclassification

Whichever outcome definition and data source are chosen, misclassification can occur. The consequences of outcome misclassification in terms of information bias are discussed in Chapter 5.

4.3.1 Imperfect laboratory result

Imperfect laboratory test

Most of the laboratory tests currently available for detecting influenza infections are highly sensitive and highly specific but not perfect (see 6.1 and 6.2). Imperfect sensitivity bears the

risk of misclassifying infected subjects as non-cases, but this will not lead to bias, unless the sensitivity of the test differs between vaccinated and unvaccinated subjects. Imperfect specificity bears the risk of wrongly classifying non-cases to be influenza-positive and may lead to serious underestimation of the vaccine effectiveness.

Lag time between symptom onset and specimen

Typically, influenza virus shedding starts a day before symptom onset and decreases substantially four days after symptom onset [33]. If a respiratory sample is taken thereafter, the virus may no longer be detectable, which introduces misclassification. Shedding peaks on the first 1-3 days of clinical illness [34-36]. Two studies on naturally acquired infection found the viral load of influenza B to be high for a longer time than for influenza A [35, 36].

Younger age has been associated with increased viral shedding (<16 vs. ≥16 years [37]; 0-5 vs. 6-15 vs. 16-64 years [38]), although this effect has not been consistently demonstrated across studies [36, 39]. No effect of vaccination on viral shedding among influenza cases was found [36, 37]. A systematic review on influenza A A(H1N1)pdm09 virus shedding found that the duration of shedding increased with increasing disease severity and decreased by timely antiviral treatment [39].

Swabs

Preferred swabbing sampling for influenza detection are nasopharyngeal aspirates and washed, followed by nasopharyngeal swabs and mid-turbinate swabs; less preferred are throat swabs [40].

Use of antivirals

Antivirals like the neuraminidase inhibitors oseltamivir and zanamivir and the M2 inhibitors amantadine and rimantadine can be used to prevent influenza (e.g. as prophylaxis for individuals exposed to influenza virus when admitted to hospital) and for the treatment of influenza in order to mitigate the associated complications [41].

The use of antivirals can reduce the duration of viral shedding [39]; consequently, respiratory samples taken after the start of antiviral treatment may already be negative, leading to misclassification.

4.3.2 Healthcare-seeking behavior

Influenza diagnosis can only be confirmed after a healthcare visit, and influenza infection cannot be ascertained in persons who do not seek care, yet the latter are assumed to not have

been infected in cohort and classic case-control studies [6]. For this reason, it is important to specify that the outcome is medically attended influenza infection.

4.3.3 Adequacy of data source

If routinely recorded data on laboratory-confirmed influenza cases, diagnostic codes or prescriptions are collected, their use in cohort studies estimating IVE must be reviewed carefully. In analogy to exposure data originating from administrative sources, it must be ensured that the records cover the entire population of interest because study subjects without an entry related to the outcome of interest are considered disease-free. However, if the chosen data source's sensitivity to detect the outcome is imperfect and influenced by a study subject's vaccination status (e.g. because of differences in healthcare-seeking behavior or differences in the physicians' testing and diagnostic preferences), differential outcome misclassification occurs. Consequently, the proportion of diseased people misclassified as non-cases might differ between the different exposure groups.

4.3.4 Data entry errors

Healthcare databases based on routinely entered medical records are prone to data entry errors and, consequently, non-differential misclassification. Differential misclassification linked to the patient's risk status can also occur, as it cannot be excluded that the content of the information entered and the scrutiny to encode this information is the same for an otherwise healthy subject compared to a patient with underlying medical conditions.

4.4 Recommendations

In studies collecting primary data, the recommended outcome is laboratory-confirmed, medically attended influenza. We suggest the collection of the following data:

- Symptoms forming the clinical syndrome of ILI or SARI, including whether hospitalization or intensive care treatment was required,
- Date of symptom onset,
- Date the respiratory specimen was taken,
- Laboratory confirmation yes/no, and if yes, influenza type and preferably also subtype/ lineage

In studies utilizing secondary data, e.g. from existing healthcare databases, the recommended outcome is laboratory-confirmed influenza, overall or stratified by clinical condition. However, this recommendation does not exclude the use of syndromic, code-based or non-specific outcome definitions discussed in 4.1, either in association with a positive influenza test or alone. We suggest the collection of the following data points:

- Clinical condition (if applicable),
- Date the respiratory specimen was taken,
- Detected influenza type and preferably also subtype/ lineage for laboratory-confirmed influenza

5 Potential biases and confounders expected across different settings

This chapter describes biases and confounders that are important to IVE studies. A systematic literature review on confounding and bias has been performed as part of WP2 [42]. A re-analysis of the DRIVE data from 2018/19 and 2019/20 to study the impact of specific confounders is in progress.

5.1 Definitions

A bias is a systematic error that leads to an incorrect effect estimate of the exposure on the outcome. Examples are selection bias and confounding.

A confounder is a variable that influences both the exposure and the outcome. Confounding can be subdivided into positive confounding, which leads to bias away from the null hypothesis (higher VE estimate), and negative confounding, which leads to bias toward the null hypothesis (lower VE estimate).

An effect modifier is a variable that differentially (positively or negatively) modifies the observed effect of the exposure on the outcome. Different groups have different risk estimates when effect modification is present [43].

It is noted that certain factors may both act as a confounder and effect modifier (see 5.3.3).

5.2 Biases and confounders

Multiple biases and confounders may be present. In TND studies, the confounders most frequently adjusted for are age, sex, calendar time and chronic conditions [44].

5.2.1 Health status

Persons who receive influenza vaccines may differ in important ways from those who do not receive them, such as in the presence of chronic underlying conditions (e.g. chronic obstructive pulmonary disease, cardiovascular diseases, metabolic disorders, renal diseases, treatment-induced immunosuppression and disease-induced immunosuppression) and frailty.

Here we differentiate between confounding by indication and the healthy vaccinee bias (which includes confounding by contraindication).

5.2.1.1 Confounding by indication (overall)

Confounding by indication occurs when a symptom or sign of disease is judged as an indication

for a given vaccine and is therefore associated with both the vaccine and a higher probability of an outcome related to the disease for which the vaccine is indicated or a specific brand is given [45]. In the case of influenza vaccination, patients who have underlying chronic conditions are more likely to be vaccinated as they are at higher risk for (severe) influenza disease, and risk groups are part of the vaccine recommendations; this can lead to a lower VE estimate [46]. Underlying conditions of interest include chronic obstructive pulmonary disease (COPD), cardiovascular diseases, metabolic disorders, renal disease, treatment-induced immunosuppression and disease-induced immunosuppression [47].

5.2.1.2 Healthy vaccinee bias and frailty bias

The healthy vaccinee bias and the frailty bias have the opposite effect of confounding by indication, i.e., vaccinated individuals may be healthier than unvaccinated individuals, leading to a higher VE.

Healthy vaccinee bias

Risk groups are part of the vaccine recommendations and are more prone to receive vaccination compared to healthy subjects (see Section 5.2.1.1). However, within these groups recommended for vaccination (e.g., those based on age), those with a healthier lifestyle may be more likely to accept influenza vaccination.

Frailty bias

Vaccine coverage has been found to be low in frailest patients, i.e. those with a low functional status and a high predicted probability of death during the upcoming influenza season [48, 49]. Healthcare providers may be reluctant to vaccinate such patients and likewise patients may “give up on preventive measures” [48]. Consequently, there may be relatively fewer severely ill patients in the exposed group [47, 49].

5.2.1.3 Confounding by indication (type- or brand-specific)

Brand-specific IVE studies from settings in multiple countries will be pooled. Vaccine type recommendations may differ between countries. For example, country 1 may use the non-adjuvanted trivalent vaccine for all adults, whereas country 2 may prescribe the adjuvanted trivalent vaccine to a specific risk group and non-adjuvanted to the rest of the population. Furthermore, a healthcare worker may decide to vaccinate with a specific brand (e.g., an adjuvanted vaccine, an intradermal vaccine, a quadrivalent vaccine) because he/she thinks a particular patient requires additional protection. This may result in potentially important differences between the exposed groups that affect the pooled VE estimate. Knowledge of

vaccine recommendations or prescription practices across settings is necessary (also refer to WP2.2 and WP3.1). Furthermore, certain clinical conditions associated with the outcome may also be associated with the receipt of specific influenza vaccine types or brands and may cause confounding by indication, as differences between VE estimates may be partially due to underlying population characteristics rather than to true differences between vaccine types or brands [50-53]. This is in part a consequence of influenza recommendations that tailor vaccine-type recommendations for specific risk groups, particularly based on age [54]. In the DRIVE 2020/21 season, this was the case in Austria, Finland, Italy, Spain and the UK. In addition, a GP survey conducted in DRIVE concluded that whilst a GP's choice of vaccine type for an individual patient is primarily driven by vaccine type availability, if GPs had a choice, older patients and patients with multiple comorbidities were more likely to be prescribed the adjuvanted vaccine [53].

5.2.2 Selection bias

Selection bias occurs when subjects are differentially enrolled in the study or analysis, or data is differentially collected based on their exposure or outcome status.

5.2.2.1 Selection bias based on vaccination status

In IVE studies, this can happen when not all subjects with acute respiratory illness are equally likely to be tested for influenza, especially in settings where testing takes place at the clinician's discretion. For example, vaccinated subjects may less frequently be tested for influenza (due to a perceived lower chance of influenza infection), which would lead to a higher VE estimate.

5.2.2.2 Selection bias based on outcome

Alternatively, the exact presentation of acute respiratory illness may – in the absence of clear sampling protocols - influence the clinician's decision to order influenza testing, which can either lower or increase the VE estimate [2].

Furthermore, in prospective studies with active enrolment that requires informed consent (i.e., those taking place outside the context of routine surveillance or medical practice), individuals who are too ill to give consent (e.g., too acutely ill or because of worsened chronic conditions) may not be enrolled, thereby biasing the outcomes captured to less severe disease.

In addition, subjects that do not fulfil the ILI/SARI definitions are excluded from TND studies, thus excluding potential cases, such as e.g., those (elderly people) with clear systemic symptoms and even radiologically confirmed pneumonia but without any signs of respiratory symptoms according to the ECDC ILI definition.

5.2.3 Healthcare-seeking behavior

Not all subjects are equally likely to seek care and hence to be diagnosed. Healthcare-seeking behavior is usually associated with exposure and the outcome (e.g., vaccinated persons may be more likely to seek medical care when diseased), and it may therefore lead to confounding. Multiple factors may influence healthcare-seeking behavior, ranging from funding of the healthcare system to personal factors and disease severity.

Disease severity may be affected by exposure status (e.g., vaccinated persons may have less severe disease, reducing their likelihood to seek care) and acute respiratory infection (ARI) etiology. If ARI severity differs by etiology, this might differentially impact the propensity to seek healthcare between cases and controls in a test-negative study [5].

Although healthcare-seeking bias will be present in each setting, the bias is likely stronger in a GP setting than a hospital setting, as patients with severe disease are more uniformly likely to seek care.

5.2.4 Information bias

Information bias arises when incorrect information about a variable is collected [22]. The most important types of information bias are exposure misclassification and outcome misclassification, although information bias is also applicable to other variables, such as confounders and covariates.

Exposure or outcome misclassification can be differential or non-differential. Misclassification is non-differential when the bias is the same for all subjects, regardless of exposure or outcome status. Non-differential misclassification usually causes a bias towards the null, leading to a lower VE. Misclassification is differential when the bias differs between exposed and non-exposed subjects or between subjects with and without the outcome. Differential misclassification can lead to bias towards or away from the null. Misclassification bias in vaccine effectiveness studies is discussed in detail by De Smedt et al. [55].

5.2.4.1 Misclassification of exposure

As described in more detail in Section 3.3, exposure misclassification can occur due to an inadequate data source, data entry errors and recall bias.

Bias in the exposure due to inadequate data source or data entry errors could be non-differential or, in rare cases, differential. Thus, in most cases, it is expected to lower the VE. Data entry errors are most likely to result in subjects erroneously classified as unvaccinated, leading to a lower VE.

Recall bias can be differential or non-differential, therefore, bias can also go either way here.

5.2.4.2 Misclassification of outcome

As described in more detail in Section 4.3, outcome misclassification can occur due to imperfect laboratory tests, lag time between symptom onset and specimen collection, lack of healthcare seeking (in cohort studies), use of antivirals, an inadequate data source and data entry errors.

Outcome misclassification due to imperfect diagnostic tests or laboratory procedures, such as lag time between symptom onset and specimen collection, use of antivirals and data entry errors, is most likely to be non-differential and thus underestimating VE. This can be overcome by using more reliable test or harmonizing the sample collection and laboratory testing procedures. Another alternative is to apply bias correction methods for TND studies that have been developed to address potential misclassification bias due to imperfect tests [56].

Bias in the outcome due to an inadequate data source can be non-differential or differential. In cohort studies conducted in the absence of a protocol or other measures ensuring a balanced case detection rate among the exposed and unexposed, differential outcome misclassification bears the risk of either under- or overestimating VE.

5.2.5 Age

Age is associated with both the exposure (some risk groups recommended for vaccination are defined by age [57]) and the outcome (the very young are more likely to be infected due to lack of immunity, and very young and elderly people are more likely to suffer severe influenza infection or to develop complications [58]). It is recommended that age is treated as an effect modifier (see Section 5.3.1); however, within each stratum, age can be treated as a confounder.

5.2.6 Gender/sex

Men and women may have different healthcare-seeking behaviors, resulting in differences in influenza vaccine uptake and likelihood of medical consultation for influenza disease.

For example, a study conducted in Spain found that, within the influenza vaccine recommended risk group, vaccine coverage was lower among women than among men [59].

Alternatively, in countries where healthcare workers are offered near compulsory influenza vaccinations or where pregnancy is considered an indication, women may have higher coverage.

5.2.7 Pregnancy

In many countries, pregnant women are recommended influenza vaccination. Furthermore, changes to the immune system, heart, and lungs during pregnancy make pregnant women more susceptible to severe disease [60].

5.2.8 Smoking behavior (or parental smoking behavior)

Smoking may be positively associated with vaccination (e.g. smokers may be more likely to be affected by chronic diseases and, therefore, belong to the risk groups particularly eligible for influenza vaccination) or negatively associated with vaccination (e.g. smokers may be more prone to ignore health-related recommendations). Furthermore, smoking may be associated with the outcome (e.g. disease may be more severe).

