Covid-Vaccine-Monitor

Evaluation of assumptions of SCRI in simulation studies

Protocol V2.1

25 August 2021

Title	Evaluation of assumptions of SCRI in simulation studies.
Protocol version identifier	2.1
Date of last version of protocol	August 25, 2021
EU PAS register number	
Active substance	*NA*
Medicinal product	*NA*
Product reference	*NA*
Procedure number	*NA*
Marketing authorisation holder(s)	*NA*
Research question and objectives	This study will evaluate assumptions of the SCRI design to assess the association between COVID-19 vaccines and adverse events of special interest in simulated datasets informed by electronic healthcare databases in European countries.
Country(-ies) of study	*NA*
Author	René Eijkemans, University Medical Center Utrecht, The Netherlands
Contributors	Olaf Klungel, Utrecht University, The Netherlands Miriam Sturkenboom, UMCU, The Netherlands

1. Table of contents

2. List of abbreviations

MCSE	Monte Carlo Standard Error
MSE	Mean Squared Error
SAP	Statistical Analysis Plan
SCCS	Self-Controlled Case Series
SCRI	Self-Controlled Risk Interval

3. Responsible parties

René Eijkemans, University Medical Center Utrecht, The Netherlands
Olaf Klungel, Utrecht University, The Netherlands
Miriam Sturkenboom, UMCU, The Netherlands

4. Abstract

Title

Evaluation of assumptions of SCRI in simulation studies
Protocol V2.1; August 25, 2021
René Eijkemans, University Medical Center Utrecht, The Netherlands
Olaf Klungel, Utrecht University, The Netherlands
Miriam Sturkenboom, UMCU, The Netherlands

Rationale and background

SCRI is a simplified version of the self-controlled case series (SCCS), including only vaccinated subjects with a short observation period. As such, it is more easily applied as a rapid signal strengthening activity. The validity of the SCRI design depends on three assumptions underlying the SCCS design: 1) independence of events within a person, 2) absence of event-dependent censoring of the observation period, 3) exposure independent of events (Whitaker H, et al. 2018). Furthermore, how the observation period itself is defined can have a large impact on event rates in the baseline time which then impacts the rate ratio. This can be important for outcomes with seasonal variation, especially if the vaccine is rapidly rolled out in a short period of time, which is the case for COVID-19 vaccination, as there is then little variability in season for the risk period. The lockdown in many countries may have led to another problem as well: events may be selectively more underreported in the control period before vaccination because patients were less likely to visit a health facility when having symptoms or complaints.

Research question and objectives

This study will evaluate assumptions of the SCRI design to assess the association between COVID-19 vaccines and adverse events of special interest in simulated datasets informed by electronic healthcare databases in European countries.

Study design

Simulation models will be built with a basic data generating mechanism and a procedure of simulation of violations of assumptions similar to the simulation model in Whitaker H, et al. 2018.

Population

Datasets will be simulated based on characteristics of populations included in the *Covid-Vaccine-Monitor Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol.

Variables

The parameters of the simulation model will be derived from the real-world datasets available from the *Covid-Vaccine-Monitor Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol

Data sources

Data sources participating in the *Covid-Vaccine-Monitor Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol

Study size

Sample sizes will be chosen according to the sample sizes that we typically will have in the datasets from the *Covid-Vaccine-Monitor Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol and that will depend on the type of event.

5. Amendments and updates

Number	Date	Section of study protocol	Amendment or update	Reason
N/A				

6. Milestones

Milestones for Covid-Vaccine-Monitor

Milestones and deliverables	Planned date
Contract signature	6 Apr 2021
Start of project	6 Apr 2021
D1 Study plan*	6 May 2021
D2 Study protocol(s)*	7 Jun 2021
Study start	7 July 2021
D4.1 Interim report*	6 Apr 2022
D4.2 Final report*	6 Apr 2023
D5 Manuscript*	6 Apr 2023

