# Supplementary document 2: Statistical analysis plan

## Contents

# Background

This statistical analysis plan (SAP) is part of the protocol of the European prospective cohort study of hospital-based healthcare workers (HCWs) to evaluate the effectiveness of COVID-19 vaccines in preventing laboratory-confirmed SARS-CoV-2 infection. The follow-up period proposed is for a minimum of three months, although investigators may wish to maintain established cohorts to continue to address vaccine effectiveness and other research questions.

All HCWs eligible to be vaccinated with COVID-19 vaccine can be enrolled in the study, including those who have already been vaccinated, intend or do not intend to be vaccinated, and those who are not sure. At enrolment, study participants should complete a baseline enrolment survey about demographics, clinical comorbidities, and work- and community-related behaviours related to infection risk. In addition, a baseline serology and a respiratory sample should be collected from all participants.

During the course of the study, participants should be actively followed for SARS-CoV-2 infection through regular monitoring:

- Molecular testing: participants should provide a weekly sample, either a nasopharyngeal swab collected by trained HCWs (self-swab following training) or a self-taken saliva sample, which should be tested for SARS-CoV-2 by RT-PCR. Study site investigators should select for genetic sequencing all or a representative proportion of SARS-CoV-2 confirmed infections in participants.
- Survey: participants should complete a brief weekly survey reporting the appearance of any COVID-19-related symptoms and any changes in vaccination and high-risk exposures to infection (both professional and personal).
- Serology: serum samples should be collected periodically (6-12 weeks) from participants. At a minimum serology should be tested for antibodies against SARS-CoV-2 by serological testing algorithms that can distinguish between vaccine-induced and infection-induced antibodies. If resources allow, further laboratory testing for correlates of disease protection such as testing for neutralising antibodies and markers of cell-mediated immunity can be undertaken on the sera or additional blood collected.

Irrespective of their participation in the study, participants should immediately follow local reporting and testing procedures if they have symptoms compatible with SARS-CoV-2 infection or if they were close contacts with confirmed COVID-19 cases. Participants diagnosed with SARS-CoV-2 infection should be followed-up for outcomes including disease severity 30 days after the symptom onset or first positive laboratory test, if the positive HCW does not report to work and re-integrated in the cohort.

Vaccine effectiveness should be analysed as described in the present plan of analysis which is being continuously updated and reviewed. However, in addition to the final analysis at the end of the study period, interim analyses at different points during the study can be undertaken.

# 1.  General approach to analysis

The primary objective is to measure product-specific COVID-19 vaccine effectiveness (VE) among hospital healthcare workers (HCWs) eligible for vaccination against all laboratory-confirmed SARS-CoV-2 infections.

For the purposes of this SAP, the following **assumptions** will be used:

- A person infected is still at risk of reinfection with Omicron variant, but not with the variants that circulated prior to Omicron. For variants that may emerge during the study period, we will define risk of re-infection based on the existing evidence.
- A person can be re-infected with the variants that circulated prior to Omicron 60 days after his/her last positive PCR (restarts contributing to person-time on day 61). Until day 61 post infection, the person does not contribute to person-time analysis.
- If a person does not have a date of onset of previous infection, we assume the infection occurred at the date of the first positive test.
- A person starts to contribute to the person-time of the vaccinated cohort, on day 14 post vaccination (first and second dose) or day 7 post-vaccination (booster doses).
- Depending on the country vaccination policy, an unvaccinated infected person does not contribute to person-time until he/she becomes eligible for vaccination (i.e. we take into account the recommendation to delay vaccination for those previously infected for three month since date of onset of the previous COVID-19 episode).

# 2. Objectives

## 2.1 Primary objective

To measure product-specific VE against PCR-confirmed SARS-CoV-2 infection among hospital health workers eligible for vaccination.

## 2.2 Secondary objectives

Depending on the sample size, secondary analysis may include measuring VE:

- against specific outcomes:
    - asymptomatic laboratory-confirmed COVID-19 infection;
    - against symptomatic laboratory-confirmed COVID-19 infection;
    - against severe laboratory-confirmed COVID-19 infection;
    - against SARS-CoV-2 variants of interest/concern;
    - re-infection with SARS-CoV-2 during the study period.
- by different exposures:
    - vaccinations status (i.e. by dose, by primary course or first/second booster);
    - vaccine product and by combination of different products;
    - time since vaccination;
    - time between vaccine doses;
    - previous COVID-19 episode/hybrid protection.
- by HCW characteristics:
    - age group;
    - gender.
- by risk groups
    - high-risk comorbidities;
    - HCW occupation and/or ward.

# 3. Definitions

## 3.1 Defining COVID-19 vaccination and intervals for analysis

Study participants will be considered as having received a vaccine against COVID-19 on the day of their first dose of vaccine. If this occurs during the study period, they will move from the unvaccinated cohort into the vaccinated cohort on this day. If participants were vaccinated before enrolment or on the day of enrolment, they will enter the study as vaccinated.

Analyses will be carried out according to the number of doses and type of the vaccine received taking into account the information from different vaccines from their product characteristics. A 'wash-out' period between 0–13 days after vaccination is considered for each dose of vaccine. (Note these definitions can evolve according to the evolving vaccination recommendations in the participating hospitals)

The following definitions will be used:

- *Unvaccinated:* Study participants contribute to the unvaccinated cohort if they did not receive any dose of COVID-19 vaccine at enrolment and stop contributing to the unvaccinated cohort on the day of receipt of any COVID-19 vaccine.
- *Partially vaccinated:* This category comprises:
  - *Vaccinated with one dose*: The 'vaccinated with one dose' cohort will begin ≥14 days after study participants received the first dose of vaccine other than COVID-19 vaccine Jcovden, previously known as COVID-19 Vaccine Janssen and until the day they received the se*cond dose*.
- Vaccinated with primary course: This category comprises:
  - *Vaccinated with one dose of COVID-19 vaccine Jcovden, previously known as COVID-19 Vaccine Janssen:* The 'vaccinated with one dose of COVID-19 vaccine Jcovden, previously known as COVID-19 Vaccine Janssen' cohort will begin ≥14 days after study participants received the first dose of vaccine.
  - *Vaccinated with two doses* (based on a two-dose primary course vaccine): The 'vaccinated with two doses' cohort will begin ≥14 days after study participants received the second dose of vaccine.
- Vaccinated with any booster dose: This category comprises:
  - *Vaccinated with first booster dose* cohort that will begin ≥7 days after study participants received the first booster dose of vaccine after three months[1] since the completion of primary course.
  - *Vaccinated with second booster dose* cohort will begin ≥7 days after study participants received the second booster dose of vaccine after four months since the completion of vaccination with the first booster and they **are in a high-risk group** that requires 3+1 schedule.
- *Vaccinated with >3 doses:* This cohort will begin ≥7 days after study participants received the fourth dose of vaccine after four months since the completion of vaccination with the first booster and **regardless of whether they are in a high-risk group** that requires a 3+1 schedule.

