

Title	A Protocol for Real-World Effectiveness of COVID-19 Vaccines in the Netherlands: Causal Inference From A Retrospective Cohort Study Using the Target Trial Emulation framework Within the PharmLines Initiative
Research question & Objectives	What is the overall and variant specific vaccine effectiveness of monovalent COVID-19 vaccine in preventing SARS-CoV-2 infection
Protocol version	V4
Last update date	01 October 2026
Contributors	Primary investigator contact information: Stefan Verweij (stefan.verweij@rug.nl; +31646884054). Unit of Pharmacotherapy, -Epidemiology and -Economics (PTEE), Department of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands. Contributor names: Guiling Zhou; Katrien Oude Rengerink; Maarten J. Bijlsma; Spyros Balafas; Jens Bos; Mirjam Knol; Peter G.M. Mol; Eelko Hak.
Study registration	Site: n.a. Identifier: n.a.
Sponsor	Organization: University of Groningen Contact: Stefan Verweij (stefan.verweij@rug.nl; +31646884054). Unit of Pharmacotherapy, -Epidemiology and -Economics (PTEE), Department of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV, Groningen, The Netherlands
Conflict of interest	Stefan Verweij received funding from the Medicines Evaluation Board to conduct his PhD research at the University of Groningen, the Netherlands

Table of contents

2. Abstract	3
3. Amendments and updates	4
4. Milestones	4
5. Rationale and background	5
6. Research question and objectives	8
Table 1 Primary and secondary research questions and objective	8
7. Research methods	12
7.1. Target Trial specifications	12
Table 2 Components of the three target trials	12
7.2. Estimand	14
Table 3 Components of the estimand	14
7.3. Study design	16
7.4. Study design diagram	16
7.5. Setting	17
7.5.1 Context and rationale for definition of time 0 (and other primary time anchors) for entry to the study population	17
Table 4. Operational Definition of Time 0 (index date) and other primary time anchors	17
7.5.2 Context and rationale for study inclusion criteria:	18
Table 5. Operational Definitions of Inclusion Criteria	18
7.5.3 Context and rationale for study exclusion criteria	18
Table 6. Operational Definitions of Exclusion Criteria	18
7.6. Variables	18
7.6.1 Context and rationale for exposure(s) of interest	18
Table 7. Operational Definitions of Exposure	19
7.6.2 Context and rationale for outcome(s) of interest	20
Table 8. Operational Definitions of Outcome	20
7.6.3 Context and rationale for follow up	21
Table 9. Operational Definitions of Follow Up	21
7.6.4 Context and rationale for covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)	22
Table 10. Operational Definitions of Covariates	22
7.7. Data analysis	23
7.7.1 Context and rationale for analysis plan	23
Table 11. Primary, secondary, and subgroup analysis specification	24
Table 12. Sensitivity analyses – rationale, strengths and limitations	28
7.8. Data sources	30
7.6.1 Context and rationale for data sources	30
Table 13. Metadata about data sources and software	30
8. Limitation of the methods	31
9. Protection of human subjects	31
Appendix A	32

2. Abstract

Background: Estimating the real-world effectiveness of COVID-19 vaccines against SARS-CoV-2 infection is challenging because observational studies are susceptible to immortal time bias, depletion of susceptibles, and time-varying confounding. Furthermore, cohort and test-negative case-control (TNCC) designs each have their advantages and methodological limitations.

Objective: To estimate the effectiveness of monovalent COVID-19 vaccines against laboratory-confirmed SARS-CoV-2 infection using a sequential target trial emulation (TTE) framework, and to compare these estimates with those obtained from a previously conducted TNCC study using the same underlying population.

Methods: We will conduct a retrospective population-based new-user cohort study using linked individual-level data from the Lifelines cohort, Statistics Netherlands (CBS), and the IADB.nl pharmacy database. Daily target trials will be sequentially emulated between June 18, 2021 and October 19, 2022. Adults without prior SARS-CoV-2 infection will be assigned to partially vaccinated, fully vaccinated, booster-vaccinated, or unvaccinated strategies according to their eligibility at each trial start. The primary outcome is time to first laboratory-confirmed SARS-CoV-2 infection determined by RT-PCR testing. Generalized additive models will estimate time-varying hazard ratios, while pooled logistic regression will estimate odds ratios for predefined post-vaccination intervals. Inverse probability of treatment and censoring weights, multiple imputation, and prespecified sensitivity analyses will address confounding, informative censoring, and missing data.

