Integrating virological data in influenza vaccine effectiveness studies: I-MOVE primary care network laboratory protocol

Last updated: November 2019

Table of content

Acronyms 3

1. Background 4

3. Methods 4

3.1 Study design 4

3.3 Definitions ILI patients, influenza cases and controls 5

3.4 Study period 5

3.5. Sample size 5

3.6. Virus selection for genetic characterisation 6

3.6.1 Proportion of influenza viruses to characterise: sampling fraction 6

3.6.2 Procedures for random selection of specimens to be characterised 7

3.7. Genetic characterisation 8

3.8 Analysis 8

3.8.1. Description of antigenic/genetic mutations of viruses included in the MCCS 8

3.8.2. Clade/genetic variant-specific vaccine effectiveness 10

3.9 Data collected 10

3.10 Data management 11

5. Limitations 11

6. References 11

Annex I Sample size calculation and estimation of sampling fraction 13

Estimating the percentage of influenza laboratory confirmed specimens to characterise to obtain sufficient precision in clade/genetic variant-specific VE estimates (Objective 2) 14

Annex II: GISAID data collection form 16

Annex III: Sampling fraction collection form 18

Acronyms

CI	Confidence Interval
HA	Haemagglutinin
ECDC	European Centre for Disease Prevention and Control
GISAID	Global Initiative on Sharing All Influenza Data
GP	General Practitioner
ILI	Influenza-Like Illness
I-MOVE	Influenza Monitoring Vaccine Effectiveness
I-MOVE+	Integrated Monitoring of Vaccine Effects in Europe
MCCS	Multicentre case-control study
TND	Test-Negative Design
VC	Vaccination Coverage
VE	Vaccine Effectiveness
WHO	World Health Organization

1. Background

Since 2008–9, the I-MOVE multicentre case-control primary care study (MCCS) provides annual estimates of influenza vaccine effectiveness (VE) by (sub)type. Between six and twelve sites have participated in the pooled analysis each season. (Sub)type-specific VE estimates often vary by site. A different distribution by clade or genetic variants by site may explain these differences.

In some seasons, the MCCS results have suggested a decreased in VE over time (1). This decrease may be due to virus changes or to the waning of the immunity conferred by the vaccine. Also, differences in age-specific VE to specific clades/genetic variants, potentially related to childhood imprinting have been reported (2,3).

The integration of virological data in the MCCS is essential to interpret VE results, particularly when low VE against a specific clade is suspected or when several clades are circulating. The distribution of clades by study site and by influenza period may help to understand the differences between site-specific VE and the changes in VE over the season.

In 2013–14, some study sites participating in the MCCS started to provide genetic and antigenic results from the specimens of a sample of patients included in the study. The samples are selected to address the objectives of the national virological surveillance system in which the studies are embedded. The priority is to characterise specimens from different age groups, settings, geographical locations and different phases of progression of influenza activity. A subset of specimens is obtained from severe cases and vaccinated cases. Therefore, the specimens characterised are not representative of the influenza cases included in the I-MOVE MCCS in terms of distribution of clades over time, vaccination status, chronic diseases, age group, etc.

Since 2015–16, a systematic approach for genetic analyses is carried out and outlined in this protocol. This laboratory protocol complements the I-MOVE generic protocol *Test Negative Design case control studies to measure pandemic and seasonal influenza vaccine effectiveness in the European Union and European Economic Area Member States*.

The aim is to define how to select a representative sample of I-MOVE MCCS specimens in order to describe genetically the viruses and to identify amino acid changes that can influence the effect of the vaccines. Ideally, the sample size should be large enough to measure a pooled influenza VE by clade and genetic variants.

2. Objectives

- **Objective 1**
  To describe, for each influenza (sub)type/lineage the viruses included in the I-MOVE MCCS (overall, by site and by influenza period) to identify key influenza virus genotypic evolutions that could affect vaccine effectiveness.

- **Objective 2**
  To measure clade/genetic variant-specific influenza VE among I-MOVE sites participating in the pilot season.

3. Methods

3.1 Study design

- **Objective 1**
  Descriptive study of viruses included in the MCCS (mutations identified, by clades, time, site, and vaccination status).

- **Objective 2**
  Multicentre test negative design (TND) case-control study using the methods described in the I-MOVE case-control generic protocol (4).

3.3 Definitions ILI patients, influenza cases and controls

The ILI case definition used, and the procedures to identify influenza cases and controls are ones defined in the I-MOVE generic protocols (4).