*Deliverable to be submitted to EMA

7. Rationale and background

SCRI is a simplified version of the self-controlled case series (SCCS), including only vaccinated subjects, studied over a short observation period. As such, SCRI is more easily applied as a rapid signal strengthening activity. The validity of the SCRI design depends on three assumptions underlying the SCCS design: 1) independence of events within a person, 2) absence of event-dependent censoring of the observation period, 3) exposure independent of events (Whitaker H, et al. 2018). Examples of violation of assumption 1) are for instance when recurrent events such as venous thromboembolism (VT) after COVID-19 vaccination are included. A history of VT is a risk factor for subsequent VTs and therefore these are not independent. Violation of assumption 2) might occur for events such as myocardial infarction, or TTS that have a high case-fatality. Examples of violation of assumption 3) include contraindications for COVID-19 vaccination such as anaphylaxis following COVID-19 vaccination. Other scenarios include fatal events which make post-event vaccination impossible, or events that put subjects in vaccine-eligible groups and increased the chance of vaccination post-event. Furthermore, how the observation period itself is defined, can have a large impact on event rates in the baseline time, which then impacts the rate ratio. This can be important for outcomes with seasonal variation, especially if the vaccine is rapidly rolled out in a short period of time, which is the case for COVID-19 vaccination, as there is then little variability in season for the risk period. An example of seasonal variation may be the occurrence of several solicited events associated with other seasonal viral infections such as the common cold or influenza. Although the SCRI design usually has short risk intervals, there still may be enough seasonal variation over the complete time span of the control and exposure intervals.

The lockdown in many countries may have led to another problem as well: events may be selectively more underreported in the control period before vaccination because patients were less likely to visit a health facility when having symptoms or complaints. The lockdown during the first COVID-19 infection wave in 2020 is not a problem because we will not include risk periods far before the start of vaccination. However, in several countries, a lockdown was also present during the second and third waves in 2021, when large-scale vaccination campaigns had been started. A formal statistical check on the potential effect of underreporting will be done using background rates of events in a pre-covid period to inform the simulation study. As this selective underreporting directly leads to a bias in the estimated relative incidence, we will not study it separately but only discuss its implications.

Assumption 1) is likely to be met, except maybe for some special events in which the first occurrence in an individual induces physiological changes that lead to change in the risk of subsequent events. A solution is then to use only the first event.

A previous study on the SCCS design has performed simulations to verify the impact of violations of assumptions 2) and 3) on the bias in the estimate of the relative incidence (Whitaker H, et al. 2018). That study also evaluated the power of statistical tests to identify violations of the assumptions. The violation of the assumptions was simulated separately for 2) and 3) and risk intervals,the proportion of violations, and the degree of violations were chosen from a range of theoretical values.

8. Research question and objectives

This study will evaluate assumptions of the SCRI design to assess the association between COVID-19 vaccines and adverse events of special interest in simulated datasets informed by electronic healthcare databases in European countries.

9. Research methods

9.1 Study design

Simulations will be performed to investigate the impact of violation of assumptions 2 and 3 on the estimated relative incidence. We will extend the previous simulation study by incorporating seasonal variation in the incidence of outcomes, and we will investigate how the choice of the length of the observation periods will affect the estimate of the relative incidence. Further, we will investigate the *combination* of violations of assumptions 2) and 3). We expect results to be potentially different from those on the SCCS design, since, in the SCRI, sampling is done on vaccinated subjects only, which could lead to exclusion of a subject, e.g., when assumption 3) is violated and the event has led to not vaccinating the subject. In the SCCS, such a subject will be included, and a biased estimate may result. Since the subject is not included in the SCRI design, the question is whether the SCRI is able to eliminate this bias. In the simulation, we will also make a comparison with an SCCS design in which nonvaccinated subjects are included as a result of the violation of assumption 3). The parameters of the simulation model will be derived from the real-world datasets available from the *Covid-Vaccine-Monitor Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol.

9.2 Setting

This study will be conducted as part of the CVM study but will rely on simulation to inspect the impact of changes in assumptions. Input parameters will be obtained from the analysis of EHR data sources that are included in the readiness phase of the *Covid-Vaccine-Monitor-Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol will be used to inform key parameters of the simulation studies.

9.3 Variables

The parameters of the simulation model will be derived from previous studies and from the output of the readiness phase of the *Covid-Vaccine-Monitor-Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol. For the scenario parameters that coincide with those in Whitaker 2018, we will start by using the same values as in that publication. These values will be replaced by the values we obtain from a previous study (ROC19). When we have

finished programming the simulation modelling and organizing the output, an alternative set of parameters for study-specific variables such as the incidence of adverse events and seasonality will be obtained from the output of the readiness phase of the *Covid-Vaccine-Monitor-Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol.