  *Note that the current recommendations of vaccination at the European level includes a four-dose schedule for vulnerable risk groups (older adults 60+years, and people all ages with high-risk conditions). Those eligible should receive the second booster at least four months after the first booster dose, with priority those who received the last vaccine dose 6 months or more. There is no recommendation for a second booster dose in immunocompetent individuals <60 years.[2]*

- *Ever vaccinated* (at least one dose vaccine): The 'ever vaccinated' cohort will begin ≥14 days after study participants received the first dose of vaccine, regardless of whether they received additional doses before enrolment or during the study.
- *Time since vaccination*: If sample size allows, the time since vaccination can be used as exposure. For this analysis, different intervals will be used to define time since vaccination (according to the current vaccination recommendations in place and knowledge acquired at the moment of the analysis: for example before and after 90 days since the last dose). Alternatively, time since vaccination can be modelled as a continuous variable, with a restricted cubic spline, or a polynomial.

---

[1] This is the current recommendation (See footnote 2). However, during the period in which the Delta variant was dominant, the interval was reduced to two months in many countries.

[2] ECDC and EMA, Updated joint statement from ECDC and EMA on additional booster doses of COVID-19 vaccines. Press release 11 July 2022. Available at: https://www.ecdc.europa.eu/en/news-events/updated-ecdc-ema-statement-additional-booster-doses-covid-19-vaccines

- *Interval between doses*: If sample size allows, the time between the different vaccine doses can be considered as exposure, considering as standard the time stipulated in the summary of the product characteristics of each brand of vaccine.

Reference groups for the VE analysis (this section will be updated according to evolution of the vaccination recommendations):

- **Unvaccinated versus vaccinated cohorts**: If the sample size allows, for the primary analysis, the events among unvaccinated person-time will be compared to events in the vaccinated person-time periods as defined as above. The 'wash-out' periods between 0–13 days after vaccination will be excluded from the analysis.
- **Primary course versus booster cohorts**: In case of high vaccination uptake before enrolment, the events during the person-time with a primary course schedule will be compared to events during the person time with any or first booster dose. The 'wash-out' periods between 0–13 days after vaccination with each dose will be excluded from the analysis.

> Note in the product characteristics of different brands of vaccines, vaccinated people are considered protected starting the seventh day after the second or third dose. For simplifying the analysis, the same 'wash-out' period for all doses of vaccines will be considered, regardless of the brand.

# 3.2 Defining previous SARS-CoV-2 infection status of participants at the time of enrolment

The previous SARS-CoV-2 infection episodes definition takes into account the previous molecular, antigenic, and serological testing as well as radiological diagnosis.

Consideration should be given to using serology tests that can distinguish between natural and vaccine-induced immunity. If a HCW has already been vaccinated when the study started, and depending on the vaccine type, it will be important to differentiate natural and vaccine-induced immunity at baseline. If all vaccines currently used in the participating study sites are spike-based proteins[3], serological tests detecting SARS-CoV-2 spike (S) and nucleocapsid (N) antibodies could be used for distinguishing a recent infection (i.e. S+/N+) from vaccine-acquired antibodies (i.e. S+/N-). However, the dynamic in antibody titers depend on the severity of the symptoms and other host related factors still need to be elucidated (1,2). Therefore, a negative antibody titre cannot exclude with certitude a previous SARS-CoV-2 episode.

## 3.2.1 Participants not previously infected

Participants will be considered **not to have been previously infected** if they meet the following conditions:

- At enrolment
    - Their PCR test at enrolment was negative, AND
    - The participant did not report any SARS-CoV-2 episode prior to enrolment AND
    - Present a negative serology test regardless of the antigen-specific antibody tested IF the HCW was unvaccinated or vaccinated with the first dose <5 days before the serology test.

A positive serological test at enrolment will be interpreted according to the vaccination status and the type of the test performed. A negative anti-N antibody test is indicative of the absence of a recent infection; however, it cannot exclude the HCW experienced a SARS-CoV-2 infection older than six months, therefore this should be interpreted with caution.

## 3.2.2 Participants previously infected with SARS-CoV-2

Participants will be considered previously infected if they meet the following conditions:

- At enrolment
    - They had a SARS-CoV-2 positive PCR test, OR
    - They reported a previous SARS-CoV-2 infection episode, any time since the beginning of the pandemic. These episodes can be self-reported, radiologically or laboratory confirmed (by PCR or rapid antigenic test), regardless of serological test results, OR
    - They had a positive serology test that is indicative of a SARS-CoV-2 infection taking into consideration the vaccination status. The following situations can be encountered:

---

[3] On 24 June 2022, the European Commission granted a marketing authorisation for COVID-19 Vaccine (inactivated, adjuvanted) Valneva for use in the primary vaccination of people from 18 to 50 years. COVID-19 Vaccine Valneva contains inactivated (killed) whole particles of the original strain of SARS-CoV-2 that cannot cause disease. However, this vaccine has so far not been used in the participating study sites. Serological assays might need to be adapted in case of including individuals vaccinated with Valneva.

- o   In HCWs unvaccinated or vaccinated fewer than five days prior to serology, any positive serology test is indicative of previous SARS-CoV-2 infection.
- o   In HCWs vaccinated with vaccines targeting the spike protein, a positive anti-N antibody test is indicative of recent infection.

**Possible previous natural infection with SARS-CoV-2**

- A participant who reported a previous SARS-CoV-2 episode prior to enrolment, without details on date or type of test, AND tested negative at serology test at enrolment or for whom serology is not available, OR
- A participant self-diagnosed with a previous COVID-19 episode with no test AND tested negative at serology test at enrolment or for whom an enrolment serological test is not available.

Study teams are encouraged to obtain more information on previous SARS-CoV-2 infection from all participants meeting the 'possible previous natural infection with SARS-CoV-2' definition. For the main analysis, participants with possible previous natural infection with SARS-CoV-2 will be included as participants with previous infection.

## 3.2.3 Hybrid immunity exposure

A combined variable vaccination and previous COVID-19 episode with four levels of-exposure will be used for investigating hybrid protection:

- booster vaccination and previous infection (booster/previous);
- primary schedule and previous infection (primary/previous);
- booster vaccination without previous infection (booster only); and
- primary schedule without previous infection (primary only).

This definition does not take into account the timing of the previous COVID-19 episode (if occurred before or after vaccination) nor the time since the previous COVID-19 episode nor the time since last vaccination dose. This level of complexity can be added if sample size allows.

# 3.3 Definition of outcomes/events

The primary outcome is represented by laboratory-confirmed SARS-CoV-2 infection detected by RT-PCR in any participant, regardless of symptoms. Both the outcome/event 'first SARS-CoV-2 infection' and 'COVID-19 reinfection' will be included in the overall analysis, depending on whether a study participant presented a previous SARS-CoV-2 infection or not.

Secondary outcomes are:

- Asymptomatic/symptomatic laboratory-confirmed COVID-19 cases;
- Severe COVID-19 cases;
- SARS-CoV-2 re-infections;
- SARS-CoV-2 infection or symptomatic COVID-19 of a specific variant of concern circulating during the study period.