We define a clade/genetic variant-specific influenza case as an ILI case positive for the specific clade/ genetic variant and a control as an ILI patient negative for any influenza virus.

3.4 Study period

For each influenza (sub)type/lineage, the National Influenza Centres will characterise strains from viruses of influenza cases recruited during the period of inclusion in the I-MOVE MCCS (from ISO week of symptom onset of first influenza positive case detected in the season up to the ISO week of symptom onset of the last positive case detected).

The proportion of strains characterised will vary according to the phase in the influenza season: the proportion will be higher in the early and late phase and lower in the peak of the season. Each site will define the start/end of each influenza phase according to the surveillance indicators used in their national surveillance system. Depending on the resources available, each site will define the proportion of strains to be characterised (sampling fraction) for each of the phases. The minimum number of strains to be characterised is described in the paragraph "sample size". Effort will be made to sequence all viruses in the early phase of the season, particularly if the circulating viruses are likely to have genetic variation.

In the pooled analysis to measure clade/genetic variant-specific VE, the study period begins during the ISO week of symptom onset of the first influenza case of that clade/genetic variant to be analysed up to the ISO week of symptom onset of the last influenza case of that clade/genetic variant to be analysed, by study site.

- *Each site to*
  - *define the influenza phases;*
  - *define the sampling fraction for each phase;*
  - *document the sampling fraction for each phase and for each subtype;*
  - *take the sampling fraction for each period and for each subtype/lineage into account if measuring clade/genetic variant specific VE for the site*

- *Epiconcept to include in the pooled clade /genetic variant-specific VE analysis, the sampling fraction by subtype, site and phase.*

3.5. Sample size

It is difficult to define the minimum sample size needed for estimating clade/genetic variant-specific VE. The minimum sample size will depend on the vaccination coverage among controls, the clade/genetic variant-specific VE and the proportion of controls. In Annex I we present the sample size for different values of these parameters.

The study sites will try to characterise the maximum number of viruses taking into account the existing resources. Each site will define the proportion of viruses to characterise. We recommend to have a higher sampling fraction in the early phase (if possible, characterise all strains, particularly if there is variation in genetic variants circulating) of the influenza season to better understand what is circulating

and to be able to provide results for the WHO vaccine strain selection committee that meets end of February each year.

3.6. Virus selection for genetic characterisation

Study sites will select viruses from ILI patients included in the MCCS testing positive for influenza. When feasible, before virus selection, study sites should verify if the ILI positive cases meet the criteria used to include cases in the MCCS pooled analysis (e.g. European Union ILI case definition, vaccination status and date documented, delay symptom onset swabbing < 8 days, etc.). Epiconcept can share with interested study sites, the algorithm and Stata commands used to exclude cases excluded from the pooled analysis.

3.6.1 Proportion of influenza viruses to characterise: sampling fraction

Each site has different resources, different (sub)type-specific viruses circulating, different proportion of clade/genetic variant viruses circulating. If a study site is not able to genetically characterize all viruses, then the sampling fraction will be based on the study site resources and on the epidemiological/virological situation. In Annex I we present the formula to compute the sampling fraction given a minimum sample size for clade-/genetic variant-specific VE calculation.

- *Each site to*
    - *define the sampling fraction used for each subtype/lineage (if all viruses characterized, then indicate 100%);*
    - *document the sampling fraction for each period and for each subtype;*
    - *take the sampling fraction for each subtype/lineage into account when measuring clade-/genetic variant-specific VE for the site (if applicable).*

- *Epiconcept to:*
    - *provide parameters for estimating sampling fraction for the pooled analysis;*
    - *include the sampling fraction by site, subtype and phase in the analysis measuring pooled clade/genetic variant-specific VE.*

Note that the sampling fraction defines the minimum number of viruses to be characterised. Study sites are welcome to characterise a higher sampling fraction, as long as the random process is adhered to (see section 3.6.2) and the sampling fraction documented.

The sample fractions can be defined for at least three influenza phases (early, peak, late). Study sites can chose more phases with specific sampling fractions in the season.

Each site will fix the periodicity to define the specimens to select (weekly, monthly, in each of the influenza phases).

- *Each site to define the periodicity to do the selection of strains to characterise and to document it*
    - *the time period for which the selection is done (e.g. every two weeks, monthly, at the end of each influenza phase);*
    - *the sampling fraction used in each of the periods*

The dates used to define the time periods will be the dates used in each of the national surveillance systems (e.g. date of onset, date of swabbing, date of notification).