Table 1: Scenario dimensions and number of different values in each dimension

Scenario dimension	Number of scenario values
Sample Size	2
Time varying confounding due to medication use or covid-19 infection	2
Seasonality, strength/variation	2
Relative incidence of outcome due to vaccination	3
Length pre-vaccination control period	2
Length pre-vaccination exclusion period	2
Assumption 2), timing of censoring	2
Assumption 2), proportion of events leading to censoring	3
Assumption 3), delay or removal of vaccination	2
Assumption 3), timing of delay or removal	3
Assumption 3), proportion of events leading to delay or removal	3

The total number of scenarios is equal to the product of all the numbers in Table 1, i.e., the combination of each value in each row will be combined with the values of all other rows in the table.

9.4 Data sources

Nine data sources in 5 countries, participating in the *Covid-Vaccine-Monitor-Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol. Parameters will be estimated from each of those data sources.

Table 2: Participating data access providers and datasources

Country	Data Access Provider	Name Data source	Experience ConcePTION CDM v2.2	AESI experience	Active population	Type of data source	Sources for diagnoses	Pregnancy data	COVID-19 vaccine data	Lag time availability key outcome data
NL	PHARMO / UMCU	PHARMO	Yes	Yes (ACCESS)	6 million	Record linkage	GP, Hospital	Yes (perined) linkage, but requires specific approval	Yes, GP and potentially CIMS	Hospital: 1 year GP <3 month Perined: 1 year
ES	AEMPS	BIFAP	Yes	Yes (ACCESS)	8 million	GP medical records	GP & hospital (larger lag)	Yes	Yes	GP: 3-6 months Hosp: 1 year
ES	IDIAPJGol	SIDIAP	Yes	Yes (ACCESS)	5.7 million	Record linkage	GP, hospital	Yes	Yes	3-6 months
IT	SoSeTe	PEDIANET	Yes	Yes (ACCESS)	0.5 million	Pediatric medical record	Primary care	No	Not yet	< 3 months
IT	ARS Toscana	ARS data	Yes	Yes (ACCESS)	3.6 million	Record linkage	Hospital	Yes (those ending in delivery)	yes	3 months
IT	Lazio	Lazio data	No	No	5.8 million	Record linkage	Hospital, emergency visits	Yes (those ending in delivery)	yes	3 months
IT	INSPIRE srl	Caserta data	No	No	1 million	Record linkage	Hospital, emergency visits	yes (those ending in delivery)	Not yet	< 3 months
UK	Utrecht University	CPRD/HES GOLD	Yes	Yes (ACCESS)	16 million	GP & Hospital medical record	GP, Hospital	Yes	Yes, but no brands	3-6 months
NO	University Oslo	Norwegian	Yes	No	5 million	Record linkage	Hospital, outpatient, and GP	Yes	Yes	>6 months

9.5 Study size

Simulations will be done based on a range of the number of expected cases, given the background rate incidences of specific events of special interest in specific data sources and overall.

9.6 Data management

This is a simulation study only and it does not contain real data.

Input parameters will be obtained from the *Covid-Vaccine-Monitor-Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care data sources* protocol. Simulations will be created in R or SAS, and analysed on computers within the UMCU firewalls. Results may be transferred to the anDREa platform, for secure collaboration with partners.

9.7 Data analysis

Simulation models will be built with a basic data generating mechanism and a procedure of simulation of violations of assumptions similar to the simulation model in Whitaker H, et al. 2018.

The number of simulations will be chosen to obtain a Monte Carlo standard error (MCSE) of the estimates of bias that is less than 0.01 on a relative scale. Sample sizes will be chosen according to the sample sizes that we typically will have in the datasets from WP3 and that will depend on the type of event.

Parameters for the baseline incidence of outcome events will be chosen to reflect the real-world data from EHR data that will be generated for CVM in the readiness phase (separate protocol). Seasonal variation will be implemented using trigonometric functions, with varying amounts of variation, which will be based on patterns in the data sources from pre-pandemic times. The date of vaccination will be placed on varying moments within the year, but with a bandwidth derived from the real-world data. Vaccination-related relative incidences will be used to reflect a range from no to strong effects, similar to Whitaker H, et al. 2018.