## 3.3.1 Definition of COVID-19-like symptomatic case

A COVID-19-like symptomatic case is a HCW who developed at least one of the following signs and symptoms compatible with COVID-19 based on EC case definition.

Clinical criteria

- cough;
- fever;
- shortness of breath;
- sudden onset of anosmia, ageusia or dysgeusia

OR

Diagnostic imaging criteria

- radiological evidence showing lesions compatible with COVID-19.

## 3.3.2 Definition of COVID-19 laboratory-confirmed case

A COVID-19 laboratory-confirmed case corresponds to any HCW having a positive RT-PCR test and at least one clinical sign or symptoms of COVID-19 (see above COVID-19-like symptomatic case).

### 3.3.3 Definition of SARS-CoV-2 asymptomatic infection

A SARS-CoV-2 asymptomatic infected HCW corresponds to any HCW who tested positive for RT-PCR and did not present any signs or symptoms 14 days before or seven days after the test (3).

In a secondary analysis, we also include the HCWs who present a positive serology test without signs or symptoms since the beginning of the epidemic (if unvaccinated). For those vaccinated, the interpretation of the serological test is done accordingly (see section 3).

### 3.3.4 Definition of COVID-19 re-infection during the study period

A SARS-CoV-2 reinfected HCW corresponds to any HCW who tested positive by RT-PCR at two different points in time during the study (at least ≥60 days apart, whether or not symptoms were present) and/or with a virus sequencing indicating different strains/lineage. For a time elapsed of <60 days between two SARS-CoV-2 positive RT-PCR tests, the classification as re-infection will be done on a case-by-case basis taking into account the Ct values, presence/absence of at least one negative PCR tests between the two episodes, and the result of sequencing if paired samples will be available. The time elapsed after the first positive test excluded from the primary analysis and included in the analysis with re-infection as outcome.

HCWs with a positive serological test since the beginning of pandemic indicating a past-infection (anti-N IgG Ab or any positive serological test among unvaccinated, see section 5) or a previous positive rapid antigen test (RAT) and a RT-PCR test positive during the study will be classified as re-infection on the case-by-case basis taking into account the presence/absence of at least one negative PCR tests between the two episodes, time elapsed between the past episode and the current episode, the availability of testing at moment of the previous episode and the result of sequencing if paired samples will be available.

*Note, the 60-day period between episodes was appropriate for SARS-CoV-2 infections that occurred before the Omicron period. For re-infection with Omicron variant regardless of lineages, this time period may not be appropriate anymore. Therefore, the re-infection definition will be updated in line with the available evidence during the study.*

### 3.3.5 Definition of severe SARS-CoV-2 infection

COVID-19 disease severity is defined as a HCW with a SARS-CoV-2 laboratory-confirmed infection with the following stages:

- Asymptomatic: no reported symptoms consistent with the COVID-19-like symptomatic case definition;
- Mild disease: reported symptoms consistent with the ECDC definition of COVID-19 requiring attendance at a medical service but requiring no further assistance/hospitalisation;
- Moderate disease: reported symptoms consistent with the ECDC definition of COVID-19 requiring either hospitalisation but not requiring oxygen treatment or not hospitalised but requiring assistance for activities of daily living;
- Severe disease: reported symptoms consistent with the ECDC definition of COVID-19 requiring hospitalisation and oxygen treatment;
- Very severe disease: reported symptoms consistent with the ECDC definition of COVID-19 requiring hospitalisation and any of the following: admittance to an intensive care unit and/or intubation or mechanical ventilation and/or additional systems/organs support (vasopressors, dialysis, ECMO) or death.

# 4. Exclusion criteria

A restriction flow chart will be created to be able to view those excluded from the study.

HCWs are excluded from the study if they:

     –     refuse to participate in the study;
     –     have contraindications for the COVID-19 vaccination.

# 5. Person-time at risk

The following section relates to follow-up time in the statistical analysis, depending on analysis objectives, rather than the time a participant is being followed up as part of the study.

**Start of person-time**

Each included study participant contributes to the person-time at risk in the single event analysis. Follow-up time for each eligible individual begins at

- enrolment date; or
- first date the participant is at risk for reinfection (see section 6.4), for participants with PCR-confirmed infection in the 60 days prior to the start of the study or during the study (before Omicron circulation); or
- when the participant becomes eligible for vaccination after a SARS-CoV-2 infection.

**End of person-time**

Participants contribute person-time at risk for a single event analysis until:

- their reported date of onset of symptoms associated with a PCR-confirmed SARS-CoV-2 infection (in or outside the study); or
- their date of positive PCR (if date of symptom onset is not available or asymptomatic infection); or
- the date of vaccination for a different COVID-19 vaccine if conducting a product-specific VE analysis, or
- the censored date.

The censored date will vary depending on the date of ad-hoc analyses and will depend on interim power calculations. Nevertheless, the censoring date is the date of the last virological test, if no event has been reported. In the case the last virological test is missing, the censoring date is the previous virological test.

**Exclusion of person-time**

We will exclude from person-time calculation:

- 0–13 days after receipt of first/second dose of vaccine and 0-6 days for the any booster dose.
- From the date of the mid-way point between the last negative serological test and the subsequent positive serological test performed at 60 days or the symptoms date if information is available for the secondary analyses including serology asymptomatic outcome (HCWs with a negative serology test followed by a positive serology test, with no PCR test or negative PCR during follow-up).

# 6. Loss to follow-up

## 6.1 Missing weekly follow-up questionnaire

Study participants are asked to fill in a weekly questionnaire on symptoms and vaccination. A study participant is defined as having

- complete loss to follow-up: if study participants do not fill in any further weekly questionnaires;
- partial loss to follow-up (intermittent missing data): if study participants miss weekly questionnaires intermittently, up to 21 days between the virology tests.

Reasons for complete loss to follow-up can include HCWs no longer wanting to participate in the study, ceasing to work at the hospital of recruitment, or illness/hospitalisation due to COVID-19 or other disease. Loss to follow-up should be minimised as much as possible, and when not possible, all attempts should be made to nevertheless obtain outcome and vaccination data on these participants. The HCWs lost to follow-up should be replaced with other HCWs who are eligible and willing to participate in the study (according to recruitment criteria).

Reasons for partial loss to follow-up can include questionnaire fatigue by HCWs, or HCWs being on holiday.

HCWs will not be followed up properly during the longer summer breaks or long annual or sick leave and should not be censored from the study. We have considered scenarios below:

- If participating HCWs are on leave for a maximum period of 3 weeks (i.e. 21 days: they miss only one follow-up for a biweekly follow-up or two follow-ups for the weekly follow-up), they will not be censored from the study if they have a NP swab and a PCR on their return to work.
- Those HCWs with longer breaks (i.e. > 21 days) will be included in sensitivity analysis if they provide a NP swab on return and complete the follow-up questionnaire during their absence. Information on symptoms and vaccination during this period without follow-up should be sought.

Those study participants with complete loss to follow-up will be **censored from the date of their last weekly questionnaire** and compared to those not lost to follow-up to understand if these populations differ by vaccination status and risk of infection.