3.6.2 Procedures for random selection of specimens to be characterised

If a study site is not genetically characterisng all viruses, then the random selection proposed should be used to select the viruses included in the MCCS (this does not apply to routine virological surveillance).

At the end of each time period defined for the selection of strains, study sites will select viruses using the Bernoulli sampling method. This method ensures that each strain has the same probability of being selected.

- *Each site to define who selects the strains (e.g. team of epidemiologists, team of virologists)*

Steps to randomly select the strains:

1st/ Sampling frame: for each subtype/lineage, create a list of all I-MOVE cases recruited for a given period with the selected influenza phase:

- I-MOVE strains already characterised by National Influenza Center during that period will be part of the sampling frame;
- strains with low viral load will be part of the sampling frame;
- if possible, sites will exclude from the sampling frame the strains from cases that would later be excluded from the I-MOVE pooled analysis (e.g. cases not meeting the European Union case definition, delay between onset and swab > 7 days, vaccination status missing).

2nd/ For each subtype/lineage list

a) order all positive cases by onset date;

b) assign a random number to each case and order the list by random number

The Excel function =RAND() can be used

[this may be different if using a different language version of Excel],

copy-paste the random number column and then sort list of cases by random number (e.g. in order of high to low)).

In Stata, you can use the syntax of the example below:

Selection of 10% of 440 cases (44 strains to characterise).

```
sort ID number
set seed 500
gen naleat=runiform(),
sort naleat
gen select=0
replace select=1 if _n<=44
list IDnumber strain select if select==1, noobs separator(44)
```

where

onset date is the date of symptoms onset,

500 is the number used to set the seed. You can select another number but it is recommended to use the “set seed“ to be able to replicate the selection

-IDnumber is the identifcation number of the case

c) Based on the sampling fraction defined for the period in a phase, select the number of cases needed: start from the first case in the list and continue

selecting the following cases until reaching the desired number (e.g. if 20 cases have been recruited in the study and the sampling fraction is 0.5, the first 10 cases in the list will be selected)

3rd/ Replacement of strains randomly selected but not characterised

Strains that cannot be characterised should be replaced using the same subtype/lineage sampling frame. The strains will be replaced by the next ones in the list. The reasons for not characterising selected specimens will be documented (e.g. low viral loads) and all study sites should document their CT threshold for sequencing (if applicable).

4th/ Increase sampling fraction if needed

If during the season, the study site decides to increase the sampling fraction for one subtype or time period (e.g. I-MOVE group decide to focus on one subtype and increase the sampling fraction), the study site should go back to the sampling frame for that subtype and time period and continue selecting the following strains from the ordered original list.

3.7. Genetic characterisation

The National Influenza Centres of the countries participating in the study will characterise the viruses. Genetic characterisation will be done by sequencing at least the hemagglutinin (HA or HA1) coding region directly on clinical samples or in cell supernatants after virus isolation in MDCK/MDCK Siat-1 cells. If possible, clinical samples should be used for sequencing even if they have a lower viral load. Viral supernatants have higher viral loads but mutations acquired by the cell-growth adaptation may be present.

Standard laboratory software programmes are used for sequence assembly, sequence alignment and virus phylogeny reconstruction. Genetic clades are determined and deduced amino acid substitutions are defined in relation to vaccine and reference strains.

3.8 Analysis

3.8.1. Description of antigenic/genetic mutations of viruses included in the MCCS

The number (and proportion) of viruses selected but excluded because they could not be characterised and the reasons for exclusion will be presented. Cases eligible, selected but not characterised will be described (age, vaccination status, date of onset). The proportion of primary specimens and isolates among sequenced viruses.

For each subtype/lineage, the following indicators will be computed

- number and proportion of viruses selected for characterisation;
- number and proportion of viruses sequenced;
- number and proportion of viruses by clades;
- distribution of clades by study site and for the pooled dataset ;
- distribution of clades by influenza phase.