For children, parental smoking status may be considered. Wilson et al. showed that children hospitalized for influenza had more severe disease if they had been exposed to second-hand smoking [61].

5.2.9 Socioeconomic status or applicable proxy

The socioeconomic status is a relevant variable to indicate access to health services. It may be associated with exposure and access to influenza vaccination in countries where influenza vaccination is not free, and with medically attended outcomes in countries where healthcare is not free, or where insurance-based systems run in parallel to public care.

5.2.10 Prior exposure to influenza vaccination and influenza infection

An important confounder is the prior exposure history of a patient, which includes both prior infection and prior vaccination status.

5.2.10.1 Prior influenza vaccination

Prior influenza vaccination may be a confounder of IVE when influenza vaccination in the current season is associated with vaccination history and when vaccination modifies the risk of natural infection in the following season [62]. Studies have found both positive and negative interference of repeated influenza vaccination on IVE [63], and this may differ by season.

The antigenic distance hypothesis proposes that negative interference may occur if the consecutive vaccines are antigenically similar, and antibodies produced in the past season may neutralize vaccine antigens of the subsequent year’s vaccine before it can trigger a full immune response, especially if the new circulating strain is antigenically different [64].

Conversely, positive interference may occur when the antigenic distance between the first vaccine and the new circulating strain is small, and pre-existing antibodies are boosted by the response to the vaccine, helping to clear the new virus [64]. Furthermore, prior vaccination may be protective because of persisting vaccination immunity, or it may modify the risk of natural infection because of a lower previous risk of natural infection [62]. IVE may be influenced by vaccination patterns over at least several seasons [65] and lower against A(H3N2) virus [66] [67].

Prior influenza vaccination is frequently predictive of influenza vaccination in the current season; therefore, collinearity may be a problem when including prior influenza vaccination in the statistical model as a confounder [68].

5.2.10.2 Prior influenza infection

Prior influenza infection can influence the choice to receive the influenza vaccine in the current year and lead to a degree of existing immunity against influenza. Exposure to the virus is believed to induce lifelong cellular and humoral immunity that not only protects against infection by the original infective strain but may also provide cross protection against antigenically similar strains. This can lower the VE estimate in the current year. For example, Saito et al. found a profound protective effect of medically attended influenza A infection in the prior season [69].

In the scientific literature, the prior influenza infection status is rarely considered because of the difficulty to collect accurate information. Furthermore, it is unknown for how many previous years this information should ideally be collected.

Little research has been done on the interplay between prior influenza infection and prior vaccination and its effect on IVE [69].

5.2.11 Concomitant administration of COVID-19, pneumococcal, Herpes zoster and childhood vaccines

The likelihood of receiving an influenza vaccination is higher among recipients of other vaccines because of behavioral aspects and the patients'/parents' opinion concerning vaccines and access to healthcare. The COVID-19 may have adjusted the behaviors and attitudes toward vaccination [70], but accounting for this component is important as this can lead to confounding, particularly in TND studies where patients with non-influenza ILI/SARI are included as part of the test-negative control group. This is particularly of interest starting from the 2021-2022 influenza season with respect to the COVID-19 vaccination, as the inclusion of SARS-CoV-2 controls in influenza TND studies may lead to underestimated IVE estimates

[71].

Younger children that are recommended for influenza vaccination may receive their recommended childhood vaccinations potentially alongside the influenza vaccine; in some elderly cohorts, the influenza vaccine is co-administered in the same healthcare visit with the pneumococcal vaccine, because of overlapping recommendations for the two vaccines. Up-to-date pneumococcal vaccination can affect the occurrence of secondary bacterial infectious complications of influenza and is therefore an important confounder for non-specific outcomes; it would lead to a higher VE estimate. Additionally, it is important to distinguish between the 23-valent pneumococcal polysaccharide vaccine (PPS23) and the 13-valent pneumococcal conjugate vaccine (PCV13) because of their differences in effectiveness against pneumonia.

5.2.12 Statins

Recent studies suggest statins may impair the antibody response and thereby reduce vaccine-induced protection [72]. Studies have shown reduced immune response to influenza vaccine [73], reduced VE to medically attended acute respiratory infection [74], and reduced IVE to some (but not all) influenza types/subtypes [72]. On the other hand, statins have been suggested to have a protective effect against infections [75].

5.2.13 Residence in long-term care facility

Residence in a long-term care facility (LTCF) may be associated with vaccination (e.g. residents of long-term care facilities are often vaccinated to minimize influenza outbreaks in the facility) and with the outcome (e.g. disease may be more severe because of underlying conditions, and influenza attack rate may be higher than in the community).

In relatively closed communities such as LTCF, disentangling the effect of individual protection and herd immunity caused by both caretakers and visiting relatives being vaccinated may be challenging [2]. For such reasons, in DRIVE, residence at long-term care institutions was an exclusion criterium.

5.2.12 Perinatal conditions

In addition to the variables for health status described above, a health status indicator for children, based on birth weight and maturity at birth, as well as some other perinatal or congenital conditions (e.g. Apgar score or Down syndrome), is relevant for children. The cut-off age depends on the nature and severity of the condition.

5.2.14 Child's adherence to the local childhood vaccination program

In case that influenza vaccine is recommended for a child, either because the child falls within a risk group recommended for influenza vaccination or the country recommends influenza vaccine for all children, adherence to the local childhood vaccination program in general is likely positively associated with influenza vaccine reception and healthcare-seeking behavior.

5.2.15 Waning immunity

The intra-seasonal waning of protection against influenza vaccination has been demonstrated through lower VE and geometric mean titers with increasing time since vaccination [76, 77]. The waning of immunity may occur at different rates for different vaccine components [76, 78] and may be of particular concern in adults aged 65 years and above [79]. Recent work has explored bias in influenza vaccine waning studies [80, 81].

5.2.16 Infection pressure

Influenza infection pressure varies among people. Factors associated with high infection pressure include being a healthcare worker, army conscript, childcare worker, or institutionalized individual (e.g., LTCF, prisons), and for children, the number of siblings.
In certain cases, vaccinated and non-vaccinated subjects may have different contact patterns leading to a differential infection pressure [6]. For example, healthcare workers, who are frequently offered occupational influenza vaccine, may have increased exposure to the influenza virus through contact with patients seeking care for influenza infection, compared to the general population. In this example, differential infection pressure would lead to a lower VE estimate.

5.2.17 Climatic factors

Influenza transmission is affected by climatic factors. Cold and dry conditions have been found to favor influenza transmission [82-85].
Lowen et. al propose several mechanisms to explain the influence of humidity and temperature on influenza [85]. First, the host may be more susceptible to respiratory virus infections as a result of desiccated nasal mucosa due to breathing dry air. Second, viral stability has been found to be very high at low relative humidity. Third, at low relative humidity, respiratory droplets carrying the influenza virus are small (as water evaporates quickly), allowing them to remain airborne for a longer time, increasing the chance of transmission of the virus. Finally, viral shedding increases at low temperatures. This could be due to reduced viral clearance as

a result of a slower mucociliary activity or increased viral stability.

5.2.18 Virus characteristics

Factors pertaining to the virus are virus virulence and level of antigenic match between the vaccine strain and the circulating strain.
Vaccine effectiveness is lower in seasons with a mismatch [86, 87]. Early news of a mismatch with the vaccine could potentially influence the level of vaccine uptake in a population. Furthermore, intra-seasonal antigenic drift may cause variations in vaccine effectiveness within the season.

5.3 Effect modifiers

5.3.1 Age

Due to immunosenescence, which refers to the gradual decline of the immune system brought naturally by age, vaccination in older adults is expected to lead to a lower immune response than in younger adults [88]. Furthermore, older adults are usually prone to develop more severe influenza disease or complications thereof. Both of these factors lead to lower VE estimates in older age groups [2].
Stratification of the data into age groups is advised. Harmonizing the age categories used in the analysis across studies will be crucial for the pooled analysis.

5.3.2 Immunosuppressed/ immunocompromised patients

Patients who are immunosuppressed or immunocompromised, whether as a result of a disease or a treatment, are at risk of complicated influenza [89]. Although patients may still benefit from a degree of protection against severe disease, the VE in this population is lower than in patients who are not immunosuppressed/immunocompromised.

5.3.3 Others

In addition to age, confounders listed above may sometimes be considered effect modifiers, such as sex [90-92], prior vaccination/infection, statin use [93] or health status (frailty). Frailty can be associated with vaccine effectiveness, especially in older adults and in risk groups. McGrath et al. state that they could not control healthy-user bias through statistical adjustment and therefore performed sample registration, which reduced much of the bias [94].

5.4 Controlling for bias and confounding

The effect of bias and confounding can be controlled at the study design level or in the analysis.

Table 5.1 shows what information would be required to control in the analysis for each bias or confounder described above and how it can be controlled in the study design.

It will not always be feasible to address each type of bias or confounding. If data has not been collected, it is impossible to account for it in the analysis. For example, if the lag time between symptom onset and specimen collection has not been collected as part of routine clinical practice, this information will not be available from administrative databases and, consequently, it will be impossible to adjust for this type of outcome misclassification.

Some types of bias are best addressed through the study design, for example, by including sampling methods in TND studies' protocols to reduce selection bias. For many confounders, an adjustment in the analysis (including matching in TND studies or the use of propensity scores in cohort studies) is likely sufficient. Separate analyses may be required across levels of effect modifier.

Table 5.1. Controlling for bias and confounding in IVE studies

Type of bias/confounder/effect modifier	Direction of bias (↓(decrease) or ↑ (increase) VE)	Controlling using study design	Information required to control in the analysis	Comments
Health status	↓: confounding by indication ↑: healthy vaccine bias, frailty bias	• Comparing groups with similar prognosis (e.g., between a GP who does and one who doesn't vaccinate patients although this may be difficult in practice [46]; or across levels of immunosuppression or frailty status) • restricting or stratifying the study population (at the cost of reduced precision due to decreased sample size) [46], • individual matching of exposed and non-exposed into main prognostic strata or propensity-score matching (although this requires a large sample size) [46].	• Data on presence of chronic underlying conditions (like chronic pulmonary disease, cardiovascular disease, metabolic disorders, renal disease), treatment-induced immunosuppression and disease-induced immunosuppression, frailty • Number of hospitalization due to chronic conditions (as indicator for disease severity) • Data on presence of perinatal conditions (for studies in children only) • Measure of functional status/frailty (probably not captured in administrative databases)	
Selection bias	↓: less testing in vaccinated subjects ↓ or ↑: disease presentation	*Studies with primary data collection:* • Standardized case definitions and specimen collection criteria (same criteria used regardless of exposure status.) • Asking for consent from next of kin for patients too ill to give consent *Studies with secondary data collection:* • Looking for selection in sampling according to vaccination/background factors (if also negative samples are available) • Using administrative databases avoids missing severely ill patients due to lack of consent		

Type of bias/confounder/effect modifier	**Direction of bias (**↓(decrease) or ↑ (increase) VE**)**	**Controlling using study design**	**Information required to control in the analysis**	**Comments**
Healthcare-seeking bias	↓ or ↑	• Using TND minimizes this bias because both cases and controls have sought medical care for an acute respiratory infection [95].	• *Studies with secondary data collection:* Number of healthcare visits in a past set period • *TND:* Disease severity, in case this differs by ILI etiology (e.g. requirement a medical visit, hospitalization, ICU, leading to death or disabilities) [5]	• Bias is more likely present in GP settings than hospital settings as healthcare-seeking behavior is likely more similar for more severe disease • VE then only holds for those people that would seek medical care for an acute respiratory infection • *TND:* Selection bias is still possible in TND IVE studies [9]. However, Jackson et al. conducted a simulation study and concluded selection bias was only meaningful when rates of care seeking between influenza ARI and non-influenza ARI were very different; therefore, selection bias is unlikely to be meaningful under conditions likely to be encountered in practice [96].
Misclassification (exposure)	↓: non-differential misclassification ↓or ↑: differential misclassification	*Studies with primary data collection:* • Using the TND reduces differential recall bias as case status is unknown at the time of recruitment [9]. • Studies using an administrative database: • For sources that record the presence of a vaccination but not a potential absence, it must be ensured that they cover the entire population of interest. Otherwise, the part of the population not reflected in the data is by default assumed to be not vaccinated. • It must be assessed to what extent the vaccination records		

Type of bias/confounder/effect modifier	Direction of bias (↓(decrease) or ↑ (increase) VE)	Controlling using study design	Information required to control in the analysis	Comments
		are expected to be complete. For example, GP records may not necessarily capture influenza vaccination at vaccination clinics or in occupational settings. • In case of aggregated vaccination information, it must be further guaranteed that this information originates only from the population of interest to avoid misclassification towards a wrongly increased number of vaccinated. • Checking interview data against vaccination registers or other data sources		
Misclassification (outcome)	↓: non-differential misclassification ↓or ↑: differential misclassification	• Choosing a test with high specificity (more important than high sensitivity) because a non-specific test is expected to increase the proportion of false negatives [97] • Imperfect tests lead to less bias in cohort and case-control than in TND design, however trivial difference when using a highly sensitive and specific test (e.g. RT-PCR) [97] • Considering inconclusive test results as negative causes less bias than considering these to be positive [97] *Studies with secondary data collection:* • it must be ensured that the records on disease status cover the entire population of interest because study subjects without	• Use of antivirals • Laboratory test used • Sensitivity and specificity of laboratory test. • Lag time between symptom onset and testing • For non–laboratory-confirmed outcomes, an analysis in different periods of the season (peak or outer ends of season), where non-specific outcomes will likely be more specific in the peak period.	