Discussion: This study will provide real-world estimates of vaccine effectiveness against SARS-CoV-2 infection across the Delta and Omicron periods while illustrating the methodological advantages and limitations of sequential target trial emulation relative to the TNCC design. The findings will contribute to improving causal inference from observational vaccine effectiveness studies and inform future real-world evidence generation.

3. Amendments and updates

Date	Version	Section of protocol	Amendment or update	Reason
01-06-2026	V1	All	Creation of the protocol	n.a.
07-07-2026	V2	All	Feedback received from contributors	Improvement of protocol before publication to RWD catalogues.
17-07-2026	V3	All	Feedback implemented	Feedback on entire protocol: V3.1: Improved rationale V3.2: Determined frequency of sequential trial emulations V3.3: Fine-tuning intercurrent events V3.4: Last feedback on Rationale and target trial implemented
23-08-2026	V4	All	RWD ready to be published	Publication ready

4. Milestones

Milestone	Date
Finalize protocol	5 August 2026
Upload protocol to EMA's RWD catalogue	1 October 2026
Start data cleaning	1 October 2026
Start data analysis	8 October 2026
Finish data analysis	1 November 2026
Publish results	1 January 2027

5. Rationale and background

What is known about the condition and exposures of interest:

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2019 created significant global public health challenges, leading to a pandemic. In response, multiple vaccines were rapidly developed and evaluated through clinical trials and post-authorization studies^{1,2,3}. These studies, including our prior systematic review of COVID-19 vaccines, demonstrated short-term efficacy and effectiveness in preventing SARS-CoV-2 infection and severe outcomes such as hospitalization, ICU admission, and death^{4,5}.

Gaps in knowledge:

Accurately estimating vaccine effectiveness (VE) against SARS-CoV-2 infection remains complex. Contributing factors include the rapid emergence of new viral variants, changing levels of vaccine coverage, evolving population behavior, and the relatively short duration of immune protection^{6,7}. Variants such as Delta and Omicron, with higher transmissibility and partial immune escape, have further challenged vaccine-induced protection⁵. Everything considered, understanding how VE changes based on vaccination status (e.g., full vs. booster doses), circulating variants, and time since vaccination is essential to guide public health responses and inform vaccine policy.

Due to mass vaccination and testing campaigns during the COVID-19 pandemic in the Netherlands, a substantial amount of observational data has been generated in relation to vaccine rates and vaccine performance, which could be used to address the previously mentioned challenges in assessing VE^{8,9}. However, in general, these observational data are potentially prone to selection bias such as depletion of susceptibles, immortal time bias, and biases from unmeasured time varying confounding by indication. For example, for immortal time bias in vaccine studies, individuals that are initially unvaccinated can only become vaccinated if they survived until vaccination. All in all, it remains uncertain which study design offers the greatest methodological advantages.

Limitations of conventional study designs:

In the context of the previous example related to vaccine research, for cohort studies, we have to align their risk sets of getting infected: someone being vaccinated at the 100th day after t_0 can only be compared to an unvaccinated individual *still alive* at day 100, which introduces survivorship bias.

In case-control studies, the risks are aligned by design through proper case-control matching. However, these designs have their own strengths and limitations. They are prone to selection bias (controls *aka* non-cases should be taken from the same population that produced cases) and, in some cases, prone to recall bias (such as in questionnaire based studies where cases may recall their risks differently as compared to non-cases). Moreover, these designs are neither able to estimate absolute risks (due to the matching process, non-cases are a sample of the source population, thus we do not know the population at risk at baseline), nor examine multiple outcomes (a case is determined by a single outcome).