In the main analysis, those with partial loss to follow-up will be **censored from the date of their last weekly questionnaire**, unless information on outcome and vaccination could be obtained. In sensitivity analyses, study participants with partial loss to follow-up will be kept in their respective group of exposure and imputation and weighting techniques will be used to account for loss to follow-up.

## 6.2 Missing RT-PCR or serological tests

Missing PCR: if a participant is missing two or more PCR tests (two or more weeks with no PCR results), then the follow-up time will be censored from the date of their last weekly questionnaire before the first missing PCR test (not enough information for the proposed imputation of asymptomatic infection).

In case that, after the PCR missing period, a positive serological test is recorded (anti-N serology or increase in the titre of anti-S antibodies), the follow-up will be censored until the date of this positive result. Negative serological results alone (with no PCR results) will not be used to determine the follow-up time in each participant.

Missing serology tests: We accept a variation of +/- 2 weeks between two serology tests.

# 7. Potential effect modifiers and confounding factors

The list of variables below is collected and are potential effect modifiers/confounding factors of the association between the vaccination and the outcome.

The variables in blue indicate time-varying variables and include only non-work-related exposures. Age is a time-varying variable, but due to the short study period (one year or less), it is only included at enrolment. Most variables listed here, including work-related exposure variables are collected at enrolment only.

- **HCW data:**
    - Demographic: age, sex, residence;
    - Underlying diseases, number of hospitalisations previous 12 months;
    - Vaccination with other vaccines (influenza, pneumococcal vaccine);
    - Illness before the start of the study/Laboratory results for last episode;
    - Previous infection;
    - Coinfection with other viruses.

- **Non-work-related exposure:**
    - Contact with confirmed COVID-19 case outside of work (date of contact);
    - Type of transport to work;
    - Number of members of household and job of the adults;
    - Compliance with recommendations at community level.
        - Frequency of:
            - use of public transport;
            - attendance of large social indoor gatherings.
        - Frequency of:
            - mask wearing in indoor settings outside of home;
            - social distancing indoors outside of home;
            - visits by people to home;
            - visits to other people's homes.

- **In hospital exposure:**
    - Full-/part-time work;
    - HCW categories (ECDC TESSy);
    - Duration of shifts;
    - Specific ward;
    - Contact with a HCW case;
    - Contact with a patient case: number of times, time;
    - Participation in aerosol generating procedures (type and number);
    - Self-assessment of PPE use and hand hygiene.
- **Time**
    - Period of each variant of concern's (VOC) circulation.

# 8. Data analysis

For the primary analysis, the overall cohort according to time when each VOC circulated is used.

For secondary analyses, the overall cohort according to the time when each VOC circulated is used and it is assessed if different covariates are effect modifiers (e.g. sex, previous COVID-19 episode) or confounder factors (age, underlying conditions, wards), when sample size allows.

For the analysis by previous COVID-19 episode, to investigate the hybrid immunity, a specific exposure variable combining previous infection and vaccination (see section 4.2.3) will be used.

## 8.1   Exclusion and missing data

The total number of eligible HCWs, the total number and proportion of HCWs who refused participation by reason for refusal will be described. The total number of potential study participants and the number of study participants excluded will be presented in an exclusion flowchart (similar to Figure 1).

**Figure 1. Example of exclusion flowchart**
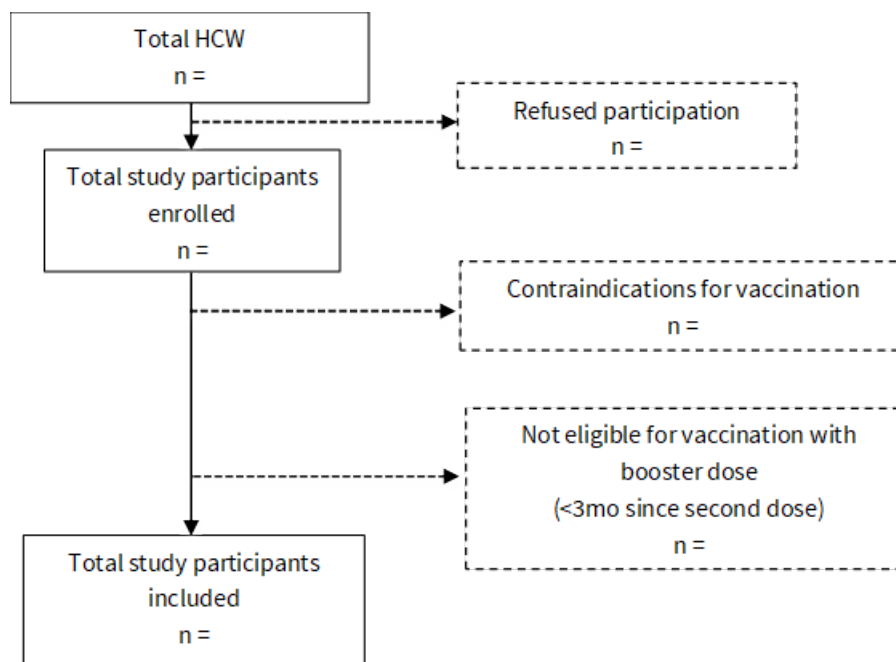


## 8.2 Minimum sample size

The sample size estimates are presented in the main protocol.

Note: those estimates do not take into account the increase in sample size needed to adjust for confounders or effect modification, and the loss to follow-up. The power will be calculated with the reached sample size.

## 8.3 Descriptive analyses

### 8.3.1 Descriptive analysis of missing data of key variables

The completeness of information will be analysed for key variables (Table 1).

**Table 1.** Completeness of key variables

| Variable | Number missing | % of total missing |
|---|---|---|
| Age | | |
| Sex | | |
| Serological test at enrolment | | |
| PCR test at enrolment | | |
| Information on previous infection | | |
| Job | | |
| Smoking status | | |
| Vaccination information (dose, date, brand, mode of ascertainment) | | |
| Community exposures | | |
| In-hospital exposures | | |

Histograms will be compiled of the following dates to identify outliers and systematic missing data:

- Dates of enrolment;
- Dates of symptom onset and dates of specimen collection for new infections;
- Dates of test of reported previous infection (distinguishing between types of tests);
- Date of vaccination (by first/second dose, vaccine brand).

## 8.3.2 Description of participant characteristics at baseline

Baseline characteristics of study participants should be tabulated. Depending on the variable type, the mean, median or proportion should be presented. The number of individuals with missing data for each variable should be presented.

Baseline characteristics tabulated should include:

- age (age histogram);
- sex;
- comorbidities and treatments;
- obesity;
- smoking history;
- COVID-19 vaccination history;
- history of other vaccines (influenza, pneumococcal vaccination);
- HCW category;
- previous SARS-CoV-2 infection;
- occupational and community-related exposures.

## 8.3.3 Description by vaccination status

Participants' characteristics and exposures by vaccination status at the end of the study period will be described according to potential confounders and effect modifiers, as displayed in Table 2.