Type of bias/confounder/effect modifier	Direction of bias (↓(decrease) or ↑ (increase) VE)	Controlling using study design	Information required to control in the analysis	Comments
		an entry related to the outcome of interest are not considered as cases		
Age		• For young children: to collect age in months	• Age as a continuous variable or in age groups [age groups to be harmonized!]	
Gender/sex	↓ or ↑		• Male/female, from healthcare records	
Previous influenza vaccine (any) in past seasons	↓ or ↑	• Yes/No, from past healthcare records or interview	•	
Concomitant administration of a COVID-19 vaccine	↓ (or ↑)		• Yes/No and date, from vaccination registry	The effect of COVID-19 vaccination on IVE estimates it is yet to be fully explored
COVID-19 vaccine	↓	• Exclusion of SARS-CoV-2 test-negative controls from TND studies • Adjustment for COVID-19 vaccination in TND studies	• Yes/No, from vaccination registry	
Previous pneumococcal vaccine in past years	↑: for non-specific outcomes (that include secondary bacterial infectious complications)		• Yes/No, from past healthcare records or from interview	
Previous laboratory-confirmed influenza infection in past seasons	↓ or ↑		• Yes/No, from past healthcare records	• Controlling for previous influenza infection is extremely difficult, as multiple factors are involved such as time since infection, influenza strain, diagnosis of past infection (depending on healthcare-seeking behavior and laboratory-confirmation), and availability of past health records.
Child's adherence to the local childhood vaccination program	↓ or ↑		• Receipt of childhood vaccines	

Type of bias/confounder/effect modifier	**Direction of bias (**↓(decrease) or ↑ (increase) VE**)**	**Controlling using study design**	**Information required to control in the analysis**	**Comments**
Infection pressure	↓ or ↑		• Healthcare worker yes/no • Army conscript yes/no • Childcare worker yes/no • Day care attendance (for pre-school children) • Number of siblings (for children)	• Controlling for differences in exposure to influenza virus is extremely difficult, as this variable is largely unobserved.
COVID-19 infection	(↓)		• Yes/No from healthcare administrative database	The effect of COVID-19 infection on IVE estimates it is yet to be explored
Pregnancy				
Smoking / parental smoking status	↓ or ↑		• From healthcare records or interview (non-smoker, ex-smoker, smoker)	
Institutionalization	↓ or ↑		• Institutionalized vs. non-institutionalized • From healthcare records or interview	
Socioeconomic status (or proxy)	↓ or ↑			
Statins	↓ or ↑		• Use of statins	
Residence in long-term care facility	↓ or ↑			
Waning immunity	↓ over time		• Time since vaccination	
Virus characteristics	VE may change over time		• Calendar time (to account for any drift)	

5.5 Recommendations

For studies using the test-negative design, in DRIVE project we suggest the collection of the following data points:

- Age in months (for children <1 year old) or years
- Gender
- Chronic underlying conditions (e.g. liver disease, heart disease, diabetes, cancer, immunodeficiency/ organ transplant, autoimmune disease, lung disease, anemia, renal disease, dementia, history of stroke, rheumatologic diseases, obesity)
- Past healthcare use (e.g., nr of GP visits, nr of hospitalizations in the past period. The impact of COVID-19 on healthcare use should be considered when choosing the period.
- COVID-19 infection and/or vaccination (yes/no and timing)
- Use of influenza antivirals (type, timing)
- Lag time between symptom onset and testing
- Calendar time

Not all data points that can be collected through primary data collection are available in secondary data. For studies using secondary data, we suggest the collection of the following data points.

- Age (stratification by age groups, minimally 6 months-14 years; 15-64 years; 65+ years)
- Gender
- Chronic underlying conditions
- Past healthcare use (e.g., nr of GP visits, nr of hospitalizations). For studies conducted during the COVID-19 pandemic, the impact of COVID-19 on healthcare use may be considered when choosing the period.

Each study's SAP should explain how the data will be used in the analysis. Information on how to adjust for confounders in the analysis can be found in Chapter 8.

6 Optimization of the value of microbiological and virological information

The WHO recommends using laboratory-confirmed outcomes as opposed to non-specific (syndromic) outcomes [2]. The decision to assess potential study subjects for laboratory-confirmed influenza virus infection should be based on pre-specified protocol guidelines, to avoid systematic misclassification of study subjects, which may arise if clinicians are allowed to decide whom to test. Whenever possible, study protocols should specify the symptoms, duration of illness, and other eligibility criteria for enrolling and testing patients for influenza (in studies based on administrative databases this cannot be done). According to the Committee for Medicinal Products for Human Use Guideline on Influenza Vaccines, Non-clinical and Clinical Module [3], cases should meet the EU ILI and influenza case definitions. An influenza case definition that includes laboratory confirmation is essential to enable estimating IVE against influenza. Available laboratory tests to confirm influenza can be grouped in direct or indirect diagnostic tests.

6.1 Direct diagnostic tests

Direct diagnostic tests to identify influenza viruses are done on nasal or throat swabs, nasopharyngeal aspirates or bronchoalveolar washes [98]. Since the average duration of virus shedding in infected persons is around five days and highest around the time of illness onset [99], the sample should ideally be collected within seven days after illness onset [2] (in studies with primary data collection) to reduce the likelihood of a false-negative test result (otherwise sensitivity would be harmed). Sensitivity may be further improved by choosing non-cases from swabs testing positive for another respiratory virus to ensure that the sample is of sufficient quality to detect virus [9] (alternatively, reference/housekeeping genes could be co-detected [100]).

6.1.1 Rapid influenza diagnostic tests

A number of easy-to-use, rapid influenza diagnostic tests (RIDTs) have become available in recent years to detect influenza virus antigens or viral enzyme activity. Results are provided in approximately 15 minutes and can be used at the point-of-care (POC) in a routine clinical setting (e.g. at the patient's bedside or at the physician's office).

A major downside is that RIDTs cannot distinguish between influenza A subtypes or, for some tests, between influenza types A and B. Moreover, they are less sensitive and considerably less specific than RT-PCR and can therefore cause disease misclassification [2]. According to the WHO recommendations on the use of rapid testing for influenza diagnosis, issued in 2005, the median sensitivity of rapid tests is 70–75%, lower than that of cell culture, while their specificity usually

exceeds 90% (median 90–95%) [101]. A more recent metanalysis of 159 studies involving 26 RIDTs found that RIDTs have a high specificity (98.2%, 95% CI 97.5% to 98.7%) and positive likelihood ratio (34.5, 95% CI, 23.8 to 45.2) and modest and highly variable sensitivity (62.3%, 95% CI 57.9% to 66.6%) for detecting influenza [102]. Because of the low sensitivity, false-negative results are a major concern with RIDTs, which tend to underestimate IVE [103, 104]. The potential for false-negative results is a concern, especially during peak influenza activity. These findings mean that a positive RIDT result in a patient with ILI provides firm support for the diagnosis of influenza, whereas a negative RIDT result has a reasonable likelihood of being a false negative and therefore should be confirmed by other laboratory diagnostic tests (RT-PCR, viral culture, or immunofluorescence) [2].

According to Tanei et al., testing too early could increase false negatives; therefore, RIDT should not be used soon after onset[105].

Since young children have higher viral loads and longer viral shedding than adults, RIDTs perform better in children, with approximately 13% higher sensitivity than adults. Similarly, RIDTs have a higher sensitivity for detecting influenza A, which causes more severe disease and therefore, usually, a higher viral load than influenza B. In the metanalysis by Chartrand et al., no single commercial brand of RIDT performed markedly better or worse than the others; however, authors cautioned that head-to-head comparisons were not made in most studies [102].

Although specificity is high, false-positive results can also occur, especially when influenza activity is low.

6.1.2 Detection of viral proteins by lab techniques

Detection of viral proteins by immunofluorescence, using direct fluorescent antibody (DFA) testing, also known as the immunofluorescent antibody test (IFA), or enzyme immunoassay (EIA), involves sedimentation of respiratory epithelial cells onto a well slide and subsequent staining with influenza-specific antibodies conjugated to fluorescent dye [106]. It is a rapid, relatively low-cost, and commercially available method. Results are available in 1-4 hours.

For seasonal influenza viruses, DFA sensitivity ranges from 60%–100%, compared to the traditional viral isolation procedures (Table 6.1. Synthesis of laboratory tests performances in terms of sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV).). The sensitivity of DFA tests for seasonal influenza appears to be higher than that of POC tests; however, DFA requires technical expertise and the availability of a fluorescence microscope

[106]. During the 2009 H1N1 pandemic, DFA tests sensitivity was variable, ranging between 38% and 93%, compared to RT-PCR-based approaches [34, 107, 108]. In the study by Bakerman et al. it missed one-third of infected individuals, as the reported test sensitivity was 65.0%; with addition of viral culture, sensitivity improved to 81.3% compared to PCR as the gold standard.

6.1.3 Viral culture

Culture of influenza virus from a respiratory specimen represented the gold standard for diagnosis in the past. Results are available in three to 10 days; this reduces its utility for patient management. Shell viral culture (SVC) is another viral culture approach in use since the early 1990s. It consists of a rapid culture method that provides results in 24-48 hours. It involves the propagation of viruses in mammalian cells grown in small 1-dram vials or shell vials, followed by staining with influenza virus-specific fluorescent monoclonal antibodies [107]. SVC has higher sensitivity compared to the traditional viral culture technique. A modified SVC method using R-mix cells, a mixture of mink lung cells and human adenocarcinoma cells, has even higher sensitivity and a turnaround time of 1.4 days.

Viral culture is almost 100% specific and is nearly as sensitive as PCR when samples have high viral titers [109], e.g. in children, as children shed virus in higher titers and for longer periods than adults. Typically, the influenza virus-specific monoclonal antibodies used for virus detection in culture target the conserved NP protein, which allows distinction between influenza type A vs type B. However, this method cannot be used routinely since it is quite cumbersome, requires expertise, specialized equipment, and a long testing time.

6.1.4 Reverse Transcriptase-Polymerase Chain Reaction

The Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) has replaced viral culture as the gold standard test for laboratory confirmation of influenza virus infection during acute illness. It involves three essential steps:

i) Extraction of viral RNA from clinical specimens;
ii) Reverse transcription of viral RNA to a single-stranded cDNA using the enzyme reverse transcriptase; and
iii) Amplification of the PCR product coupled to fluorescent detection of labeled PCR products [107].

Beyond laboratory confirmation of the presence of influenza virus infection, RT-PCR provides a direct and complete identification of influenza A viruses viral type and subtype and of influenza B

viruses lineage, so it allows IVE estimate stratification by virus type/subtype or lineage. Nearly all RCTs use RT-PCR assays for influenza testing [110], based on the fact that it is currently the most sensitive (sensitivity is near to 100%, see Table 6.1 and Table 6.2; lower limit of detection, 1–10 infectious units [106]) and specific (specificity ranges between 91.1% and 100%) method for detection of influenza viruses [99, 111]. It has a 2% to 13% higher detection rate than viral culture, and results can be obtained within hours [101]. Moreover, it enables appreciation of the genetic variability of influenza viruses.

The multiplex PCR identifies other viruses than influenza, the most important of which are respiratory syncytial virus A and B, rhinovirus/enteroviruses, and human metapneumovirus; followed by parainfluenza viruses, adenoviruses, coronaviruses and bocaviruses.

Potential disadvantages of RT-PCR are represented by the fact that it requires technical expertise and expensive equipment and that its ability to detect the virus depends on the amount of virus in a sample: the less virus, the more cycles required to identify influenza products. The US Centers for Disease Control and Prevention suggested using 37 cycles as the limit for classifying a sample as positive because positive results detected using a higher number of cycles may be false positives (reduced specificity) [99]. Finally, in case of novel virus appearance, rapid production and validation of new primer and probe sets may be required [106].

6.1.5 Other amplification techniques

An alternative molecular technique can be uses for a sensitive detection of viral RNA, including LAMP and TMA technologies. These technologies have the same sensitivity and specificity as compared to RT-PCR. These techniques can be of high throughput (TMA) or rapid (LAMP).
The droplet digital PCR (ddPCR) is another technique available. However, the ddPCR is not adapted for diagnostic purposes and, as a consequence, should not be used for IVE studies.

6.1.6 Sequencing techniques

Sequencing of influenza hemmaglutinin is important to investigate the putative virological causes of vaccine failure during IVE studies. Despite the implementation of this new technology in numerous laboratories, sequencing is not adapted for detection purposes and should be used only for virus characterization to analyze the match between vaccine and circulating strains.

6.2 Indirect diagnostic tests

6.2.1 Serology on paired blood samples

The approach based on virologically-confirmed human influenza cases has the disadvantage that virus shedding in infected persons typically lasts only a week and has often diminished or ended by the time of sampling [112]. In addition, infections may cause only mild illness, leading to cases possibly remaining undetected. Studies based on the serological evidence of infection have a wider window of detection. Data, however, need to be interpreted with caution due to cross-reactivity of antibodies among and within virus subtypes and sensitivity decrease when used to detect antibodies against novel influenza subtypes [113].

Serology consists of collecting a first blood sample at symptoms onset and a second two to three weeks later. Influenza infection is defined by at least a 4-fold rise in specific antibody titer between paired sera. Antibody titer can be obtained by different methods. Those most commonly used are the haemagglutination inhibition assay (HAI), microneutralization or virus neutralization assay (VN), single radial hemolysis (SRH), complement fixation assay, enzyme-linked immunoabsorbent assay (ELISA) and Western blotting [107]. One major concern of ELISA-based tests is the lower sensitivity compared to nucleic acid-based tests (NATs). A novel europium nanoparticle-based immunoassay for rapid detection using monoclonal antibodies directed against the nucleoprotein from influenza A and influenza B viruses showed a sensitivity of 90.7% for influenza A viruses and 81.80 % for influenza B viruses with 100% specificity [114].

This test is unlikely to become relevant in DRIVE.

Table 6.1. Synthesis of laboratory tests performances in terms of sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV).