A derivative of the case-control design, the test-negative case-control design, was developed to avoid a specific type of selection bias, bias due to healthcare seeking behaviour, by including controls (non-cases) who also sought care (or at least took a test) with comparable symptoms, but have received a negative test result for the disease studied. We previously designed a test-negative case-control study¹⁰ to estimate COVID-19 vaccine effectiveness in participants of the Pharmlines initiative in the Northern part of the Netherlands, of which the results are currently submitted for peer-reviewed publication. However, the TNCC design in this particular study came with specific time-related challenges. For example, due to the particular matching-procedure in this study, we could not incorporate the time passed since most recent eligible COVID-19 vaccination as subgroup-analysis; since that would heavily affect the informativeness of the matched strata, where strata need at least comparative outcomes (case and non-cases) or comparative exposures. Using this particular variable as covariate was not possible either, due to structural dependence being introduced by the unvaccinated participants and 'time passed since most recent eligible COVID-19 vaccination' = *NA* (because they have never been vaccinated). Ultimately, an interaction effect

¹ Le, T.T., et al., *The COVID-19 vaccine development landscape*. Nat Rev Drug Discov, 2020. **19**(5): p. 305-306.

² Ndwandwe, D. and C.S. Wiysonge, *COVID-19 vaccines*. Current opinion in immunology, 2021. **71**: p. 111-116.

³ Henry, D.A., et al., *Effectiveness of COVID-19 vaccines: findings from real world studies*. The Medical Journal of Australia, 2021. 215(4): p. 149.

⁴ Pormohammad, A., et al., *Efficacy and safety of COVID-19 vaccines: a systematic review and meta-analysis of randomized clinical trials*. Vaccines, 2021. 9(5): p. 467.

⁵ Zhou, G., et al., *Effectiveness of COVID-19 vaccines against SARS-CoV-2 infection and severe outcomes in adults: a systematic review and meta-analysis of European studies published up to 22 January 2024*. European Respiratory Review, 2025. 34(175).

⁶ Haque, A. and A.B. Pant, *Mitigating Covid-19 in the face of emerging virus variants, breakthrough infections and vaccine hesitancy*. Journal of autoimmunity, 2022. 127: p. 102792.

⁷ Tregoning, J.S., et al., *Progress of the COVID-19 vaccine effort: viruses, vaccines and variants versus efficacy, effectiveness and escape*. Nature reviews immunology, 2021. 21(10): p. 626-636.

⁸ Statistics Netherlands (CBS), *Overledenen; Belangrijke Doorsoorzaken (Korte Lijst), Leeftijd, Geslacht*. 2024. Accessed on April 19th, 2025.

⁹ Lanooij, S., et al., *Vaccinatiegraad COVID-19 vaccinatie Nederland, 2022*. National Institute for Public Health (RIVM): 2024-03-12.

¹⁰ Verweij, S., et al., *Real-World Effectiveness of COVID-19 Vaccines in the Netherlands: A Test-Negative Case-Control Study within the PharmLines Initiative*. (to be published).

variable was created to estimate the VE per discrete time interval, making estimation of the overall VE impossible. Therefore, the proposed follow-up study has been designed following the sequential target trial emulation framework to compare how estimates differ from the aforementioned test-negative case-control study. The target trial emulation framework has been extensively described in literature and has been applied in other COVID-19 effectiveness studies^{11,12,13,14}. Using the TTE design, we not only avoid the time-related challenges due to matching in the TNCC design, but (using calendar-based stratification) may also identify the effect of survivorship bias, which, as mentioned above, is often seen in cohort studies not yet applying the TTE framework; and allow estimation of absolute risks, multiple outcomes and the overall VE. This study specifically aims to address how effect estimates differ from the previous TNCC study, to get better insights into the definition of the estimands, as defined in ICH E9 (R1)¹⁵, and the possible applications of this study design.

Rationale of the target trial:

Developing a thorough and complete target trial protocol in the context of rapidly evolving infectious disease outbreaks covering multiple consecutive dominant circulating viral variants – as was the case during the COVID-19 pandemic – is challenging due to complexities around the calendar period, and is therefore complex to define for TTE use-cases aiming to investigate the real-world vaccine effectiveness of COVID-19 vaccines over a longer period. The COVID-19-related VE in the real-world is strongly influenced by time-varying factors such as the dominant circulating viral variant, which is impossible to accurately predict and thus hard to capture in experimental trial scenarios. Therefore, to avoid these complexities, we will only describe a simplified version of the hypothetical target trial, using target trial components (such as the eligibility criteria, treatment strategies & assignment, follow-up, outcomes, causal contrast and statistical analysis) unrelated to calendar-related components, such as pandemic peaks and societal stringency measurements.