Example of table: Influenza subtype/lineage viruses characterised by clade and study site. I-MOVE multicentre case control study, influenza season XXX

Characterised viruses	Clade	Site 1 n (%)	Site 2 n (%)	Site 3 n (%)	Total n (%)

A(H3N2)
A/HongKong/5738/2014
A/Switzerland/9715293/2013
A/Samara/73/2013
A/Newcastle/22/2014
Total A(H3N2)
A(H1N1)pdm09
A/SouthAfrica/3626/2013
B/Yamagata
B/Phuket/3073/2013
B/Massachusetts/02/2012
Total B/Yamagata
B/Victoria
B/Brisbane/60/2008

Example of table: Distribution of Clade/genetic variant XX by site and influenza phase. I-MOVE multicentre case control study, influenza season XXX

	Wk 40-52 Oct-Dec	**Wk 1-5** Jan	**Wk 6-9** Feb	**Wk 10-20** Mar-May
I-MOVE				
Site 1				
Site 2				
Site 3				
Site 4				

We will compare the sequences to the corresponding reassortant viruses used in the vaccines. We will describe amino acid substitutions.

We will use the neighbour-joining method and Kimura 2-parameter nucleotide substitution model for phylogenetic analysis.

Virologists from the Portuguese and Spanish MCCS team will summarise the virological results.

3.8.2. Clade/genetic variant-specific vaccine effectiveness

If sample size allows, Epiconcept will compute a pooled clade/genetic variant-specific VE and 95% CI overall and by age group following the methods described in the generic I-MOVE case-control study. The computation will take into account the sampling fractions.

3.9 Data collected

For each strain characterised, at least the following information should be available *(Annex II GISAID data collection form):*

- Country
- I-MOVE ID number
- HAISD (GISAID sequence database accession number)
- Selected for characterisation? (Y/N)
- Reasons for not characterising?
- CT value
- Type of sample (primary specimen or isolate)

From the HAISD number, the virologist in charge of the summary will extract the sequence information from GISAID[1].

For each time period of selection, study sites should document the start/end of the time period and the sampling fraction *(Annex III: Sampling fraction collection form)*

[1] Y. Shu and J. McCauley, 'GISAID: Global initiative on sharing all influenza data – from vision to reality', *Eurosurveillance*, vol. 22, no. 13, p. 30494, Mar. 2017.

3.10 Data management

We present here the data management related to the genetic characterisation information *(note that the data transfer and management of the clinical and epidemiological pooled database is described in the I-MOVE protocol).*

Study sites should send to Francisco Pozo the data collection form including individual information for each strain characterised *(Annex II).*
Francisco Pozo will extract the needed information from GISAID and send it to Epiconcept (individual data).

After each period of strain selection, study sites should send to Epiconcept the sampling fraction used and the start/end of the period *(Annex III).*

Epiconcept will link the virological data (e.g. clade) to the epidemiological data collected by GPs (e.g. patient characteristics, vaccination status) using the patients MCCS identification code. This code is unique for each patient and defined by each study site.
Epiconcept will check if the strains correspond to patients included in the MCCS analysis. Epiconcept will send out a query to the study site if a record in the virological data does not match a record in the epidemiological data.

5. Limitations

This study is subject to logistical and methodological constraints and limitations. Study sites are encouraged to document all of them. They include

- Possible low sample sizes for the
  - Descriptive objective
  - Clade-specific /genetic variant-specific VE studies
- Specimens excluded due to low viral load (potential lack of representativeness)
- Timeliness of results
- Technical issues with sequencing viruses

6. Communications

Communications, e.g. scientific manuscripts will follow the procedures described in the I-MOVE primary care generic protocol.

7. References

1. Kissling E, Valenciano M, Larrauri A, Oroszi B, Cohen JM, Nunes B, et al. Low and decreasing vaccine effectiveness against influenza A(H3) in 2011/12 among vaccination target groups in Europe: results from the I-MOVE multicentre case-control study. Eurosurveillance. 2013;18(5).

2. Skowronski DM, Chambers C, Sabaiduc S, De Serres G, Winter A-L, Dickinson JA, et al. Beyond Antigenic Match: Possible Agent-Host and Immuno-epidemiological Influences on Influenza Vaccine Effectiveness During the 2015-2016 Season in Canada. J Infect Dis. 2017 Dec 19;216(12):1487–500.