PUBLICATION	TEST	SE (%)	SP (%)	PPV (%)	NPV (%)	Notes
Baker man 2011 [115] Nasopharyngeal swabs	DFA	65.0 (59.0, 71.0)	99.6 (99.1, 100)	98.8 (97.0, 100)	85.9 (83.2, 88.7)	
Chartrand 2012 [101] viral culture and RT-PCR used as references	RIDT	62.3 (57.9-66.6) There was a considerable overlap among the accuracy estimates for the RIDTs. Directigen Flu A had the highest pooled sensitivity (76.7% [CI, 63.8% to 86.0%]), followed by QuickVue Influenza test, although the difference from the overall estimate was not statistically significant. However, BinaxNOW, Directigen Flu A+B, and QuickVue Influenza A+B had a lower sensitivity compared to the overall estimate (57.0%, 57.2%, and 48.8%, respectively).	98.2 (97.5-98.7) Specificity was consistent among most RIDTs	34.5 (23.8-45.2)	0.38	POOLED ANALYSIS RIDTs had a significantly higher pooled sensitivity when compared to viral culture rather than RT-PCR because of the increased accuracy of the latter. Subgroup analyses showed that RIDTs had a significantly higher pooled sensitivity in children (66%, 95% CI 61.6% to 71.7%; 60 datasets) than in adults (53.9%, 95% CI 47.9% to 59.8%; 33 datasets); specificities were similar between the age groups. RIDTs were associated with a significantly higher sensitivity for detecting influenza A (64.6%, 95% CI 59.0% to 70.1%; 72 datasets) than for detecting influenza B (52.2%, 95% CI 45% to 59.3%; 27 datasets). Results were unchanged when RIDT brand, specimen type and reference
Ganzenmueller 2010 [105] Respiratory specimens (nasopharyngeal swabs, pharyngeal washes and bronchoalveolar lavage samples), using RT-PCR as gold standard	Point-of-care (POC) RIDT	18.2	100	100	78.1	Detection of novel 2009 influenza A (H1N1)
	DFA	38.7	100	100	82.2	
	Virus isolation	45.7	99.8	95.5	94.8	
Ginocchio 2009 [108] Detection of	DFA	47.2	99.6	90.6	96.2	Ages ranged from four days to 98 years; authors did not differentiate between adult and paediatric populations
	RIDT	21.2	99.5	76.5	94.5	

novel 2009 influenza A (H1N1)	Viral culture	98.4	100	100	99.9	
Hindiyeh 2005 [116] Throat and nasal swabs Viral culture as reference	Multiplex TaqMan	Influenza A: 100 Influenza B: 95.7	Influenza A: 91.1 Influenza B: 98.7	Influenza A: 84.7 Influenza B: 94.3	Influenza A: 100 Influenza B: 99	The authors concluded that the multiplex TaqMan assay is highly suitable for the rapid diagnosis of influenza virus infections both in well-established molecular biology laboratories and in reference clinical laboratories.
	Multiplex RT-PCR	Influenza A: 100 Influenza B: 100	Influenza A: 93.1 Influenza B: 98.3	87.8 Influenza B: 93.2	Influenza A: 100 Influenza B: 100	
	IF	Influenza A: 84.4 Influenza B: 33.3	Influenza A: 98.8 Influenza B: 100	Influenza A: 97.2 Influenza B: 100	Influenza A: 92.8 Influenza B: 86.7	
Kenmoe. 2014 [117] Nasal swabs RT-PCR as gold standard	RIDT	29.4	100	100	89.5	
Landry 2014 [118] Nasopharyngeal swabs Laboratory-developed TaqMan PCR methods used as reference	DFA	Influenza A: 62.5 (53.6 to 70.7) Influenza B: 69.7 (52.5 to 82.8)	Influenza A: 100 (98.6 to 100) Influenza B: 100 (98.9 to 100)			
	RT-PCR Simplexa Flu A/B*	Influenza A 2µl sample: 83.3 (75.6 to 89.0) 5µl extract 89.2 (82.3 to 93.7) Influenza B: 2µl sample: 72.7 (55.6 to 85.1) 5µl extract 84.9 (68.6 to 93.8)	Influenza A 2µl sample: 100 (98.6 to 100) 5µl extract: 100 (98.6 to 100) Influenza B: 2µl sample: 100 (98.9 to 100) 5µl extract: 100 (98.9 to 100)			Positive samples with cycle threshold (*CT*) values of_38 were accepted as positive. For *CT* values of >38, amplification was repeated in duplicate and accepted if one replicate was positive. Although Simplexa was less sensitive than current LDT assays, it was simpler, required minimal hands-on time, included an internal control, and had a shorter assay time. Samples missed by Simplexa using extracted samples had very low viral loads. More than 95% of the discrepant results were from adults and, with one exception, were tested late in the course of the illness, when patients presented with secondary complications.

						Thus, the authors suggested that the clinical impact of missing low-viral-load samples might be minimal and that a higher detection rate with Simplexa is anticipated in settings where samples from patients presenting early in illness are tested.
Leonardi 2010 [119] assay performance and nasopharyngeal swab	RIDT (EZ Flu)	66.7	100			Detection of novel 2009 influenza A (H1N1)
	Rapid shell vial culture	100	100			
	Traditional tube culture	100	100			
	DFA	80	100			
Pollock 2009 [110] nasopharyngeal swabs or aspiration RT-PCR used as reference	DFA	93	97	95	96	Specimens collected only from symptomatic HCW or patients who met CDC criteria for influenza-like illness, as part of their routine clinical evaluation.
Reina 2010 [120] nasopharyngeal swabs RT-PCR used as reference	EIA	52.9	100	100	79.7	
	Viral culture	94.1	100	100	96.9	
Scheuller 2015 [109] nasal swab, throat swab, or nasal wash	RT-PCR	93 (all) 92 (BMT) 95 (Non-BMT)				The goal of the study was to better understand how influenza diagnostic tests perform in the basic military trainees (BMT) population, and how this performance differs from the general population (Non-BMT)
	EIA	57 (all) 51 (BMT) 60 (Non-BMT)				
	Viral culture	51 (all) 63 (BMT) 41 (Non-BMT)				
Tanei 2013 [104]	RIDT	72.9 (95 CI 61.5 to 84.2)	91.3 (79.7 to 102.8)	95.6 (89.5 to 101.6)	56.8 (40.8 to 72.7)	
Tuuminen 2013 [121] nasopharyngeal	RIDT (mariPOC)	85.7 (69.7-95.2) (aspirates) 77.3 (54.6-92.2) (swabs)	100 (aspirates) 98.3 (swabs)			The rapid and automated test system mariPOC is based on two-photon excitation fluorometry and the concentrations of antigens and fluorescent

aspirates and swab samples DFA as the primary reference method		Both aspirates and swabs can be analyzed with the mariPOC, although aspirates yielded better sensitivity than swabs.				tracer on microspheres by antigen-antibody reactions. The technology utilizes microvolume reaction chambers and separation-free fluorescence measurement; it allows real-time follow-up of reaction kinetics, and in this application, the test results are read at approximately 20 min and 2 h from the beginning of the reactions. Strong positive samples can be revealed very rapidly, and even the lowest positive samples can be detected at the point of care.
Uyeki 2009 [122] Nasal swab confirmatory influenza testing by RT-PCR (all sites) and viral culture (sites 1 and 2) for all specimens tested at each site throughout the study period	RIDT	27 median 19–32 range	97 median 96–99.6 range	87.5 median 80.0–90.9 range	69.4 median 62.5–79.1 range	
WHO. 2005 [123] Viral culture as gold standard	RIDT	70-75 median	90-95 Median			
	IF	70-100 range	80-100 Range	85-94 range	96-100 range	
Zhang 2104 [113]	ELISA	Influenza A 90.7 (95% CI 86 to 96) Influenza B 81.8 (95% CI 61 to 100)	Influenza A 100 Influenza B: 100			

DFA: Direct florescent antibody testing; EIA enzyme immunoassay; ELISA: Enzyme Linked Immunoabsorbant Assay; HAI: hemagglutination inhibition assay; IF: immunofluorescence; RIDT: Rapid influenza diagnostic test.

Table 6.2. Ranges of sensitivity, specificity, PPV and NPV levels per each test (Source: results from studies listed in Table 6.1)

TEST	Range SE (%)	Range SP (%)	Range PPV (%)	Range NPV (%)
DFA	47.2-93	97-100	90.6-100	82.2-96.2
EIA	52.9-60	100	100	79.7
ELISA	81.8-90.7	100		
IF	70-100	80-100	85-94	86.7-100
RIDT	18.2-85.7	90-100	76.5-100	56.8-94.5
RT-PCR	72.7-100	91.1-100	84.7-94.3	99-100
Viral culture	45.7-100	99.8-100	95.5-100	94.8-100

DFA: Direct florescent antibody testing; EIA enzyme immunoassay; ELISA: Enzyme Linked Immunoabsorbent Assay; HAI: hemagglutination inhibition assay; IF: immunofluorescence; RIDT: Rapid influenza diagnostic test.

6.3 Recommendations

Vaccine effectiveness studies require the identification of viruses. This should be carried out with sensitive and specific techniques and provide detailed relevant biological information about the causative agent to avoid confounders. Consequently, this detection should use the most up-to-date diagnostic tools.

The clinical network involved in the DRIVE studies should be able to collect nasal and/or nasopharyngeal swabs and send these specimens to a corresponding virological lab.

Since evidence shows that virus shedding is significantly reduced three or four days after disease onset, information regarding the delay between disease onset and the specimen collection should be collected in studies with primary data collection, and stratification of IVE estimates according to this delay can be considered. It is recommended to exclude swabs collected more than seven days after start of symptom onset.

For studies performing primary data collection, we suggest using labs that:

- Are able to detect influenza by RT-PCR (first line of screening)
- Further characterize the detected virus by sub-typing (for Influenza A viruses) and lineage determination (Influenza B viruses).
- Have their performance assessed by participation in External Quality Assessment (EQA), as those provided by Quality Control for Molecular Diagnostics (QCMD) [124].

If possible, the lab should also be in the capacity to carry out additional influenza testing:

- Genotyping of the virus (HA and NA gene sequencing, by Sanger or NGS, for genetic clade determination, full genome sequencing should also be an objective). This can be very helpful for comparing strains and interpreting IVE results.
- Strain characterization for the identification of potential antigenic variants. This means being able to grow influenza viruses on MDCK cells and subsequently determine their antigenic profile with ferret sera. This will allow the complete antigenic characterization of the influenza viruses, according to the WHO standards as described by the CDC [115].

In addition, the labs may detect with the same techniques other respiratory viruses such as COVID-19, RSV, rhinoviruses, human metapneumoviruses, adenoviruses and parainfluenza viruses; all these viruses can co-circulate during the influenza epidemic and may be responsible for ILI presentation.

For studies using secondary data collection, the preferred lab method to test for influenza is also RT-PCR.

7 Methods for rapid assessment of IVE

7.1 Purpose of rapid IVE assessment

Due to the continuous evolution of the influenza viruses and potential mismatches between the circulating and the vaccine strains [125], the effectiveness of different seasons' influenza vaccines can differ a lot [2]. Rapid or real-time assessment of IVE before the end of the epidemic is important for several reasons, including to contribute to the monitoring of the benefit-risk assessment of newly reformulated influenza vaccines without undue delay [22].

In case early estimates during a given season are available and indicate a low IVE, additional preventive measures can be put into practice to ensure a high level of protection in the population, e.g. recommendations to continue the use of vaccination, complementary to the use of antivirals to mitigate the influenza-associated complications. Likewise, early knowledge about a highly effective vaccine might further increase vaccine uptake in the population. Either way, rapidly available impact measures communicated to the population during an epidemic can strengthen individual health and public health. At the same time, rapid or real-time IVE figures calculated before the end of the epidemic may be misleading for the public opinion if not accompanied by appropriate guidance and precautionary measures. An influenza epidemic is often characterized by more than one wave caused by different viruses circulating in different periods. The vaccines' effectiveness might differ between those waves and thus an estimate based on the first wave is not necessarily a good estimate for the rest of the season.

Moreover, early IVE estimates can also benefit future vaccine compositions for the other hemisphere. The WHO regularly reselects the influenza strains to be included in the vaccines [15]. The recommendation for the Northern Hemisphere is usually made by the end of February [123], which consequently marks the time point when the results of the rapid IVE assessment would be expected, at the latest.

7.2 Applied methods for rapid IVE assessment

The estimation of the IVE during an ongoing influenza epidemic lays high demands on the study design, data sources, and the data collection process. However, the only major difference between end-season and rapid (mid-season) IVE assessment in terms of data analysis is the length of the study period. By analyzing only those data collected until a respective cut-off day before the end of the influenza epidemic, several recently published studies present ‘early’, ‘interim’, or ‘mid-season’ estimates [7, 126-131]. The majority of these studies applied the test-negative design [7, 126, 127, 129-131], while only one utilized the cohort design [128]. In the context of rapid IVE assessment, the pros and cons of the two approaches remain as described in 2.3.1 and 2.1, and one would expect no

difference between the study designs, as long as the different sources of bias are correctly controlled for (see 5.4).

The challenges of such studies are related to the rapidness in which they must be conducted and to statistical power considerations. While respiratory specimens should always be analyzed in a timely manner to benefit a patient's treatment, data collection and statistical analysis only need to be accelerated for a rapid IVE assessment. Accordingly, there is no time for extensive data validation steps. Moreover, the data must be collected in real-time or with a known delay to ensure that the study period is completely covered by the data on the day of analysis. Because the study period is ended before all cases of an influenza epidemic have occurred, many rapid IVE estimations are characterized by the lack of statistical power due to small numbers of cases. Generally, it is considered that the longer the study period, the more precise the estimates.

This trade-off between rapidness and statistical power and thus reliability of estimates also affects the decision of when to conduct IVE analyses. All the studies from the Northern hemisphere referenced above had finished their interim analyses in the first half of February 2017 [7, 126-130], i.e. before the WHO Consultation and Information Meeting on the Composition of Influenza Virus Vaccines for Use in the 2017-2018 Influenza Season [132]. Furthermore, three studies indicated that they had ended the study period about two weeks after the peak of the epidemic curve [128, 129, 131].

7.3 Recommendations for future near real-time IVE assessment

In the future, any study design that has been proven to yield valid and reliable estimates can be chosen to rapidly assess IVE in near real time. Large test-negative case-control studies implemented on top of multiple, existing influenza surveillance/sentinel systems and large cohort studies based on automated (secondary) data collections might be the most feasible strategies. The appropriate timing within the season depends on the actual purpose of the IVE figures and the course of the epidemic. It seems advisable to assess IVE early but after the peak of the influenza-type specific epidemic curve, since it has been shown that such studies can reliably predict even end-season IVE [133, 134]. However, in order to provide data to the WHO to support the decision on future vaccine compositions, one might also preliminarily end the study period shortly before or at the top of an intense epidemic that started late.

In DRIVE, no interim analyses were performed due to a lack of statistical power for brand-specific estimates.

8 Data analysis for individual studies

8.1 Vaccine effectiveness measures

The effect measures to be estimated and quantified in DRIVE are brand-specific IVEs. IVE is usually defined as

$$VE = 1 - RR,$$

where *RR* denotes the relative risk of the outcome for vaccinated individuals versus unvaccinated individuals [135]. *VE* is estimated by either one minus a confounder-adjusted estimator of the ratio of the influenza attack rates in a cohort or by one minus a confounder-adjusted estimator of the ratio of the influenza incidence rates in a cohort or a dynamic population. When the study is a case-control design, dependent on the sampling design, either the attack rate ratio or the incidence rate ratio can be estimated by the sample odds ratios. If the study is a cohort design, the attack rate ratio can be estimated using an attack rate or risk ratio estimator, and to estimate the incidence rate ratio, the sample incidence rate ratio or hazard rate ratio is generally used.

In Table 8.1, the estimator type and the estimator are given for a number of study designs in Chapter 2.