From here onwards, scenarios will be implemented in which the length of the pre-vaccination control period will be varied, together with all possible combinations of violations of assumptions 2) and 3) as in (Whitaker H, et al. 2018). For assumption 2), event-dependent censoring of the observation period, different scenarios will be designed with regard to the timing of censoring, combined with different values for the proportion of events that will lead to censoring, as in (Whitaker H, et al. 2018). Likewise, for assumption 3), exposure independent of events, different scenarios will be designed with delay or removal of vaccination. The scenario of adding exposures, as used in Whitaker H, et al. 2018 will not be considered here, as vaccination for COVID-19 is scheduled in either one or two doses, according to the brand.

In all the principal scenarios, we have left the number of separate scenarios, as well as the values for parameters unspecified as yet. The precise number of scenarios and values for parameters will be determined in close cooperation with other WP's. In Table 1, we give an overview of all the parameters that will be varied in the simulation.

For the time-varying confounding, we will use the simulation model according to the design of a pre-vaccination control period and a risk period following vaccination. As a sensitivity analysis, we will

compare the results of the simulation with a simulation where an active comparator is used, as this seems to reduce bias due to time-varying confounding (Hallas J, Whitaker H, et al. 2021).

The impact of the violations on the results of standard statistical analyses, as well as on alternative analysis, will be evaluated. Comparisons will always be made between the scenarios and the baseline model, which contains no violation of assumptions, in the bias and Mean Squared Error (MSE) in the log relative incidence.

The results of the simulation studies with the many scenarios will be summarized in tables and graphically. Based on a margin of bias in the log relative incidence that is considered acceptable, we will identify which amount of violation in which assumptions are critical. These critical values will be fed into the analytical strategy of SCRI analysis The margin of bias and MSE that is considered acceptable will be determined by expert consensus among key members from WP3 and WP4. The amount of bias that is acceptable should be related to its impact on the statistical power and on the chance of false findings. We will employ the Delphi method, with a facilitator, anonymised experts, and multiple rounds to achieve consensus.

Recommendations for rapid SCRI conduct or alternative statistical analyses will be made based on the results from the simulation studies.

9.8 Quality control

Rigorous quality control (QC) will be applied to all deliverables. Standard operating procedures or internal process guidance will be used to guide the conduct of the study. These procedures include rules for secure and confidential data storage, backup, and recovery; methods to maintain and archive project documents; QC procedures for programming; standards for writing analysis plans; and requirements for scientific review by senior staff.

9.9 Limitations of the research methods

Simulation studies are an interpretation of real world scenarios and only a limited number of scenarios can be included, in order to reduce this limitation we will use parameter settings that are informed by real world data in the CVM study.

9.10 Other aspects

Not applicable

10. Protection of human subjects

Not applicable

11. Management and reporting of adverse events/adverse reactions

Not applicable

12. Plans for disseminating and communicating study results

The study protocol will be converted to a SAP and posted on the EU PAS register. Upon study completion and finalization of the study report, the results of this non-interventional study will be submitted for publication and posted in the EU PAS publicly accessible database of results. Publications will comply with the International Committee of Medical Journal Editors (ICMJE) guidelines.

13. References

Whitaker HJ, Ghebremichael-Weldeselassie Y, Douglas IJ, Smeeth L, Farrington CP. Investigating the assumptions of the self-controlled case series method. Stat Med. 2018 Feb 20;37(4):643-658. doi: 10.1002/sim.7536. Epub 2017 Nov 2. PMID: 29094391.

Hallas J, Whitaker H, Delaney JA, Cadarette SM, Pratt N, Maclure M. The Use of active Comparators in self-controlled Designs. Am J Epidemiol. 2021 Apr 16:kwab110. doi: 10.1093/aje/kwab110. Epub ahead of print. PMID: 33861309.

Annex 1: Role and contribution of WP4 in developing protocols

Role and contribution of WP4 in developing protocols, statistical analysis plans, conduct of statistical analyses and reporting with regard to protocols of WP1, WP2, and WP3.

Protocol WP1.

As per D1 study plan task 4.1 will participate in Observed/ Expected analyses by comparing observed rates with the rates in EHR data where possible.
WP4 will not work on its own, but with WP1 as statistical and methodological support.

Collaboration is shown in the Wp1 protocol:
Section 9.7.1. Main analysis: The statistical analysis plan will be developed jointly with WP1 and WP4 with special attention for control for confounding in comparisons of incidence rates between vaccine brands.

Section 9.7.2. Sensitivity analyses: This part of the statistical analysis plan will be developed by WP4 in collaboration with WP1.