**Table 2. Participants' characteristics and exposures by vaccination status**

| Characteristic | Unexposed (Unvaccinated or Completed primary vaccination course) | | Exposed (Unvaccinated / Received a booster dose) | |
|---|---|---|---|---|
| | N | % | N | % |
| **Sex** | | | | |
| Female | | | | |
| Male | | | | |
| **Age group (years)** | | | | |
| 18-34 | | | | |
| 35-39 | | | | |
| 40-44 | | | | |
| 45-49 | | | | |
| 50-54 | | | | |
| 55 and older | | | | |
| **Role** | | | | |
| Medical doctor | | | | |
| Nurse | | | | |
| Administration/reception | | | | |
| Ancillary | | | | |
| Allied | | | | |
| Laboratory | | | | |
| Other | | | | |
| **Smoking** | | | | |
| Never smoked | | | | |
| Ex-smoker | | | | |
| Current smoker | | | | |
| **Chronic condition** | | | | |
| Reporting chronic conditions | | | | |
| Reporting >1 condition | | | | |
| **Regular medication** | | | | |
| Reporting taking medication | | | | |
| Taking >1 medication | | | | |
| **Previous COVID-19 episode** | | | | |
| Yes | | | | |
| Median time since last previous COVID-19 episode (range) | | | | |
| No | | | | |
| Missing/Unknown | | | | |
| **Time since last dose vaccination** | | | | |
| Median time (range) | | | | |
| <90 days | | | | |
| 90+ days | | | | |
| …. | | | | |

In addition, proportions of symptoms among symptomatic participants will be described by outcome. This will be done overall, and by vaccination status.

# 8.4 Regression choice

A person time multiple regression (Cox proportional hazards model) analysis will be carried out to analyse the effect of multiple variables and identify and analyse effect modification and control for confounding.

CoxI regression is at present preferred over Poisson regression as there is no assumption of stability of the COVID-19 rate over time. A priori confounders and the change in point estimate for inclusion of other covariates in the adjusted analysis should be outlined and clinical relevance noted, if necessary. The presence of effect modification/interaction terms should be explored.

## 8.5 Univariate analysis

In accordance with our study objectives, we will build our main results tables as displayed in Table 3.

**Table 3.** **Incidence rate ratio for SARS-CoV-2 infection, personal characteristics and in-hospital and community exposures by vaccination status***

| | N events/person time | | Hazard Ratio (95%CI) |
|---|---|---|---|
| | **Vaccinated/Booster vaccination** | **Unvaccinated/ Complete primary course** | |
| **Overall** | | | |
| | | | |
| **Previous infection** | | | |
| Reported | | | |
| Not reported | | | |
| **Vaccine brand** | | | |
| Spikevax | | | |
| Comirnaty | | | |
| Vaxzevria | | | |
| JCovden (1 dose only) | | | |
| **Sex** | | | |
| Male | | | |
| Female | | | |
| **Age group (years)** | | | |
| <50 | | | |
| 50+ | | | |
| **Chronic conditions** | | | |
| **Study site** | | | |
| ... | | | |

*Same table for each outcome, each analysis period.*

## 8.6 Complete case analysis

The primary approach will be a complete case analysis. In the event of 10-20% or greater amounts of missing data in key variables (age, sex, underlying conditions, previous COVID-19 episode), a multiple imputation using chained equations will also be carried out.

## 8.7 VE analysis

VE will be calculated at univariable and multivariable levels. Primary analyses will be carried out in the overall cohort restricted to the time of circulation of each VOC, due to their different epidemiology. In the secondary analyses we will use the same approach as well as undertake separate analyses, if sample size allows, for participants who did and did not report a previous infection at enrolment.

VE will be estimated as (1-HR)*100. If possible, the primary analysis will compare the incidence rate/hazard rate in unvaccinated and that in vaccinated. In light of the very high COVID-19 vaccine coverage rates, with nearly all HCWs having received at least one dose of vaccine before enrolment, comparing the incidence of SARS-CoV-2 infection or COVID-19 disease in vaccinated and unvaccinated groups may not be possible. Therefore, the rates in HCWs who received a booster dose with those of the HCWs who did not receive a booster dose but received a primary course of vaccination (relative VE) will be compared.

To investigate the hybrid immunity, SARS-CoV-2 incidence in HCWs with previous infection and first booster vaccination is compared with those with no previous infection and primary course vaccination using a four-level variable (see section 4.3).

If sample size allows, VE can be measured for varying time since vaccination. Intervals will be established according to the available knowledge and vaccination recommendations, depending on sample size. As study time continues, further intervals can be included. Alternatively, time since vaccination can be modelled as a continuous variable, with a restricted cubic spline, or a polynomial.

**Multivariable analysis model building approach**

Priority potential confounders to be included in the model are:

- Age;
- Sex;
- Comorbidities (the coding of which will be determined during model building);

- Study site/hospital;
- Prior infection status (in the overall analysis);
- Calendar time.

Other potential confounders are listed in section 10. Due to the expected high proportion of participants with previous infection in this particular study, a low number of events may be observed, and care should be taken not to overfit the model with too many parameters per event. Therefore, a parsimonious approach to model-building will be sought.

A backwards stepwise model fitting approach will be used, and those potential confounders modifying the adjusted VE by 5% absolute or more will be retained in the model.

## 8.8 Modelling continuous variables

Continuous variables include age and calendar time. Calendar time will be modelled in categories/intervals to more easily split the data. For age, the AIC[4], as well as a qualitative assessment of complexity of the functional form, will be used to determine the best functional form (age groups, age as a linear term, a restricted cubic spline or a polynomial of a higher order than 1).

## 8.9 Sparse data

Models will be assessed to assure there are 10 or more events per parameter (mainly in the stratified and adjusted analyses). Parsimonious model building will be carried out. If needed, penalised regression methods could be considered.

## 8.10 Collinearity

Variables will be assessed for collinearity using the variance inflation factor.

## 8.11 Proportional hazards assumption

The proportional hazards assumption of Cox regression will be assessed using graphical approaches and tests based on Schoenfeld residuals. If there is evidence of non-proportionality, then a proportional hazards model may not be appropriate. A full set of frailty mixture models should be fitted to assess the appropriate method to measure VE.

## 8.12 Controlling for clustering by hospital

To control for a clustering effect by health facility, a mixed model could be considered, including the hospital as a random intercept. If no effect is found, the hospital will be included in the model as a fixed effect.

## 8.13 Sensitivity analyses

Sensitivity analyses will be carried out if sample size allows and relevant:

- Considering vaccinated status to be reached at 7 days after the second dose of a two-dose vaccine, or each booster dose;
- Using different outcomes and combination of outcomes (PCR, serology);
- Correcting for sensitivity and specificity of various outcomes;
- By previous infection using different definitions of previous infection;
- Using different delays between symptom onset and specimen collection (e.g. <3/3+ days);
- Using different delays for defining vaccination status;
- Calculating E-values to quantify the potential for bias due to unmeasured confounding;
- VE by time since vaccination; including Farrington/Longini/Halloran methods for analysing/correcting for biases due to cumulative incidence risk; stable or variable incidence rate over time;
- VE by time since previous COVID-19 episode.