Table 8.1: Estimator for the analysis of IVE data, by study design and outcome data

Study design	Type of control sampling	Outcome data	Estimator
Cohort design		Count data (cases, non-cases)	Attack rate ratio
		Time-to-event data (also person-time data)	Hazard rate ratio or incidence rate ratio
Nested case-control design	Cumulative sampling	Count data (cases, controls)	Odds ratio
	Density sampling	Count data (cases, controls)	Odds ratio
Case-cohort	Case-base sampling	Count data (cases, referents)	Attack rate ratio with pseudo-denominators [136]
Case-control design with density sampling	Density sampling	Count data (cases, controls)	Odds ratio
Test-negative design	No sampling	Count data (cases, controls)	Odds ratio

The effect measure being estimated in the analysis of IVE data is not *VE*, but the attack rate ratio (cohort design, cumulative sampling) or the incidence rate ratio *IRR* (test-negative design, density sampling) If $LL_{(I)RR}$ and $UL_{(I)RR}$ are the lower and the upper limit of a confidence interval for *(I)RR*, then 1- $UL_{(I)RR}$ and 1-$LL_{(I)RR}$ are the lower and the upper limit for a confidence interval for *VE.*

8.2 Power and sample size

When planning a VE study, the power and sample size calculations can be strongly influenced by a number of particular parameters. First, in case brand-specific VE is of interest, the expected overall vaccination coverage, the share of each brand, and the assumed VE of each brand have a large influence on these calculations. Similarly, when the goal is to obtain strain-specific VE estimates, one has to take into account the distribution of the different strains. To accommodate these options within DRIVE, a dashboard has been developed to perform sample size calculations for brand- and strain-specific VE cohort or (test-negative) case-control studies (https://shinyproxy.p-95.com/app/drivesamplesize).

8.3 Adjusting for confounders

8.3.1 Statistical methods

In the statistical analysis, confounders can be adjusted for (controlled) either by stratification or by regression. Stratified analysis works best if the number of confounders to adjust for is small. For each possible combination of confounder levels, a separate stratum must be created, with the risk of a large amount of sparsely populated strata with too little data to estimate the association between vaccination and prevention of the outcome with any reasonable degree of precision. Given that in IVE studies the number of confounders to be adjusted for is usually non-small, stratification to adjust for confounding is not advised. On the other hand, stratification is an important tool to inspect effect modification, see Section 8.4.

Alternative approaches to confounder adjustment are regression and propensity scoring.

8.3.2 Regression models

Confounders are adjusted for by including them as covariates in the regression model.
In Table 8.2, appropriate regression models for the analysis of IVE data are listed by source population, study design and type of data.

Table 8.2: Regression models for the analysis of IVE data, by study design and type of outcome data

Study design	Outcome data	Regression model
Cohort design	Count data (cases, non-cases)	General log-linked binomial regression model
	Time-to-event data	Cox regression Poisson regression
Case-control design with cumulative sampling	Count data (cases, controls)	Unconditional logistic regression
Case-control design with density sampling	Count data (cases, controls)	Unconditional logistic regression with adjustment for calendar time Conditional logistic regression with matching on calendar time
Case-cohort design	Count data (cases, controls)	Unconditional logistic regression with pseudo-likelihood [137]
Case-control design with density sampling	Count data (cases, controls)	Conditional logistic regression with matching on calendar time Unconditional logistic regression with adjustment for calendar time
Test negative design	(cases, controls)	Conditional logistic regression with matching on calendar time Unconditional logistic regression with adjustment for calendar time

Logistic regression for case-control designs with cumulative sampling

Logistic regression is the standard regression model for the analysis of nested case-control data. The parameter of interest is the odds ratio (*OR*). If the influenza attack rates in the cohort are low, $OR \approx RR$.

Logistic regression for case-control designs with density sampling

Just as in logistic regression for nested case-control designs, in logistic regression for case-control designs with density sampling, the input data are numbers of cases and controls. However, there are some subtle differences to be aware of:

- the parameter being estimated is the sample odds ratio, but in this design, sample OR estimates the population incidence rate ratio
- controls represent person-time, individuals may be at different times sampled more than once as control or may become a case after being sampled as a control; these should be treated as independent observations [138]
- a person who has been sampled twice at different times may have changed his/her vaccination status in between
- calendar time should be included in the regression model to control for seasonal patterns or time-dependent differences in the vaccinated/unvaccinated ratio

Logistic regression for test-negative design studies

In case the control subjects are sampled using sampling scheme mimicking density sampling, the regression analysis can proceed as for the previously described case-control studies with density

sampling.

Cox proportional hazards regression

The Cox proportional hazards regression model is widely used to analyze time-to-event data. The parameter of interest is the hazard ratio *HR*, which is often interpreted as the influenza incidence rate ratio. The time-to-event is defined as the time-span between the start of the influenza season and the occurrence of the outcome, assuming that all vaccinations were given prior to the start of the influenza season; alternatively, vaccination can be modelled as a time-varying variable. For subjects who did not experience an event, the time-to-event is right-censored at death, moving out of the catchment areas, receiving an influenza vaccination other than the defined exposure (when studying type or brand-specific IVE), having a (confirmed) influenza infection other than defined outcome (only applicable for influenza type- or strain-specific analysis), or the study period, whichever comes first.

Poisson regression

When time-to-event data is available, an alternative to Cox proportional hazard models are Poisson regression models in which the follow-up time is included as an offset term. It can be shown that the estimated coefficients from a Poisson regression model with a piecewise constant function of calendar time will be identical to those of a proportional hazards model with a piecewise constant baseline hazard [139, 140]. An attractive property of Poisson regression models in multi-site studies is that they can be used even when only aggregated data can be shared, i.e., number of events and total follow-up time.

8.3.2.1 Representing confounders in the regression model

The approach advised to represent categorical (dichotomous, nominal and ordinal) confounders in the regression model is using indicator variates because this approach maximizes the thoroughness of control [141]. If a categorical confounder has *k* categories, then (*k*-1) indicator variates must be defined:

X_2 = indicator for the second category

.

.

X_k = indicator for the k^{th} category

The first category is the reference category.

Incorrect modelling of a continuous confounder can result in residual confounding. If the association between the continuous confounder and risk of influenza infection is not linear, but, for example, U-

or J-shaped, the assumption of a linear relation between the confounder and influenza infection can result in substantial residual confounding. It has been shown that modelling the relation between a continuous variable and an outcome using fractional polynomials [142] and restricted cubic splines [143]. or stratification of the confounder in five strata, tends to limit residual confounding and lead to similar results [144]. Within DRIVE, the focus is mainly on smoothed restricted cubic splines as the methodology is well developed and does not strongly rely on the selection of knots or cut points.

8.3.2.2 Selecting confounders in the regression model

Ideally, confounder selection is partly based on previously published literature and expert knowledge. However, we acknowledge that such information might be unavailable and incomplete and data-driven methods can provide additional insights.

One commonly used data-driven confounder selection method is the following change in estimate strategy:

- all *known* confounders specified in the protocol (dictated by the knowledge of the disease, the medical understanding) should be adjusted for, that is, be included (“forced”) in the regression model, regardless of their significance in the specific study.
- select possible confounders stepwise, one by one in the model, based on the change-in-estimate criterion, that is, at each step add that confounder that leads to the greatest change in the estimate of the relative risk
- stop adding variables to the model if the changes in the relative risk estimate become non-meaningful; popular choices for cut-offs for non-meaningful changes are 5% and 10%

An alternative to the forward-selection strategy is the backward-deletion strategy. The forward-selection strategy is advised because the backward-deletion strategy cannot be implemented when the problem of sparse data (see below) occurs. However, note that this approach, and many other data-driven approaches, have a number of drawbacks, including:

- Stepwise selection methods are generally unable to differentiate between different causal mechanisms such as confounding, collider effects, etc.
- Stepwise selection methods are often based on easily criticized criteria, e.g., the selection of a cut-off value for the change in estimate is not straightforward [145].
- Unless the data is divided into discovery and test data, advanced post-selection inference methods have to be used to obtain valid confidence intervals and unbiased estimators.

8.3.2.3 Disadvantages of regression adjustment to confounder control

The regression approach to controlling for confounding variables has a few major disadvantages.

First, generally, these models assume that the effect of the confounders on the outcome has a particular functional form. Additionally, one should be aware of the risk of sparse data bias. This bias occurs when there are no or only a few infected cases for some combinations of confounders, and it can occur even in quite large data sets. If sparse data bias is suspected, reducing the number of confounders may be attempted, or penalized models can be utilized [146]. A diagnostic test for the risk of sparse data bias is that the total number of cases divided by the total number of variables in the model should not be lower than 7.

8.3.3 Propensity score methods

The propensity score method removes confounding caused by the observed covariates by balancing baseline covariates values between vaccinated and unvaccinated subjects [147]. This is achieved by assigning each subject a so-called 'propensity score.' The propensity score is then the predicted probability of being vaccinated. VE estimates are obtained using a inverse probability of treatment weighted analysis, adjusting for the propensity score as a linear or categorical variable, or by matching subjects with similar propensity scores. While in many cases similar results will be obtained, there are important potential advantages to propensity scoring over conventional regression. For example, with propensity scoring, one needs not to be concerned with overparameterization and can include non-linear terms and interactions. When influenza attack rates are low but vaccination is common, propensity scoring may be better than logistic regression if many confounders must be adjusted for. Finally, propensity scoring tends to be the more robust method. For an example of vaccine effectiveness with propensity scoring, see Simpson et al. [148].

8.4 Effect modification

An effect modifier is a variable that differentially (positively or negatively) modifies the observed effect of the exposure on the outcome. Different groups have different risk estimates when effect modification is present [43]. A known effect modifier is age. The standard approach to study effect modification is to divide the effect modifier in two or more distinct strata and adding the appropriate product (interaction) terms (vaccination times stratum) to the statistical model.

8.5 Missing covariate data

Missing covariate data can be handled by multiple imputation (MI) [149, 150]. Simulation results suggest the application of predictive mean matching after regression switching generally performs well, unless 50% or more of the subjects have missing data, or the missing data are 'missing not at random' (MNAR) [151]. When data are MNAR, exploring other methods such as the pattern-mixture model (PMM) approach is advised. For a discussion of this approach as well as publicly available

SAS macros, see chapter 7 of the book by O'Kelly and Ratitch [152].

8.6 Statistical models to deal with limitations of laboratory tests

In order to adjust for non-differential disease misclassification, correction equations, such as the one showed in Figure 8.1, (in which π_{Other} is the risk of disease due to other pathogens than those targeted by the vaccine; p_0, the observed disease prevalence among the subjects indicated as unvaccinated; p_1, the observed prevalence among the subjects indicated as vaccinated; and SP_d, the disease specificity of the case definition), which requires the disease specificity estimate, should be applied to the estimate of vaccine effectiveness [153].

$$VE_{\pi} = 1 - \frac{p_1 - (1 - SP_d)\pi_{Other}}{p_0 - (1 - SP_d)\pi_{Other}}.$$

Figure 8.1 Correction equation of IVE adjusting for disease misclassification. Source: [153]

8.7 Recommendations

When analyzing data of an individual VE study, the study design should be taken into account because the parameter to estimate the VE differs between designs. Confounders of the relationship between vaccination and prevention of infection should be adjusted for, which can be achieved by regression or propensity scoring. Known confounders should be included in the statistical model, regardless of their significance. If additional variables are considered potential confounders, data-driven regression or propensity score methods can be explored, but care should be taken when interpreting their results. Within DRIVE, logistic regression analyses are used for the TND studies, and Poisson regression is used for the cohort studies. All analyses are stratified by age group, as in this setting, age is generally considered both an effect modifier and a confounder. Additionally, confounder control is primarily done through regression adjustment of known confounders and sensitivity analyses performance to evaluate the impact of other potential confounding variables. Finally, in DRIVE, the regression models tend to use penalized splines to model continuous variables such as age and time of disease onset.

9 Pooling data from different data sources

9.0 Prerequisites

Brand-specific vaccine effectiveness is expected with near certainty to need pooling of data across studies. Differences between studies should be limited where possible to allow pooling. The following aspects should be taken into account:

- Limit difference between individual study designs by:
  - Reducing choices in local study design
  - Clear definitions and optimal compliance
- Pool data across settings with similar characteristics (i.e. primary care vs hospital setting)
- Define upfront the minimum framework to pool the data
- Define quality criteria
- Align local designs by using a generic protocol to ensure homogeneity
  - Define common screening/eligibility criteria used to identify cases/controls (Reasons for GP consultations/hospital admission)
- Define a minimum set of data to be collected

9.1 One-stage vs. two-stage pooling

There are two statistical approaches for pooling data: a one-stage or a two-stage pooling approach. The two-stage approach refers to the classical meta-analytical approach, also called aggregated data meta-analysis (AD-MA). In this approach, the patient-level or minimally aggregated data from each study are analyzed separately in order to obtain the effect estimates of interest (here vaccine effectiveness) and the corresponding confidence intervals (CIs). Then, in the second step, the effect estimates are combined by an appropriate meta-analysis model to obtain the meta-analytical (pooled) estimate. The one-stage pooling approach analyzes all the combined patient-level or minimally aggregated data from the different data sources in a single step. This approach is also called the individual participant data meta-analysis (IPD-MA).

9.1.1.1 Two-stage pooling

Two-stage pooling or AD-MA is the mainstay of systematic reviews [154]. This is the ‘classical’ meta-regression approach, by which aggregated data (typically effect measures) are combined into a pooled estimate and heterogeneity is quantified and possibly explained. There are two popular statistical models for AD-MA, the fixed-effect model and the random-effects model [155]. A fixed-effect meta-analysis assumes all studies are ‘replicates’ estimating exactly the same effect, with the

differences between study estimates solely explained by sampling variability. The random-effects meta-analysis assumes the observed estimates can vary across studies because of 'true' effect differences across populations, differences in the conduct of the study, etc., on top of sampling variability. Heterogeneity between studies is typically quantified, and the sources of heterogeneity can be explored using meta-regression or stratified analysis.

This approach is more flexible than the one-stage pooling approach, as sites sharing only aggregated data can be included.

If studies are comparable in terms of population, exposure and case definitions and control for confounding, pooling is considered appropriate irrespective of study design. Several examples of two-stage pooling of IVE estimates, including estimates from studies with different study designs, exist [156, 157].

Two-stage pooling was used for the DRIVE analyses.