Protocol WP2.

As per D1 study plan task 4.1, WP4 will participate in Observed/ Expected analyses by comparing observed rates with the rates in EHR data where possible

Section 7. Data Analysis
Main analysis: The statistical analysis plan will be developed jointly with WP1 and WP4 with special attention for control for confounding (channeling) in comparisons of aggregated incidence rates between vaccine brands.
Sensitivity analyses: This part of the statistical analysis plan will be developed by WP4 in collaboration with WP2.

Protocol WP3.

WP4 plays an important role in the work of WP3. WP4 will

- Specify the required format of the analytical data sets (D3-pipeline) that data engineers need to produce from the CDM to do all the work on the sensitivity analyses and methods work
- Write the R-scripts that runs on the D3 datatable for the analytical parts and the required sensitivity analyses for the negative control analyses (cohort and SCRI)
- Development of the SAP to detail the different sensitivity analyses and variable constructs

WP4 tasks 4.2-4.5 have been included in section 9.7.3 as detailed sensitivity analyses
We specified WP4 wherever we had included only methods group in the prior proposal, now it is indicated in 15 places in the protocol to be specific what they will contribute.

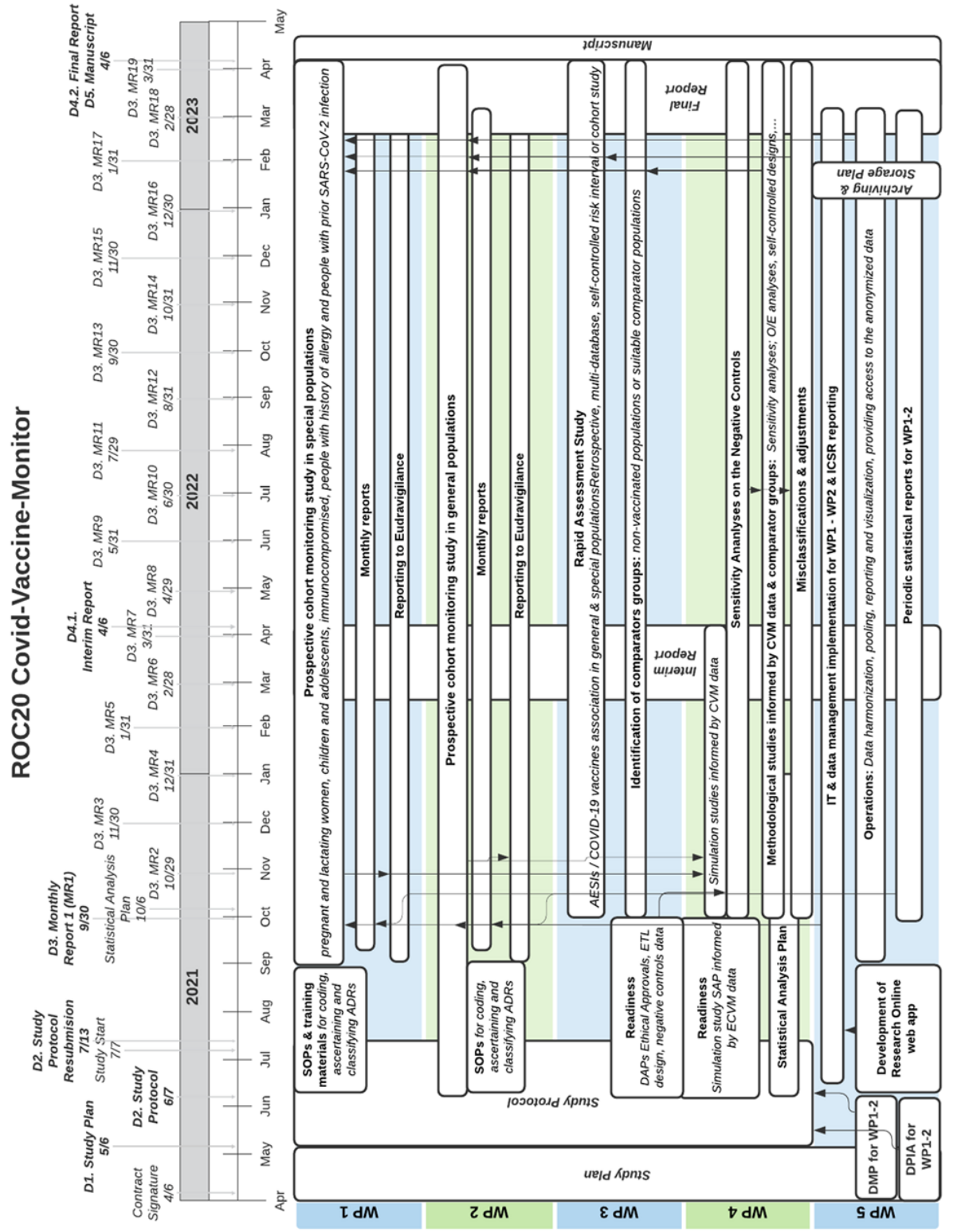
Section 9.7.1. Main summary measures
Comparators: the consideration of potential comparators, depending on the event of interest, are elaborated by WP4 in section 9.9. Other aspects and will be developed jointly with WP3 and WP4 as part of the statistical analysis plan.
Section 9.7.3. Sensitivity analyses: This part of the statistical analysis plan will be developed by WP4 in collaboration with WP3.