3. Flannery B, Smith C, Garten RJ, Levine MZ, Chung JR, Jackson ML, et al. Influence of Birth Cohort on Effectiveness of 2015-2016 Influenza Vaccine Against Medically Attended Illness Due to

2009 Pandemic Influenza A(H1N1) Virus in the United States. J Infect Dis. 2018 Jun 20;218(2):189–96.

4. ECDC. Protocol for case control studies to measure pandemic and seasonal vaccine effectiveness in the European Union and European Economic Area [Internet]. Stockholm: European Centre for Disease Prevention and Control; 2010. Available from: _http://www.ecdc.europa.eu/en/publications/Publications/0907_TED_Influenza_AH1N1_Measuring_Influenza_Vaccine_Effectiveness_Protocol_Case_Control_Studies.pdf_

5. Lemeshow S, World Health Organization, editors. Adequacy of sample size in health studies. Chichester [England] ; New York : New York, NY, USA: Published on behalf of the World Health Organization by Wiley ; Distributed in the U.S.A., Canada, and Japan by Liss; 1990. 239 p.

Annex I Sample size calculation and estimation of sampling fraction

We will calculate the sample size to obtain a clade /genetic variant-specific VE with sufficient precision using the following parameters:

- an estimated seasonal vaccine coverage (VC) among the source population (or among controls),
- an estimated control/case ratio and an estimated clade/genetic variant-specific VE (estimated early in the season if possible).

In the table below, we present the estimated sample sizes needed using a VC ranging from 10% to 70%, a control/case ratio of 3 or 4 and for three different crude VE of:

- ✓ 75 % with a lower bound of the 95% confidence interval above 50%;
- ✓ 50 % with an upper bound of the 95% confidence interval lower than 75%;
- ✓ 25 % with an upper bound of the 95% confidence interval lower than 50%

The sample size formula is adapted from Lemeshow, Hosmer and Klar (5).

Table I. Sample size estimation for pooled clade-specific crude VE (in each stratum)

Alpha	Confidence Interval (CI)	Controls/ case in pooled analysis	Vaccine effectiveness	Vaccine coverage in controls	Number of cases in the pooled analysis	Number of controls
0.05	upper limit of CI lower than 0.5	3	0.25	0.1	466	1,397
0.05	upper limit of CI lower than 0.5	3	0.25	0.2	252	755
0.05	upper limit of CI lower than 0.5	3	0.25	0.3	184	552
0.05	upper limit of CI lower than 0.5	3	0.25	0.4	154	462
0.05	upper limit of CI lower than 0.5	3	0.25	0.5	142	426
0.05	upper limit of CI lower than 0.5	3	0.25	0.6	141	424
0.05	upper limit of CI lower than 0.5	3	0.25	0.7	155	464
0.05	upper limit of CI lower than 0.5	4	0.25	0.1	442	1,766
0.05	upper limit of CI lower than 0.5	4	0.25	0.2	238	952
0.05	upper limit of CI lower than 0.5	4	0.25	0.3	178	694
0.05	upper limit of CI lower than 0.5	4	0.25	0.4	145	581
0.05	upper limit of CI lower than 0.5	4	0.25	0.5	133	533
0.05	upper limit of CI lower than 0.5	4	0.25	0.6	132	530
0.05	upper limit of CI lower than 0.5	4	0.25	0.7	144	577
0.05	upper limit of CI lower than 0.75	3	0.50	0.1	324	972
0.05	upper limit of CI lower than 0.75	3	0.50	0.2	167	500
0.05	upper limit of CI lower than 0.75	3	0.50	0.3	116	347
0.05	upper limit of CI lower than 0.75	3	0.50	0.4	92	275
0.05	upper limit of CI lower than 0.75	3	0.50	0.5	80	239
0.05	upper limit of CI lower than 0.75	3	0.50	0.6	75	224
0.05	upper limit of CI lower than 0.75	3	0.50	0.7	77	230
0.05	upper limit of CI lower than 0.75	4	0.50	0.1	312	1,246
0.05	upper limit of CI lower than 0.75	4	0.50	0.2	159	638
0.05	upper limit of CI lower than 0.75	4	0.50	0.3	110	440
0.05	upper limit of CI lower than 0.75	4	0.50	0.4	87	348
0.05	upper limit of CI lower than 0.75	4	0.50	0.5	75	300
0.05	upper limit of CI lower than 0.75	4	0.50	0.6	70	280
0.05	upper limit of CI lower than 0.75	4	0.50	0.7	71	285
0.05	lower limit of CI above 0.5	3	0.75	0.1	354	1,062