9.1.1.2 One-stage pooling

The application of one-stage pooling or IPD-MA has increased over the last decade, with many examples of combining clinical trial data, particularly in the area of cancer and cardiovascular disease interventions. In 2015, PRISMA-IPD guidelines for reporting systematic reviews and meta-analyses of IPD were published [158]. Advantages of IPD-MA compared to literature-based AD-MA include checking and transforming data to common sources or measures, standardizing analysis and increased flexibility in performing statistical analyses and common reporting. Major disadvantages of IPD-MA include being very time- and resource-intensive and requiring high levels of (international) collaboration [154]; furthermore, data protection standards may restrict the feasibility of this type of analysis. Random effects (multilevel) regression models are used to jointly analyze IPD from all studies while accounting for the (within-study) clustering of subjects [159]. Different model specifications are possible, including correlated and independent random effects, as well as stratified random-effects models [160].

9.2 Equivalence of one-stage and two-stage pooling

The one-stage and two-stage approaches have shown to be equivalent in many situations using theoretical considerations [161-163], simulations [160] or empirical comparisons [164], provided the same data was used. However, some studies have shown different results for both approaches on some occasions [160]. Burke et al. explain why there might be differences [165]. The reasons relate

to different modelling assumptions, parameter specifications and estimation methods, which might sometimes be subtle. In case of a small number of studies or a small number of events, the IPD-MA may be preferred in order to be able to use exact likelihood methods and avoid having to make incorrect assumptions about the between-study variance [165]. However, usually, the AD-MA will suffice [164].

In the DRIVE pilot season 2017/18, the one-stage and two-stage pooling approaches were compared. In line with expectations based on statistical theory, identical to very similar results in the main effect and 95%CI for all models were found [166].

9.3 Combining both approaches

Subject-level data are not always available for all relevant studies, and individual-level and aggregated data need to be combined. There are three approaches for combining IPD and AD in a meta-analysis [167]. First, the available IPD is reduced to AD and then pooled with the other AD. Second, it might be possible to construct IPD from published aggregate information (based on 2x2 tables), which is subsequently combined with the other IPD for analysis. Third, in hierarchical-related regression, the IPD and AD data are analyzed jointly.

9.4 Recommendation

The objective of DRIVE is to estimate brand-specific IVE in Europe by combining data from different study sites. Some site-specific studies will adopt the test-negative case-control study design, while others will use a cohort design. These different study designs imply that different data will be collected and that different statistical analyses are needed to analyze these data (i.e. logistic or conditional logistic regression for case-control versus Poisson or Cox regression for cohort studies). Given the statistical equivalence of AD-MA and IPD-MA and given the additional complexity or even impossibility of performing IPD-MA when data are collected using different study designs or only aggregated data are shared, the AD-MA is the preferred method for combining data from different study sites. Furthermore, many of the mentioned advantages of IPD-MA (i.e. transforming data to common sources or measures and standardizing analysis) can also be achieved through harmonization/standardization of the individual site-specific studies.

Within AD-MA, the random-effects meta-analysis approach is preferred to the fixed-effects approach for combining the data from the site-specific studies on IVE. The assumption that the between-study variability is explained by sampling variability only, which underlies the fixed-effects meta-analysis, is not realistic for the studies on IVE in Europe. There are many differences between these studies, including differences in population, design, exposure- and disease ascertainment as well as in

covariate adjustment. For DRIVE, we used a random effects meta-analysis approach for every effect measure of interest, stratified by age and clinical outcome, while at the same time standardizing the design and conduct of the site-specific studies to the extent possible.

10 Summary of recommendations

In this section, the recommendations provided at the end of each chapter are listed.

Study design

For studies using primary data to monitor IVE, we suggest using the test-negative case-control design, with an appropriate choice of the control group and implementation of sampling protocols [14].

For studies using secondary data, we suggest using the cohort design.

Exposure

The following data on exposure should be collected, both for studies collecting primary data and those using secondary data:

- Vaccine brand(s) or type(s) of all influenza vaccinations given during the index season (i.e. the season for which IVE is being estimated); *for studies on brand- or type-specific IVE*
- Vaccination date(s) of all influenza vaccinations given during the index season, or if not available, the sequential order and relative timings of exposure and outcome
- How the vaccination status was ascertained and whether it was confirmed, e.g. through medical records

Outcome

In studies collecting primary data, the recommended outcome is laboratory-confirmed, medically attended influenza. We suggest the collection of the following data:

- Symptoms forming the clinical syndrome of ILI or SARI, including the information whether hospitalization or intensive care treatment was required,
- Date of symptom onset,
- Date the respiratory specimen was taken,
- Laboratory confirmation yes/no, and if yes, influenza type and preferably also subtype/ lineage.

In studies utilizing secondary data, e.g. from existing healthcare databases, the recommended outcome is laboratory-confirmed influenza, overall or stratified by clinical condition. However, this recommendation does not exclude the use of syndromic or non-specific outcome definitions

discussed in 4.1, either in association with a positive influenza test or alone. We suggest the collection of the following data points:

- Clinical condition (if applicable),
- Date the respiratory specimen was taken,
- Detected influenza type and preferably also subtype/lineage for laboratory-confirmed influenza.

Potential biases and confounders

For studies using the test-negative design, we suggest the collection of the following data points:

- Age by month (for children <1 year old) or year
- Gender
- Chronic underlying conditions (e.g. liver disease, heart disease, diabetes, cancer, immunodeficiency/ organ transplant, autoimmune disease, lung disease, anemia, renal disease, dementia, history of stroke, rheumatologic diseases, obesity)
- Past healthcare use (e.g. nr of GP visits, nr of hospitalizations in the past period). The impact of COVID-19 on healthcare use should be considered when choosing the period.
- COVID-19 infection and/or vaccination (yes/no and timing)
- Use of influenza antivirals (type, timing)
- Lag time between symptom onset and testing
- Calendar time

Not all data points that can be collected through primary data collection are available in secondary data. For studies using secondary data, we suggest collecting the following data points.

- Age (stratification by age groups, minimally 6 months-14 years; 15-64 years; 65+ years)
- Gender
- Chronic underlying conditions
- Past healthcare use (e.g. nr of GP visits, nr of hospitalizations). The impact of COVID-19 on healthcare use should be considered when choosing the period.

Each study's SAP should explain how the data will be used in the analysis. Information on how to adjust for confounders in the analysis can be found in Chapter 8.

Optimization of the value of microbiological and virological information

Vaccine effectiveness studies require the identification of viruses. This should be carried out with sensitive and specific techniques and provide detailed relevant biological information about the

causative agent to avoid confounders. Consequently, this detection should use the most up-to-date diagnostic tools.

The clinical network involved in the DRIVE studies should be able to collect nasal and/or nasopharyngeal swabs and send these specimens to a corresponding virological lab.

Since evidence shows that virus shedding is significantly reduced three or four days after disease onset, information regarding the delay between disease onset and the specimen collection should be collected in studies with primary data collection, and stratification of IVE estimates according to this delay can be considered. It is recommended to exclude swabs collected more than seven days after start of symptom onset.

For studies performing primary data collection, we suggest using labs that:

- Are able to detect influenza by RT-PCR (first line of screening).
- Further characterize the detected virus by sub-typing (for Influenza A viruses) and lineage determination (Influenza B viruses).
- Have their performance assessed by participation in External Quality Assessment (EQA), as those provided by Quality Control for Molecular Diagnostics (QCMD) [124].

If possible, the lab should also be in the capacity to carry out additional influenza testing:

- Genotyping of the virus (HA and NA gene sequencing, by Sanger or NGS, for genetic clade determination, full genome sequencing should also be an objective). This can be very helpful for the comparing strains and interpreting IVE results.
- Strain characterization for the identification of potential antigenic variants. This means being able to grow influenza viruses on MDCK cells and subsequently determine their antigenic profile with ferret sera. This will allow the complete antigenic characterization of the influenza viruses, according to the WHO standards, as described by the CDC [115].

In addition, the labs may detect with the same techniques other respiratory viruses such as RSV, rhinoviruses, human metapneumoviruses, adenoviruses and parainfluenza viruses; all these viruses can co-circulate during the influenza epidemic and may be responsible for ILI presentation.

For studies using secondary data collection, the preferred lab method to test for influenza is RT-PCR.

Methods for near-related time assessment of IVE

In the future, any study design that has been proven to yield valid and reliable estimates can be chosen to rapidly assess IVE in near real time. Large test-negative case-control studies implemented on top of multiple, existing influenza surveillance/sentinel systems and large cohort studies based on automated (secondary) data collections might be the most feasible strategies. The appropriate timing

within the season depends on the actual purpose of the IVE figures and the course of the epidemic. It seems advisable to assess IVE early but after the peak of the influenza-type specific epidemic curve, since it has been shown that such studies can reliably predict even end-season IVE [133, 134]. However, in order to provide data to the WHO to support the decision on future vaccine compositions, one might also preliminarily end the study period shortly before or at the top of an intense epidemic that started late. In DRIVE, no interim analyses were performed due to a lack of statistical power for brand-specific estimates.

Data analysis for individual studies

When analyzing data of an individual VE study, the study design should be taken into account because the parameter that can be used to estimate the VE differs between designs. Confounders of the relationship between vaccination and prevention of infection should be adjusted for, which can be achieved by regression or propensity scoring. Known confounders should be included in the statistical model, regardless of their significance. If additional variables are considered potential confounders, data-driven regression or propensity score methods can be explored, but care should be taken when interpreting their results. Within DRIVE, logistic regression analyses are used for the TND studies, and Poisson regression is used for the cohort studies. All analyses are stratified by age group, as in this setting, age is generally considered both an effect modifier and a confounder. Additionally, confounder control is primarily done through regression adjustment of known confounders and sensitivity analyses performance to evaluate the impact of other potential confounding variables. Finally, in DRIVE, the regression models tend to use penalized splines to model continuous variables such as age and time of disease onset.

Pooling data from different data sources

The objective of DRIVE is to estimate brand-specific IVE in Europe by combining data from different study sites. Some site-specific studies will adopt the test-negative case-control study design, while others will use a cohort design. These different study designs imply that different data will be collected and that different statistical analyses are needed to analyze these data (i.e. logistic or conditional logistic regression for case-control versus Poisson or Cox regression for cohort studies). Given the statistical equivalence of AD-MA and IPD-MA and given the additional complexity or even impossibility of performing IPD-MA when data are collected using different study designs or only aggregated data are shared, the AD-MA is the preferred method for combining data from different study sites. Furthermore, many of the mentioned advantages of IPD-MA (i.e. transforming data to common sources or measures and standardizing analysis) can also be achieved through harmonization/standardization of the individual site-specific studies.

Within AD-MA, the random-effects meta-analysis approach is preferred to the fixed-effects approach for combining the data from the site-specific studies on IVE. The assumption that the between-study variability is explained by sampling variability only, which underlies the fixed effects meta-analysis, is not realistic for the studies on IVE in Europe. There are many differences between these studies, including differences in population, design, exposure- and disease ascertainment as well as in covariate adjustment. For DRIVE, we used a random effects meta-analysis approach for every effect measure of interest, stratified by age and clinical outcome, while at the same time standardizing the design and conduct of the site-specific studies to the extent possible.

11 References

1. I-MOVE, Marta Valenciano, Alain Moren, Bernadette Gergonne, Camelia Savulescu. Methods for measuring Influenza vaccine effectiveness during influenza seasons and pandemics in the European Union. 2008.

2. WHO. Evaluation of influenza vaccine effectiveness - A guide to the design and interpretation of observational studies. Gevena: World Health Organization; 2017.

3. Committee for Medicinal Products for Human Use. Guideline on Influenza Vaccines - Non-clinical and Clinical Module. EMA/CHMP/BWP/310834/2012. London: Eur Med Agency; 2016.

4. Pharmacovigilance Risk Assessment Committee. Interim guidance on enhanced safety surveillance for seasonal influenza vaccines in the EU. In: EMA, editor. 10 April 2014.

5. Foppa IM, Haber M, Ferdinands JM, Shay DK. The case test-negative design for studies of the effectiveness of influenza vaccine. Vaccine. 2013;31(30):3104-9.

6. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. Vaccine. 2013;31(17):2165-8.

7. Kissling E, Rondy M, team IMIMs. Early 2016/17 vaccine effectiveness estimates against influenza A(H3N2): I-MOVE multicentre case control studies at primary care and hospital levels in Europe. Euro Surveill. 2017;22(7).

8. Kissling E, Valenciano M, Pozo F, Vilcu AM, 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. 2017.

9. Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. American Journal of Epidemiology. 2016;184(5):345-53.

10. Hak E, Wei F, Grobbee DE, Nichol KL. A nested case-control study of influenza vaccination was a cost-effective alternative to a full cohort analysis. J Clin Epidemiol. 2004;57(9):875-80.

11. R.L. P. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika. 1986;73:12.

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

13. Remschmidt C, Rieck T, Bodeker B, Wichmann O. Application of the screening method to monitor influenza vaccine effectiveness among the elderly in Germany. BMC Infect Dis. 2015;15:137.

14. Fukushima W, Hirota Y. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. Vaccine. 2017;35(36):4796-800.

15. WHO. Influenza. Geneva: WHO; 2017 [Feb 5, 2018]]. Available from: http://www.who.int/biologicals/vaccines/influenza/en/

16. ECDC. Types of seasonal influenza vaccine Stockholm: ECDC; 2017 [cited May 24, 2018]. Available from: https://ecdc.europa.eu/en/seasonal-influenza/prevention-and-control/vaccines/types-of-seasonal-influenza-vaccine.
17. WHO. Influenza (seasonal) Fact sheet. Geneva: WHO; 2018 [cited Feb 5, 2018]. Available from: http://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)
18. Dos Santos G, Liyanage H, Stuurman A. D2.4 Development of guidelines for the identification of vaccine status and brand in study participants DRIVE; 2019.
19. Stuurman A, Haag M, Rizzo C. D3.1 Report on the sources for usage of specific influenza vaccine brands and accessibility. DRIVE; 2017.
20. Stuurman A, Liyanage H, Haag M. D3.3 Report on the brand availability and usage of specific influenza vaccine brands - update. DRIVE; 2021.
21. Stuurman AL, Rizzo C, Haag M. Investigating the procurement system for understanding seasonal influenza vaccine brand availability in Europe. PloS one. 2021;16(4):e0248943.
22. ENCEPP. Guide on Methodological Standards in Pharmacoepidemiology (Revision 6). EMA/95098/2010. London: Eur Med Agency; 2010 [cited November 22, 2017]. Available from: http://www.encepp.eu/standards_and_guidances/documents/ENCePPGuideofMethStandardsinPE_Rev6.pdf.
23. Kissling E, Moren A, M. V. Protocol for cohort database studies to measure pandemic and seasonal influenza vaccine effectiveness in the European Union and European Economic Area Member States. . 2009; Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0907_TER_Influenza_AH1N1_Measuring_Influenza_Vaccine_Effectiveness_Protocol_Cohort_Database_Studies.pdf. Accessed 11/20, 2017.
24. Kissling E MA, Valenciano M. . Protocol for case-control studies to measure pandemic and seasonal influenza vaccine effectiveness in the European Union and European Economic Area Member States. . 2009; Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0907_TED_Influenza_AH1N1_Measuring_Influenza_Vaccine_Effectiveness_Protocol_Case_Control_Studies.pdf. Accessed 11/22, 2017.
25. I-MOVE+. Protocol for hospital-based test negative case control studies to measure seasonal influenza vaccine effectiveness against influenza laboratory confirmed SARI hospitalisation among the elderly across the European Union and European Economic Area Member States. 2016 [cited Feb 9, 2018]. Available from: https://drive.google.com/file/d/0B54XpZN4SY65OGRSTHlvOVZTMFE/view.
26. I-MOVE+. Generic Protocol for the register based studies to measure seasonal influenza vaccine effectiveness against laboratory confirmed influenza and pneumonia among elderly in primary care.