## 8.14 Serology analysis

Longitudinal serology test results at enrolment and at each time a test is conducted will be described. Time since enrolment and test result will be calculated, and tests will be grouped into follow-up time points (i.e. first time point will be after 4-6 weeks, second time point will be 8-12 weeks, depending on the study site testing

---

[4] Akaike Information Criterion

frequency). Serology results will be presented by study site, due to the differing testing regimes (ranging from 4-6 weeks to every 12 weeks).

Results will be split by type of antibody test (anti-S, anti-N). Any individual who had a serology test done outside of regular follow-up (i.e. due to symptoms, or other reasons unknown) will be excluded. Individuals will be split by former infection status, specifying VOC if possible. Anti-S antibody tests will give an overall assessment of protection from either COVID-19 vaccination and/or natural infection, while anti-N antibody tests provide information on those with previous natural infection only.

Specific focus will be made on identifying:

- Any individuals who seroconvert (anti-N) during the study period in the absence of a positive PCR result indicate a natural SARS-CoV-2 infection which has not been detected by regular RT-PCR testing within the study period. These individuals will be included as cases in a secondary analysis.
- Serology results and titres for cases, vaccinated and previous infections will be described.
- Due to heterogeneity of reporting of serology results, adjustments will be made using set criteria for study sites that report antibody titres.

# 8.15 Influenza and COVID-19 vaccination

A sub-analysis will be conducted to assess whether there is any additive benefit of influenza vaccination (during the most recent influenza season, at time of enrolment) and COVID-19 vaccination.

Previous studies have shown a protective effect (with widely varying VE estimates) of recent influenza vaccination against SARS-CoV-2 infection, however there is currently little evidence on the added benefit of both influenza and COVID-19 vaccination against SARS-CoV-2 infection.

# 8.16 Pooled analysis

A pooled analysis is the primary objective of the ECDC multi-country study and will be performed at central level. Data validation, cleaning, and verification will be carried out at study site and coordination/central levels. A hospital identifier will be included in each record and each record will be given a unique and persistent identifier so that the enrolment and follow-up questionnaires, laboratory and serology forms can be linked. This identifier will also be included in the study team's database and will be used by the coordinating team and the study teams during pooling, so that records can be traced back while maintaining anonymity if there should be any further queries. Tracing back will be performed by the participating hospital teams, not by the central level/coordinating team.

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

A study-site specific flowchart of exclusions and restrictions will be shared with each of the study sites. Variables will be recoded and new variables generated. The recoded data will be stored separately from the crude data and recoding will be documented.

Study-specific crude and adjusted HRs and their CIs will be plotted in separate forest plots. Following the core protocol minimises heterogeneity between studies. However, adherence to the protocol and study design and study quality characteristics may influence the results. A qualitative decision will be taken if one or more studies are substantially different from the other. Statistical heterogeneity between studies will be tested using Q-test and the $I^2$ index.

If sample sizes are too small to measure vaccine effectiveness controlling for all potential confounders for each individual study site, a one-stage pooled approach will be used for analysis. Individual study data will be pooled into one dataset and analysed as a one-stage model with study as a fixed effect. This assumes not only that the underlying true exposure effect is the same in all studies, but also that the association of all covariates with the outcome is the same in all studies.

If heterogeneity is found between study sites when using a one-stage fixed effects approach, reasons for heterogeneity need to be thoroughly investigated and the assumptions underlying the one-stage pooling approach need to be revisited.

If adequate sample size by study is achieved to obtain an adjusted HR by study site individually, then a two-stage approach to pooled analysis will be taken. Study site-specific adjusted HRs and confidence intervals obtained from the individual studies will be combined in a model that incorporates random effects of the study site-specific studies, to account for unmeasured country- and site-specific factors that differ between studies.

The study site-specific exposure-disease effects (HRs) are then weighted by the inverse of their marginal variances in a meta-analysis that will provide a pooled HR and confidence intervals.

# 9. Potential biases

There are many potential biases for the main analysis in this study. They include:

- Selection bias:
    - those unvaccinated may be less willing to participate and may be less likely to complete follow-up;
    - differential reporting of symptoms (and therefore probability of being tested) between exposed and unexposed (vaccinated/unvaccinated or booster/primary course vaccination);
    - more likely to be vaccinated and participate in the study if the participants have an underlying condition.
- Differences in exposure to SARS-CoV-2 between vaccinated/unvaccinated or booster/primary course that are not accounted for in the demographic information collected in the questionnaire;
- Imperfect sensitivity and specificity of serology tests and, to a lesser extent, PCR tests, leading to misclassification of outcomes and previous SARS-CoV-2 definitions;
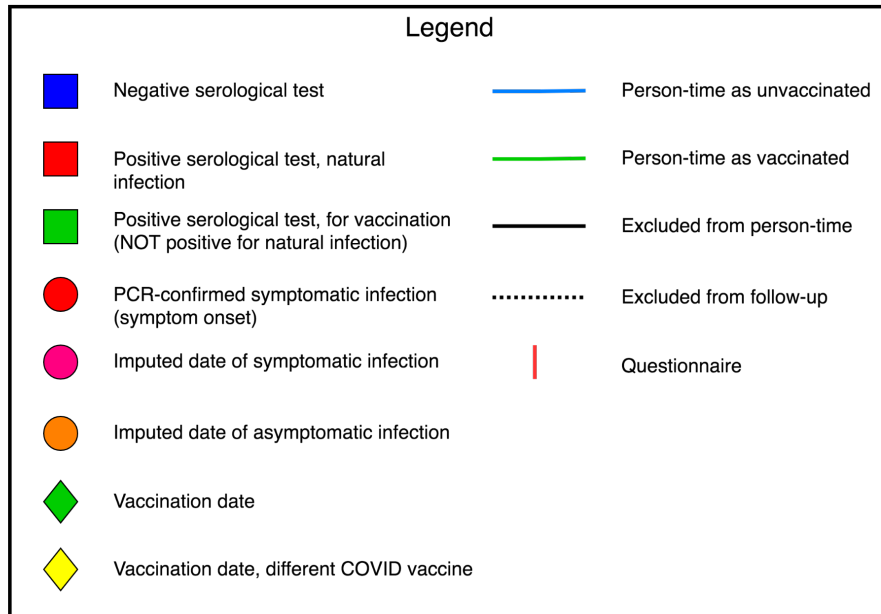- Deferral bias (those with recent SARS-CoV-2 infection may get vaccinated later).

# References

1.  Lumley SF, Wei J, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, et al. The Duration, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers. Clin Infect Dis Off Publ Infect Dis Soc Am. 2021 02;73(3):e699–709.

2.  Van Elslande J, Oyaert M, Lorent N, Vande Weygaerde Y, Van Pottelbergh G, Godderis L, et al. Lower persistence of anti-nucleocapsid compared to anti-spike antibodies up to one year after SARS-CoV-2 infection. Diagn Microbiol Infect Dis. 2022 May;103(1):115659.