0.05	lower limit of CI above 0.5	3	0.75	0.2	171	513
0.05	lower limit of CI above 0.5	3	0.75	0.3	111	323
0.05	lower limit of CI above 0.5	3	0.75	0.4	81	243
0.05	lower limit of CI above 0.5	3	0.75	0.5	64	193
0.05	lower limit of CI above 0.5	3	0.75	0.6	55	164
0.05	lower limit of CI above 0.5	3	0.75	0.7	50	150
0.05	lower limit of CI above 0.5	4	0.75	0.1	346	1,385
0.05	lower limit of CI above 0.5	4	0.75	0.2	166	666
0.05	lower limit of CI above 0.5	4	0.75	0.3	107	429
0.05	lower limit of CI above 0.5	4	0.75	0.4	78	312
0.05	lower limit of CI above 0.5	4	0.75	0.5	62	246
0.05	lower limit of CI above 0.5	4	0.75	0.6	52	206
0.05	lower limit of CI above 0.5	4	0.75	0.7	47	186

To provide clade/genetic variant-specific adjusted VE or to stratify by age-group or other variables, the sample size should be increased.
Study sites that expect sufficient sample size to measure clade-specific /genetic variant-specific VE can estimate the study-specific sample size using study-specific parameters of VC, control/case ratio and clade/genetic variant-specific VE.
To obtain the parameters needed for a precise sample size calculation at pooled level, each study site should provide the estimated vaccination coverage, the proportion of cases among ILI recruited for the I-MOVE study and an estimation of the VE expected for the season (e.g. using previous season estimates).

Estimating the percentage of influenza laboratory confirmed specimens to characterise to obtain sufficient precision in clade/genetic variant-specific VE estimates (Objective 2)

For each subtype/lineage the proportion of influenza laboratory confirmed specimens to be characterised will be calculated using the formula below*:*

$$\begin{aligned} &Sampling\ fraction\ (subtype\ or\ lineage) \\ &\quad = number\ of\ IMOVE\ \ cases\ belonging\ to\ clade\ needed\ for\ VE\ estimate(1) \\ &\quad /(clade\ proportion \\ &\quad *\ expected\ number\ of\ subtype/lineage\ cases\ during\ the\ season) \end{aligned}$$

(1) Obtained in the sample size calculation section 3.5.

The sampling fraction will be calculated at pooled level, using the pooled parameters (VE, VC and proportion of cases) provided by study sites.

Where clade/genetic variant-specific VE estimates at study-site level are feasible, estimates of the number of I-MOVE influenza cases, clade proportion and expected number of subtype/lineage cases during the season can be obtained from the study site.

- ✓ Study sites with sufficient sample size, may calculate the proportion of specimens to be characterised based on the viruses available by type and subtype in the study site.

- ✓ Sites with low sample size may have insufficient cases to calculate a sampling fraction at study site level. They can either use the pooled sampling fraction, or choose a different method, (e.g. characterising all specimens of the chosen subtype/lineage).

  - ➢ *Each site to*

- *define the sampling fraction used for each subtype/lineage;*
- *document the sampling fraction for each period and for each subtype;*
- *take the sampling fraction for each subtype/lineage into account when measuring clade-/genetic variant-specific VE for the site (if applicable).*

➢ *EpiConcept to:*
- *provide parameters for estimating sampling fraction for the pooled analysis;*
- *include the sampling fraction by site, subtype and phase in the analysis measuring pooled clade/genetic variant-specific VE.*

Note that the sampling fraction defines the minimum number of strains to be characterised. In addition, the computation of the sampling fraction is based on assumptions that may vary from reality. Study sites are welcome to characterise a higher sampling fraction, as long as the random process is adhered to (see section 3.6.1) and the sampling fraction documented.

Annex II: GISAID data collection form

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

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

Row for strain with AA substitutions compared with vaccine reference strain																		

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

	Country	ID number I-MOVE+ case-control study	GISAID number	Selected for characterisation?	Reasons for not characterising?	CT value	Type of sample (primary specimen or isolate)
Row for 2018/19 vaccine reference strain							
Row for strain with AA substitutions compared with vaccine reference strain							
Row for strain with AA substitutions compared with vaccine reference strain							

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

Annex III: Sampling fraction collection form

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

Time period	First date of time period	Last date of time period	Sampling fraction used	Date used for definition of time unit (onset date, swab date, other)	Comments
A(H3N2)					
1					
2					
Example1	*01/10/2016*	*31/12/2016*	*1*	*Date of onset*	*(this is only an example; all specimens were characterised)*
Example2	*01/01/2017*	*15/02/2017*	*0.2*	*Date of onset*	*(this is only an example; 20% of all specimens were characterised)*
A(H1N1)					
1					
2					
B/Victoria					
1					
2					
B/Yamagata					
1					
2					

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