2016 [cited Feb 9, 2018]. Available from: https://drive.google.com/file/d/0Byv9pYYPpY4PYlZ4bjk1d3pxaUE/view. .
27. Baum U, Sundman J, Jääskeläinen S, Nohynek H, Puumalainen T, Jokinen J. Establishing and maintaining the National Vaccination Register in Finland. Eurosurveillance. 2017;22(17).
28. Nauta J, Beyer WE, Kimp EP. Toward a better understanding of the relationship between influenza vaccine efficacy against specific and non-specific endpoints and vaccine efficacy against influenza infection. Epidemiology, Biostatistics and Public Health. 2017;14(4).
29. European Commission. EU case definitions. Official Journal of the European Union 2012/506/EU.
30. . !!! INVALID CITATION !!! [27-29].
31. Fitzner J, Qasmieh S, Mounts AW, Alexander B, Besselaar T, Briand S, et al. Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. Bulletin of the World Health Organization. 2018;96(2):122.
32. Military Health System. Influenza-Like Illness. . AFHSC Standard Case Definitions. 2015.
33. CDC. Guidance for Clinicians on the Use of RT-PCR and Other Molecular Assays for Diagnosis of Influenza Virus Infection 2016 [cited Feb 19, 2018]. Available from: https://www.cdc.gov/flu/professionals/diagnosis/molecular-assays.htm.
34. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, et al. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. American journal of epidemiology. 2008;167(7):775-85.
35. Ip DK, Lau LL, Chan K-H, Fang VJ, Leung GM, Peiris MJ, et al. The dynamic relationship between clinical symptomatology and viral shedding in naturally acquired seasonal and pandemic influenza virus infections. Clinical Infectious Diseases. 2015;62(4):431-7.
36. Suess T, Remschmidt C, Schink SB, Schweiger B, Heider A, Milde J, et al. Comparison of shedding characteristics of seasonal influenza virus (sub) types and influenza A (H1N1) pdm09; Germany, 2007–2011. PloS one. 2012;7(12):e51653.
37. Wang B, Russell ML, Fonseca K, Earn DJ, Horsman G, Van Caeseele P, et al. Predictors of influenza a molecular viral shedding in Hutterite communities. Influenza and other respiratory viruses. 2017;11(3):254-62.
38. Ng S, Lopez R, Kuan G, Gresh L, Balmaseda A, Harris E, et al. The timeline of influenza virus shedding in children and adults in a household transmission study of influenza in Managua, Nicaragua. The Pediatric infectious disease journal. 2016;35(5):583.
39. Fielding JE, Kelly HA, Mercer GN, Glass K. Systematic review of influenza A (H1N1) pdm09 virus shedding: duration is affected by severity, but not age. Influenza and other respiratory viruses. 2014;8(2):142-50.
40. Landry ML. Diagnostic tests for influenza infection. Current opinion in pediatrics. 2011;23(1):91-7.

41. EMA. Antiviral medicines for pandemic influenza London: Eur Med Agency; [cited Apr 30, 2018]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000463.jsp&mid=WC0b01ac058004bf56.
42. Dos Santos G, Van Boxmeer J, Levi M, Stuurman A, Turunen T. D2.2 Systematic review of the sources of confounding, bias and strategies to manage their impact in influenza vaccine effectiveness studies DRIVE; 2018.
43. Glasser SP. Bias, confounding, and effect modification (interaction). Essentials of Clinical Research: Springer; 2014. p. 363-72.
44. Chua H, Feng S, Lewnard JA, Sullivan SG, Blyth CC, Lipsitch M, et al. The Use of Test-negative Controls to Monitor Vaccine Effectiveness. Epidemiology. 2020;31(1):43-64.
45. Miquel Porta (editor). A Dictionary of Epidemiology. 6th ed. Oxford: Oxford University Press; 2014.
46. Hak E, Verheij TJ, Grobbee D, Nichol K, Hoes A. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. Journal of Epidemiology & Community Health. 2002;56(12):951-5.
47. McElhaney JE, Andrew MK, McNeil SA. Estimating influenza vaccine effectiveness: Evolution of methods to better understand effects of confounding in older adults. Vaccine. 2017.
48. Fireman B, Lee J, Lewis N, Bembom O, Van Der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. American journal of epidemiology. 2009;170(5):650-6.
49. 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. International journal of epidemiology. 2005;35(2):345-52.
50. Kahn KE, Santibanez TA, Zhai Y, Singleton JA. Influenza vaccination type, live, attenuated influenza vaccine (LAIV) versus inactivated influenza vaccine (IIV), received by children, United States, 2011– 12 through 2013– 14 influenza seasons. Vaccine. 2015;33(39):5196-203.
51. Talbot HK, Nian H, Zhu Y, Chen Q, Williams JV, Griffin MR. Clinical effectiveness of split-virion versus subunit trivalent influenza vaccines in older adults. Clinical Infectious Diseases. 2015;60(8):1170-5.
52. Zimmerman RK, Nowalk MP, Chung J, Jackson ML, Jackson LA, Petrie JG, et al. 2014–2015 influenza vaccine effectiveness in the United States by vaccine type. Clinical Infectious Diseases. 2016:ciw635.
53. Stuurman AL, Ciampini S, Vannacci A, Bella A, Rizzo C, Muñoz-Quiles C, et al. Factors driving choices between types and brands of influenza vaccines in general practice in Austria, Italy, Spain and the UK. Plos one. 2021;16(6):e0252836.
54. Stuurman A, Biccler J, Carmona A, Levi M, Descamps A. D7.8 Brand-specific influenza vaccine effectiveness in Europe Season 2020/21. DRIVE consortium; 2021.

55. De Smedt T, Merrall E, Macina D, Perez-Vilar S, Andrews N, Bollaerts K. Bias due to differential and non-differential disease-and exposure misclassification in studies of vaccine effectiveness. PloS one. 2018;13(6):e0199180.
56. Endo A, Funk S, Kucharski AJ. Bias correction methods for test-negative designs in the presence of misclassification. Epidemiology & Infection. 2020;148.
57. Mereckiene J. Seasonal influenza vaccination in Europe, 2007–2008 to 2014–2015. Stockholm: ECDC; 2017.
58. ECDC. Factsheet about seasonal influenza Stockholm 2017 [Available from: https://ecdc.europa.eu/en/seasonal-influenza/facts/factsheet.
59. Jiménez-García R, Hernández-Barrera V, de Andres AL, Jimenez-Trujillo I, Esteban-Hernández J, Carrasco-Garrido P. Gender influence in influenza vaccine uptake in Spain: time trends analysis (1995–2006). Vaccine. 2010;28(38):6169-75.
60. CDC. Pregnant Women & Influenza. 2017 [cited May 28, 2018]. Available from: https://www.cdc.gov/flu/protect/vaccine/pregnant.htm.
61. Wilson KM, Pier JC, Wesgate SC, Cohen JM, Blumkin AK. Secondhand tobacco smoke exposure and severity of influenza in hospitalized children. The Journal of pediatrics. 2013;162(1):16-21.
62. Foppa IM, Ferdinands JM, Chung J, Flannery B, Fry AM. Vaccination history as a confounder of studies of influenza vaccine effectiveness. Vaccine: X. 2019;1:100008.
63. Ramsay LC, Buchan SA, Stirling RG, Cowling BJ, Feng S, Kwong JC, et al. The impact of repeated vaccination on influenza vaccine effectiveness: a systematic review and meta-analysis. BMC Med. 2017;15(1):159.
64. Smith DJ, Forrest S, Ackley DH, Perelson AS. Variable efficacy of repeated annual influenza vaccination. Proceedings of the National Academy of Sciences. 1999;96(24):14001-6.
65. Belongia EA, Skowronski DM, McLean HQ, Chambers C, Sundaram ME, De Serres G. Repeated annual influenza vaccination and vaccine effectiveness: review of evidence. Expert review of vaccines. 2017;16(7):723-36.
66. Gherasim A, Martínez-Baz I, Castilla J, Pozo F, Larrauri A, group cw. Effect of previous and current vaccination against influenza A (H1N1) pdm09, A (H3N2), and B during the post-pandemic period 2010-2016 in Spain. PloS one. 2017;12(6):e0179160.
67. Belongia EA, Simpson MD, King JP, Sundaram ME, Kelley NS, Osterholm MT, et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. The Lancet Infectious Diseases. 2016;16(8):942-51.
68. Lane C, Carville KS, Pierse N, Kelly H. Seasonal influenza vaccine effectiveness estimates: Development of a parsimonious case test negative model using a causal approach. Vaccine. 2016;34(8):1070-6.

69. Saito N, Komori K, Suzuki M, Morimoto K, Kishikawa T, Yasaka T, et al. Negative impact of prior influenza vaccination on current influenza vaccination among people infected and not infected in prior season: A test-negative case-control study in Japan. Vaccine. 2017;35(4):687-93.
70. Cascini F, Pantovic A, Al-Ajlouni Y, Failla G, Ricciardi W. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review. EClinicalMedicine. 2021;40:101113.
71. Doll MK, Pettigrew SM, Ma J, Verma A. Effects of Confounding Bias in COVID-19 and Influenza Vaccine Effectiveness Test-Negative Designs Due to Correlated Influenza and COVID-19 Vaccination Behaviors. medRxiv. 2021.
72. McLean HQ, Chow BD, VanWormer JJ, King JP, Belongia EA. Effect of statin use on influenza vaccine effectiveness. The Journal of infectious diseases. 2016;214(8):1150-8.
73. Black S, Nicolay U, Del Giudice G, Rappuoli R. Influence of statins on influenza vaccine response in elderly individuals. The Journal of infectious diseases. 2015;213(8):1224-8.
74. Omer SB, Phadke VK, Bednarczyk RA, Chamberlain AT, Brosseau JL, Orenstein WA. Impact of statins on influenza vaccine effectiveness against medically attended acute respiratory illness. The Journal of infectious diseases. 2015;213(8):1216-23.
75. Tleyjeh IM, Kashour T, Hakim FA, Zimmerman VA, Erwin PJ, Sutton AJ, et al. Statins for the prevention and treatment of infections: a systematic review and meta-analysis. Archives of internal medicine. 2009;169(18):1658-67.
76. 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? Eurosurveillance. 2016;21(16).
77. Pebody R, Andrews N, McMenamin J, Durnall H, Ellis J, Thompson C, et al. Vaccine effectiveness of 2011/12 trivalent seasonal influenza vaccine in preventing laboratory-confirmed influenza in primary care in the United Kingdom: evidence of waning intra-seasonal protection. Eurosurveillance. 2013;18(5):20389.
78. Song JY, Cheong HJ, Hwang IS, Choi WS, Jo YM, Park DW, et al. Long-term immunogenicity of influenza vaccine among the elderly: risk factors for poor immune response and persistence. Vaccine. 2010;28(23):3929-35.
79. Puig-Barberà J, Mira-Iglesias A, Tortajada-Girbés M, López-Labrador F, Librero-López J, Díez-Domingo J, et al. Waning protection of influenza vaccination during four influenza seasons, 2011/2012 to 2014/2015. Vaccine. 2017;35(43):5799-807.
80. Lipsitch M, Goldstein E, Ray GT, Fireman B. Depletion-of-susceptibles bias in influenza vaccine waning studies: how to ensure robust results. Epidemiology & Infection. 2019;147.

81. Ray GT, Lewis N, Klein NP, Daley MF, Lipsitch M, Fireman B. Depletion-of-susceptibles bias in analyses of intra-season waning of influenza vaccine effectiveness. Clinical Infectious Diseases. 2020;70(7):1484-6.
82. Tamerius JD, Shaman J, Alonso WJ, Bloom-Feshbach K, Uejio CK, Comrie A, et al. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. PLoS pathogens. 2013;9(3):e1003194.
83. Jaakkola K, Saukkoriipi A, Jokelainen J, Juvonen R, Kauppila J, Vainio O, et al. Decline in temperature and humidity increases the occurrence of influenza in cold climate. Environmental Health. 2014;13(1):22.
84. Gomez-Barroso D, León-Gómez I, Delgado-Sanz C, Larrauri A. Climatic Factors and Influenza Transmission, Spain, 2010–2015. International journal of environmental research and public health. 2017;14(12):1469.
85. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS pathogens. 2007;3(10):e151.
86. Jefferson T, Di Pietrantonj C, Al-Ansary LA, Ferroni E, Thorning S, Thomas RE. Vaccines for preventing influenza in the elderly. The Cochrane Library. 2010.
87. Tricco AC, Chit A, Soobiah C, Hallett D, Meier G, Chen MH, et al. Comparing influenza vaccine efficacy against mismatched and matched strains: a systematic review and meta-analysis. BMC medicine. 2013;11(1):153.
88. Haq K, McElhaney JE. Immunosenescence: influenza vaccination and the elderly. Current opinion in immunology. 2014;29:38-42.
89. Bosaeed M, Kumar D. Seasonal influenza vaccine in immunocompromised persons. Human vaccines & immunotherapeutics. 2018:1-12.
90. Nichol KL, Nordin JD, Nelson DB, Mullooly JP, Hak E. Effectiveness of influenza vaccine in the community-dwelling elderly. New England Journal of Medicine. 2007;357(14):1373-81.
91. Vila-Córcoles A, Rodriguez T, de Diego C, Ochoa O, Valdivieso A, Salsench E, et al. Effect of influenza vaccine status on winter mortality in Spanish community-dwelling elderly people during 2002–2005 influenza periods. Vaccine. 2007;25(37-38):6699-707.
92. Chambers C, Skowronski DM, Rose C, Serres GD, Winter A-L, Dickinson JA, et al., editors. Should sex be considered an effect modifier in the evaluation of influenza vaccine effectiveness? Open forum infectious diseases; 2018: Oxford University Press US.
93. Havers FP, Chung JR, Belongia EA, McLean HQ, Gaglani M, Murthy K, et al. Influenza vaccine effectiveness and statin use among adults in the United States, 2011–2017. Clinical Infectious Diseases. 2019;68(10):1616-22.