3.  Meyerowitz EA, Richterman A, Gandhi RT, Sax PE. Transmission of SARS-CoV-2: A Review of Viral, Host, and Environmental Factors. Ann Intern Med. 2021;174(1):69–79.

# Annex 1. Scenarios of person-time at risk contribution in a single event analysis

This annex explains how to integrate the results of serological tests to define the contribution of individuals to the person-time at risk, presenting different possible scenarios. We assumed that serological tests were taken every 13 weeks.

Here are the symbols that will be used to represent the different elements in the scenarios:

## Legend

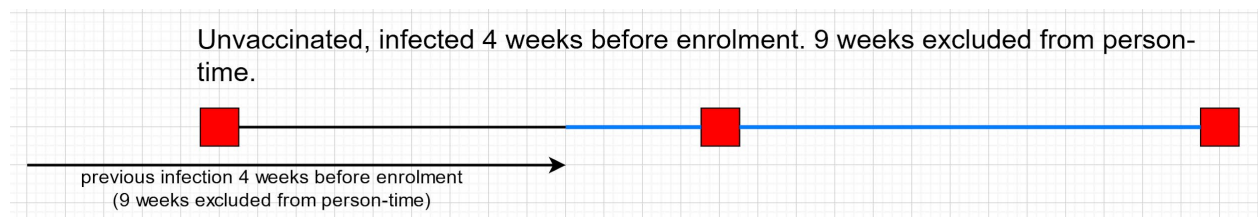| Symbol | Description | Symbol | Description |
|---|---|---|---|
| Blue square | Negative serological test | Blue line | Person-time as unvaccinated |
| Red square | Positive serological test, natural infection | Green line | Person-time as vaccinated |
| Green square | Positive serological test, for vaccination (NOT positive for natural infection) | Black line | Excluded from person-time |
| Red circle | PCR-confirmed symptomatic infection (symptom onset) | Dotted line | Excluded from follow-up |
| Pink circle | Imputed date of symptomatic infection | Red vertical bar | Questionnaire |
| Orange circle | Imputed date of asymptomatic infection | | |
| Green diamond | Vaccination date | | |
| Yellow diamond | Vaccination date, different COVID vaccine | | |

## *Start of person-time*

E1) The participant contributes to person-time as unvaccinated, all serological tests are negative.

Unvaccinated, not previously infected, not infected during the follow-up.
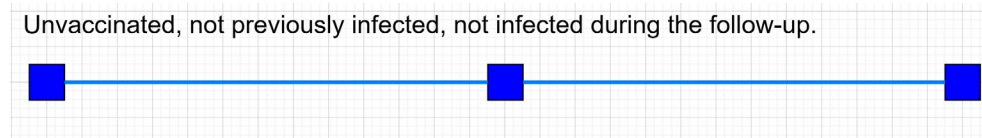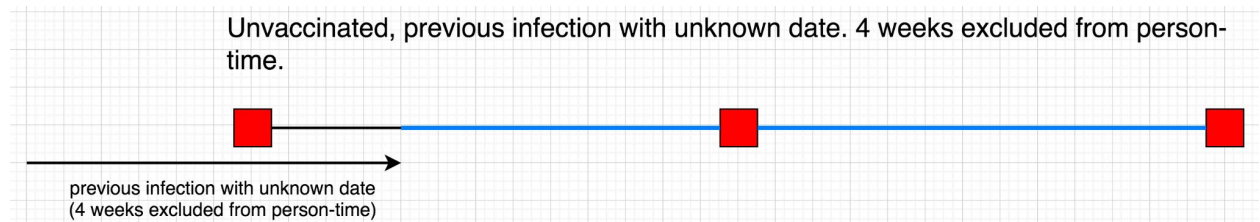
E2) The participant reports an infection 13 or more weeks before enrolment. We are interested in evaluating any possible reinfection. The assumption is that 60 days are required to assure that new, positive PCR tests and symptoms correspond to a new infection.
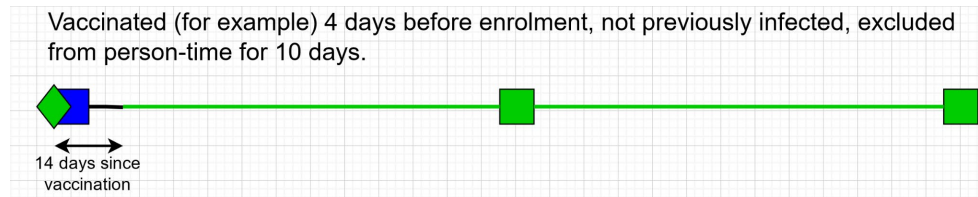
Unvaccinated, infected 13 or more weeks before enrolment. Never excluded from person-time.

previous infection >= 13 weeks before enrolment
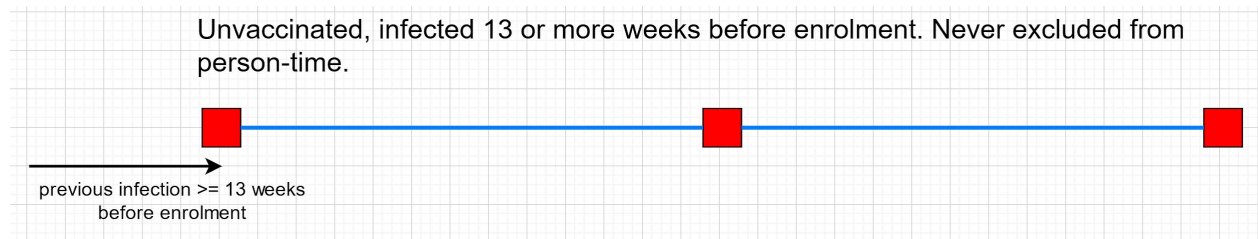
E3) The participant reports an infection less than 13 weeks before enrolment.

Unvaccinated, infected 4 weeks before enrolment. 9 weeks excluded from person-time.

previous infection 4 weeks before enrolment (9 weeks excluded from person-time)

E4) The participant reports a previous infection with an unknown date. We assumed it occurred 9 weeks before enrolment.

Unvaccinated, previous infection with unknown date. 4 weeks excluded from person-time.

previous infection with unknown date
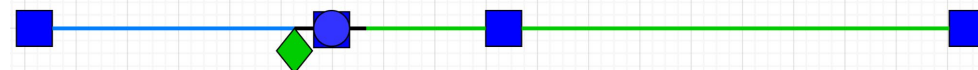(4 weeks excluded from person-time)

E5) The participant contributes to person-time as vaccinated, only 14 days after vaccination. All subsequent serological tests will be positive due to vaccination not for natural infection.