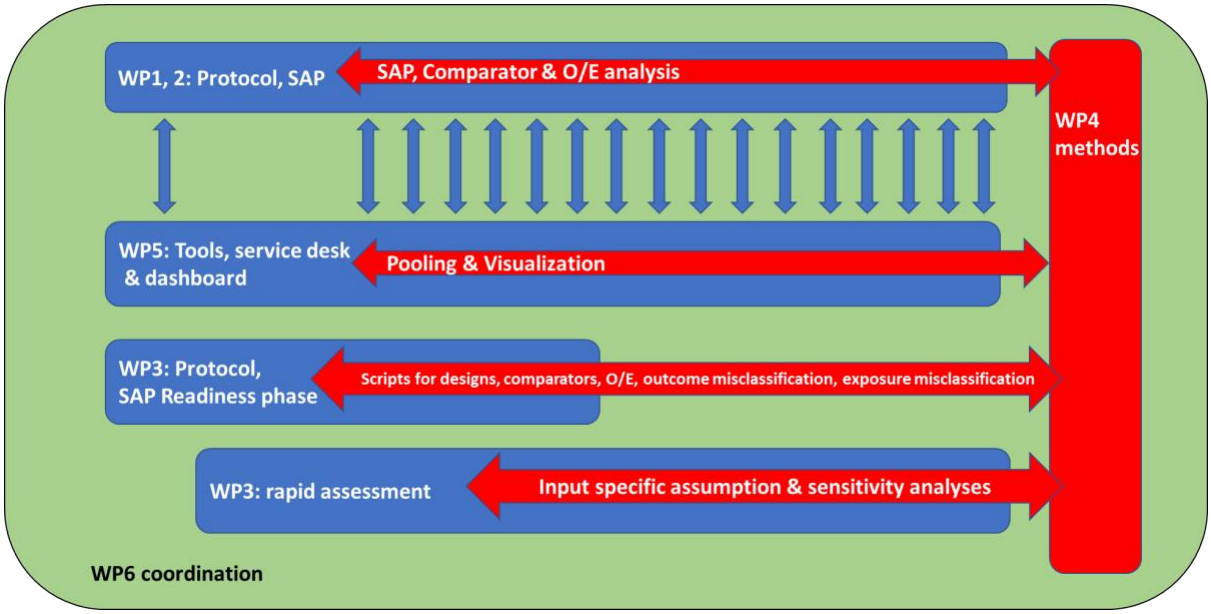
Section 9.9. Other aspects: Comparators and negative controls includes detailed considerations of selection of potential comparators and negative control outcomes to be considered in the readiness phase.

Protocol WP4.

This protocol details the work from task 4.3 that will be done on simulations and not on real data, all other work from WP4 on real data is part of WP1-3 protocols. The results from the simulation study will inform the analyses of WP3.

Annex 2: WPs collaboration and timelines scheme




WP1, 2: Protocol, SAP
SAP, Comparator & O/E analysis
WP4 methods
WP5: Tools, service desk & dashboard
Pooling & Visualization
WP3: Protocol, SAP Readiness phase
Scripts for designs, comparators, O/E, outcome misclassification, exposure misclassification
WP3: rapid assessment
Input specific assumption & sensitivity analyses
WP6 coordination