94. McGrath LJ, Ellis AR, Brookhart MA. Controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. American journal of epidemiology. 2015;182(1):17-25.
95. Lipsitch M, Jha A, Simonsen L. Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. International journal of epidemiology. 2016;45(6):2060-74.
96. Jackson ML, Phillips CH, Benoit J, Kiniry E, Madziwa L, Nelson JC, et al. The impact of selection bias on vaccine effectiveness estimates from test-negative studies. Vaccine. 2017.
97. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. Vaccine. 2015;33(11):1313-6.
98. Heikkinen T, Marttila J, Salmi AA, Ruuskanen O. Nasal swab versus nasopharyngeal aspirate for isolation of respiratory viruses. Journal of clinical microbiology. 2002;40(11):4337-9.
99. Sullivan SG, Feng S, Cowling BJ. Potential of the test-negative design for measuring influenza vaccine effectiveness: a systematic review. Expert review of vaccines. 2014;13(12):1571-91.
100. Kuchipudi SV, Tellabati M, Nelli RK, White GA, Perez BB, Sebastian S, et al. 18S rRNA is a reliable normalisation gene for real time PCR based on influenza virus infected cells. Virology journal. 2012;9(1):230.
101. Chartrand C, Leeflang MM, Minion J, Brewer T, Pai M. Accuracy of rapid influenza diagnostic tests: a meta-analysis. Annals of internal medicine. 2012;156(7):500-11.
102. 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;36(3):623-31.
103. Ferdinands JM, Shay DK. Magnitude of potential biases in a simulated case-control study of the effectiveness of influenza vaccination. Clinical infectious diseases. 2011;54(1):25-32.
104. Tanei M, Yokokawa H, Murai K, Sakamoto R, Amari Y, Boku S, et al. Factors influencing the diagnostic accuracy of the rapid influenza antigen detection test (RIADT): a cross-sectional study. BMJ open. 2014;4(1):e003885.
105. Ganzenmueller T, Kluba J, Hilfrich B, Puppe W, Verhagen W, Heim A, et al. Comparison of the performance of direct fluorescent antibody staining, a point-of-care rapid antigen test and virus isolation with that of RT-PCR for the detection of novel 2009 influenza A (H1N1) virus in respiratory specimens. Journal of medical microbiology. 2010;59(6):713-7.
106. Weinberg GA, Erdman DD, Edwards KM, Hall CB, Walker FJ, Griffin MR, et al. Superiority of reverse-transcription polymerase chain reaction to conventional viral culture in the diagnosis of acute respiratory tract infections in children. The Journal of infectious diseases. 2004;189(4):706-10.
107. Vemula SV, Zhao J, Liu J, Wang X, Biswas S, Hewlett I. Current approaches for diagnosis of influenza virus infections in humans. Viruses. 2016;8(4):96.

108. Ginocchio CC, Zhang F, Manji R, Arora S, Bornfreund M, Falk L, et al. Evaluation of multiple test methods for the detection of the novel 2009 influenza A (H1N1) during the New York City outbreak. Journal of Clinical Virology. 2009;45(3):191-5.
109. Scheuller HS, Lott L, Haas R, Tavish M, Danaher P. Comparison of diagnostic test performance in a population of high risk young adults versus a general population presenting with influenza. Journal of Clinical Virology. 2015;68:49-52.
110. Pollock NR, Duong S, Cheng A, Han LL, Smole S, Kirby JE. Ruling out novel H1N1 influenza virus infection with direct fluorescent antigen testing. Clinical Infectious Diseases. 2009;49(6):e66-e8.
111. WHO. WHO recommendations on the use of rapid testing for influenza diagnosis. Geneva: WHO; 2005 [updated July 2005]. Available from: http://www.who.int/influenza/resources/documents/RapidTestInfluenza_WebVersion.pdf?ua=1.
112. Sikkema RS, Freidl GS, de Bruin E, Koopmans M. Weighing serological evidence of human exposure to animal influenza viruses− a literature review. Eurosurveillance. 2016;21(44).
113. Zhang P, Vemula SV, Zhao J, Du B, Mohan H, Liu J, et al. A highly sensitive europium nanoparticle-based immunoassay for detection of influenza A/B virus antigen in clinical specimens. Journal of clinical microbiology. 2014;52(12):4385-7.
114. Loens K, Van Loon A, Coenjaerts F, Van Aarle Y, Goossens H, Wallace P, et al. Performance of different mono-and multiplex nucleic acid amplification tests on a multipathogen external quality assessment panel. Journal of clinical microbiology. 2012;50(3):977-87.
115. Bakerman P, Balasuriya L, Fried O, Tellez D, Garcia-Filion P, Dalton H. Direct fluorescent-antibody testing followed by culture for diagnosis of 2009 H1N1 influenza A. Journal of clinical microbiology. 2011;49(10):3673-4.
116. Hindiyeh M, Levy V, Azar R, Varsano N, Regev L, Shalev Y, et al. Evaluation of a multiplex real-time reverse transcriptase PCR assay for detection and differentiation of influenza viruses A and B during the 2001-2002 influenza season in Israel. Journal of clinical microbiology. 2005;43(2):589-95.
117. Kenmoe S, Tchendjou P, Moyo Tetang S, Mossus T, Njankouo Ripa M, Guillet M, et al. Evaluating the performance of a rapid antigen test for the detection of influenza virus in clinical specimens from children in Cameroon. Influenza and other respiratory viruses. 2014;8(2):131-4.
118. Landry ML, Ferguson D. Comparison of Simplexa Flu A/B & RSV PCR with cytospin-immunofluorescence and laboratory-developed TaqMan PCR in predominantly adult hospitalized patients. Journal of clinical microbiology. 2014;52(8):3057-9.
119. Leonardi GP, Mitrache I, Pigal A, Freedman L. Public hospital-based laboratory experience during an outbreak of pandemic influenza A (H1N1) virus infections. Journal of clinical microbiology. 2010;48(4):1189-94.

120. Reina J, Plasencia V, Leyes M, Nicolau A, Galmés A, Arbona G. Estudio comparativo entre una técnica de reacción en cadena de la polimerasa en transcripción reversa en tiempo real, un método de enzimoinmunoanálisis y el cultivo shell-vial en la detección de virus gripales A y B en pacientes adultos. Enfermedades Infecciosas y Microbiología Clínica. 2010;28(2):95-8.
121. Tuuminen T, Suomala P, Koskinen JO. Evaluation of the automated multianalyte point-of-care mariPOC® test for the detection of influenza A virus and respiratory syncytial virus. Journal of medical virology. 2013;85(9):1598-601.
122. Uyeki TM, Prasad R, Vukotich C, Stebbins S, Rinaldo CR, Ferng Y-h, et al. Low sensitivity of rapid diagnostic test for influenza. Clinical Infectious Diseases. 2009;48(9):e89-e92.
123. WHO. Influenza: WHO recommendations on the composition of influenza virus vaccines Geneva: WHO; 2017 [cited December 12, 2017]. Available from: http://www.who.int/influenza/vaccines/virus/recommendations/en/.
124. CDC. Influenza Antigenic Characterization. Atlanta: CDC; 2017 [updated Sept 27, 2017]. Available from: https://www.cdc.gov/flu/professionals/laboratory/antigenic.htm.
125. Dos Santos G, Neumeier E, Bekkat-Berkani R. Influenza: Can we cope better with the unpredictable? Human vaccines & immunotherapeutics. 2016;12(3):699-708.
126. Castilla J, Navascués A, Casado I, Díaz-González J, Pérez-García A, Fernandino L, et al. Combined effectiveness of prior and current season influenza vaccination in northern Spain: 2016/17 mid-season analysis. Eurosurveillance. 2017;22(7).
127. Flannery B, Chung JR, Thaker SN, Monto AS, Martin ET, Belongia EA, et al. Interim Estimates of 2016-17 Seasonal Influenza Vaccine Effectiveness-United States, February 2017. Morbidity and Mortality Weekly Report. 2017;66(6):167-71.
128. Hergens M-P, Baum U, Brytting M, Ikonen N, Haveri A, Wiman Å, et al. Mid-season real-time estimates of seasonal influenza vaccine effectiveness in persons 65 years and older in register-based surveillance, Stockholm County, Sweden, and Finland, January 2017. Eurosurveillance. 2017;22(8).
129. Noh JY, Lim S, Song JY, Choi WS, Jeong HW, Heo JY, et al. Interim estimates of the effectiveness of the influenza vaccine against A (H3N2) influenza in adults in South Korea, 2016–2017 season. PLoS One. 2017;12(5):e0178010.
130. Skowronski DM, Chambers C, Sabaiduc S, Dickinson JA, Winter A-L, De Serres G, et al. Interim estimates of 2016/17 vaccine effectiveness against influenza A (H3N2), Canada, January 2017. Eurosurveillance. 2017;22(6).
131. Sullivan SG, Chilver MB, Carville KS, Deng Y-M, Grant KA, Higgins G, et al. Low interim influenza vaccine effectiveness, Australia, 1 May to 24 September 2017. Eurosurveillance. 2017;22(43).
132. WHO. WHO consultation and information meeting on the composition of influenza virus vaccines for use in the 2017-2018 northern hemisphere influenza season. Geneva: WHO; 2017 [updated March 2,

2017; cited Feb 5, 2018]. Available from: http://www.who.int/influenza/vaccines/virus/recommendations/2017_18_north/en/.
133. Leung VK, Cowling BJ, Feng S, Sullivan SG. Concordance of interim and final estimates of influenza vaccine effectiveness: a systematic review. Eurosurveillance. 2016;21(16):30202.
134. Sullivan S, Kelly H. Late season interim estimates of influenza vaccine effectiveness reliably predict end of season estimates in Victoria, Australia, 2007 to 2012. Eurosurveillance. 2013;18(41):20605.
135. Halloran ME, Longini IM, Struchiner CJ, Longini IM. Design and analysis of vaccine studies: Springer; 2010.
136. Sato T. Maximum likelihood estimation of the risk ratio in case-cohort studies. Biometrics. 1992:1215-21.
137. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika. 1986;73(1):1-11.
138. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 2008.
139. Holford TR. The analysis of rates and of survivorship using log-linear models. Biometrics. 1980:299-305.
140. Aalen O, Borgan O, Gjessing H. Survival and event history analysis: a process point of view: Springer Science & Business Media; 2008.
141. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine: Wiley New York; 1985.
142. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables: John Wiley & Sons; 2008.
143. Harrell Jr. FE. Regression modelling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. . 2nd ed. Cham: Springer; 2015.
144. Groenwold RH, Klungel OH, Altman DG, van der Graaf Y, Hoes AW, Moons KG. Adjustment for continuous confounders: an example of how to prevent residual confounding. Canadian Medical Association Journal. 2013;185(5):401-6.
145. Lee PH. Is a cutoff of 10% appropriate for the change-in-estimate criterion of confounder identification? Journal of epidemiology. 2014:JE20130062.
146. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Statistics in medicine. 2002;21(16):2409-19.
147. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research. 2011;46(3):399-424.
148. Simpson C, Lone N, Kavanagh K, Ritchie L, Robertson C, Sheikh A, et al. Seasonal Influenza Vaccine Effectiveness (SIVE): an observational retrospective cohort study–exploitation of a unique community-

based national-linked database to determine the effectiveness of the seasonal trivalent influenza vaccine. 2013.
149. Molenberghs G, Kenward M. Missing data in clinical studies: John Wiley & Sons; 2007.
150. O'Kelly M, Ratitch B. Clinical trials with missing data: A guide for practitioners: John Wiley & Sons; 2014.
151. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC medical research methodology. 2010;10(1):7.
152. O'Kelly M., B. R. Clinical trials with mossing data: A Guide for Practitioners. Chichester: John Wiley and Sons; 2014.
153. Bollaerts K, Weil J, the WP4 working group. D4 . 4 Report on tested methods for accelerated assessment of vaccination coverage , vaccine benefits , V1 Final.: Accelerated Development of VAccine beNefit-risk Collaboration in Europe Grant Agreement no115557 2017.
154. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. BMC Med Res Methodol. 2005;5:14.
155. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97-111.
156. Osterholm MT, Kelley NS, Sommer A, Belongia EA. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. Lancet Infect Dis. 2012;12(1):36-44.
157. Rondy M, El Omeiri N, Thompson MG, Leveque A, Moren A, Sullivan SG. Effectiveness of influenza vaccines in preventing severe influenza illness among adults: A systematic review and meta-analysis of test-negative design case-control studies. J Infect. 2017;75(5):381-94.
158. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015;313(16):1657-65.
159. Abo-Zaid G, Guo B, Deeks JJ, Debray TP, Steyerberg EW, Moons KG, et al. Individual participant data meta-analyses should not ignore clustering. J Clin Epidemiol. 2013;66(8):865-73 e4.
160. Debray TP, Moons KG, Abo-Zaid GM, Koffijberg H, Riley RD. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? PLoS One. 2013;8(4):e60650.
161. Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. Biometrics. 1999;55(4):3.
162. Mathew T, Nordstrom K. Comparison of one-step and two-step meta-analysis models using individual patient data. Biometrical Journal. 2010;52(2):17.
163. Olkin I, Sampson A. Comparison of meta-analysis versis analysis of variance of individual patient data. Biometrics. 1998;54(1):5.

164. Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. PLoS One. 2012;7(10):e46042.
165. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. Stat Med. 2017;36(5):855-75.
166. Stuurman A, Riera M, K B, Alexandridou M, Turunen T, Rizzo C, et al. D7.4 Setting up brand-specific influenza vaccine effectiveness studies in Europe – results of the pilot season 2017/18. DRIVE; 2018.
167. Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. J Clin Epidemiol. 2007;60(5):431-9.