Vaccinated (for example) 4 days before enrolment, not previously infected, excluded from person-time for 10 days.
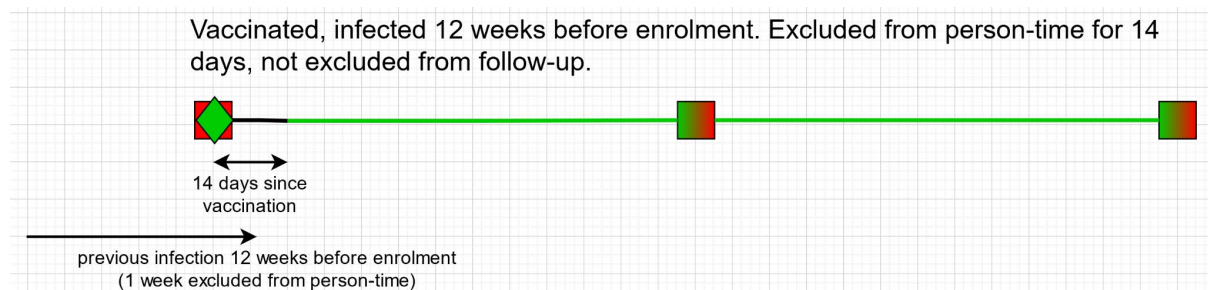
14 days since vaccination

E6) The participant was vaccinated during the follow-up period.

Not previously infected, vaccinated in week 8, followed by 14 days excluded from person-time. Not infected during the follow-up. Additional PCR test and serology test after vaccination.

E7) The participant was infected 12 weeks before enrolment and was vaccinated at enrolment. We excluded 14 days from person-time, assuming immunization will be reached after this period. The following positive serological tests would be positive for both natural infection and vaccination.

Vaccinated, infected 12 weeks before enrolment. Excluded from person-time for 14 days, not excluded from follow-up.

14 days since vaccination

previous infection 12 weeks before enrolment
(1 week excluded from person-time)

## End of person-time

E8) The participant has a PCR-confirmed symptomatic infection. The person-time ends on the date of symptoms. They are excluded from follow-up for 90 days (13 weeks), except for a questionnaire at 30 days.

Unvaccinated, not previously infected, PCR-confirmed symptomatic infection, excluded from follow-up for 90 days (13 weeks), except for a questionnaire at 30 days.
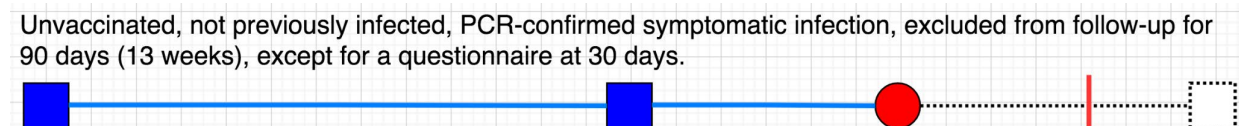
E9) The participant received a different vaccine (if doing a **product-specific** VE analysis). The person-time contribution ends at the vaccination date.
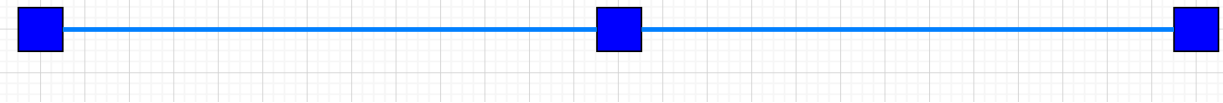
Unvaccinated, not previously infected, vaccinated with a different vaccine, excluded from persons-time for 90 days (13 weeks)

23

E10) The participant contributes to person-time as unvaccinated until the last serological test (interim/final analysis).

Unvaccinated, not previously infected, not infected during the follow-up.

E11) Last serological test is missing. The person-time contribution ends at the previous serological test.

Unvaccinated, not previously infected, not infected during the follow-up, missing serological test.
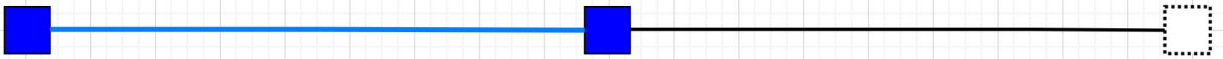
## *Periods excluded from person-time*

E13) The participant presents a positive serology without PCR confirmation or symptoms. The date of an asymptomatic infection will be imputed on the mid-point of the number of weeks during which the serological test detects infections more reliably. If we assume that a serological test detects infections that occurred at least 3 weeks before more reliably, then a positive serological test could be from infection between t=-3 and t = 10. The mid-point would be on week 3.5 after the first serological test.

Unvaccinated, not previously infected, seropositive after asymptomatic infection, excluded from person-time for 13 weeks.

E14) The participant presents a serological positive without PCR confirmation but reports symptoms. The date of symptomatic infection will be imputed in the symptoms date.

Unvaccinated, not previously infected, symptomatic infection imputed on the date of symptom onset, seropositive after asymptomatic infection.
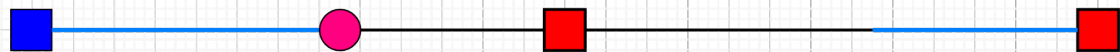
E15) The participant was infected at enrolment or within 7 days of enrolment, the participant is excluded from follow-up during 13 weeks, except for a questionnaire at 30 days.

Unvaccinated, infected at enrolment or within 7 days of enrolment. 13 weeks excluded from follow-up (questionnaire at 30 days).

infected at enrolment or within 7 days of enrolment 13 weeks excluded from follow-up
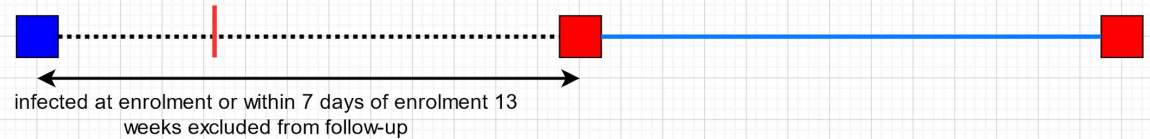
## *Partially lost to follow-up*

The participant did not send weekly questionnaires for a period of time and then returned to fill in the weekly questionnaires. Here we present different scenarios to show how events could be imputed, if additional information about symptoms or serological tests could be recovered. We also show which periods must be excluded from person-time.

E16) Possible asymptomatic infection

Negative serological test at enrolment. No symptoms reported. No information on PCR test. Asymptomatic infection imputed on week 3.5 (mid-point). Also if symptoms reported within the 3 weeks prior to the next serological test (even if a positive PCR test is reported).



E17) Possible symptomatic infection

Negative serological test at enrolment. Symptoms onset 8 weeks after serological test at enrolment (date of symptom onset does not fall within the 3 weeks preceding the next serological test). No information on PCR test. Symptomatic infection imputed on the date of symptoms onset.
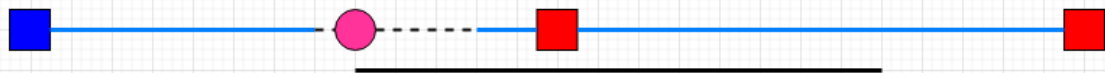


E18) Previous infection

Positive serological test at enrolment. Symptoms onset 8 weeks after serological test at enrolment. No positive PCR reported by the participant. We assume no event has occurred.