Another challenge that arises in this target trial scenario is our aim to estimate vaccine efficacy of both participants who received the primary series as well as booster-vaccinated participants, in accordance with the primary analysis of our previous TNCC (which compared both exposure types to the unvaccinated). Those who received a booster and those who completed only the primary series differ from each other, since booster participants should already have completed their primary series before trial enrollment to be considered eligible. To estimate the true effect of the booster, their control counterparts should therefore also have completed the primary series before trial enrollment. This makes them ineligible for comparison to the primary series arm, where participants are only eligible if they have *not* completed the primary series before trial enrollment; their control counterparts should therefore not have completed the primary series either, which leads to the existence of two separate control arms (or even two separate target trials). This issue – overlooked in our previous TNCC study – is described in detail in this TTE protocol within the target trial components (section 7.1, Table 2).

Rationale of the sequential target trial emulation

In accordance with the primary analysis of the previous TNCC study (which compared all vaccination statuses to the unvaccinated), our primary analysis aims to estimate VE of partially, fully and booster vaccinated as compared to the unvaccinated. Since this control group (i.e., the reference group) exists of unvaccinated participants, multiple timepoints arise on their timelines in which they will be eligible for participation in the trial emulation, as opposed to vaccinated individuals, who will only be eligible once for participation in the exposure group at the date of vaccination (i.e., the date of becoming partially vaccinated is the date of being eligible for the partially vaccinated group, the date of completing the primary series is the date of being eligible for the fully vaccinated group, and the date of the first booster vaccination is the date of being eligible for the booster vaccinated group). Therefore, to address the issue of multiple time-zeros for unvaccinated individuals, we will apply a sequential target trial emulation approach, where we will initiate a trial emulation on successive calendar dates of a given period (i.e., daily between July 2021 and October 2022). Each unvaccinated participant that is still unvaccinated on the date a new trial emulation starts will be assigned to the control group. If a participant gets vaccinated in the period thereafter, they will not be eligible anymore for the control group of the following trial emulation, but in turn will become eligible for the particular exposure group. Due to selecting June 18, 2021 as our start date, the positivity violation as observed in the TNCC study is avoided, since COVID-19 vaccines became available to all adult citizens of the Netherlands on that specific date.

The participants assigned to the primary series arm will only be compared to the unvaccinated participants, whereas the booster group will be compared to both those that have completed the primary series (in accordance to the target trial, as described in the previous

¹¹ Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8), 758-764.

¹² Ioannou, G. N., Locke, E. R., O'Hare, A. M., Bohnert, A. S., Boyko, E. J., Hynes, D. M., & Berry, K. (2022). COVID-19 vaccination effectiveness against infection or death in a national US health care system: a target trial emulation study. *Annals of Internal Medicine*, 175(3), 352-361.

¹³ Li, G., Gerlovin, H., Muñoz, M. J. F., Wise, J. K., Madenci, A. L., Robins, J. M., ... & Dickerman, B. A. (2024). Comparison of the test-negative design and cohort design with explicit target trial emulation for evaluating COVID-19 vaccine effectiveness. *Epidemiology*, 35(2), 137-149.

¹⁴ Huijberts, A. J., Smagge, B., van Werkhoven, H., de Gier, B., Monge, S., de Melker, H. E., ... & Knol, M. J. (2026). Comparison of methodological approaches in COVID-19 vaccine effectiveness estimation using observational data.

¹⁵ European Medicines Agency, *ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials*. 2020.

paragraph), and the unvaccinated participants (as to compare our results with a test-negative case-control study that utilized the same dataset, see the expected contribution below).

What is the expected contribution of this study?

In light of these considerations, to get an estimate of the effect of COVID-19 vaccination in the Northern part of the Netherlands, a retrospective cohort study which incorporates the sequential target trial emulation framework is designed to not only allow for a comparison with the test-negative case-control study utilizing the same dataset (study results are currently under submission at the time of publication of this protocol), but also to identify how a target trial emulation study differs in study design from a test-negative case-control design. The evidence generated with the proposed study, in combination with the previous results from our Test-Negative Case-Control study, can offer a comprehensive assessment of vaccine effectiveness against both SARS-CoV-2 infection and severe COVID-19 disease, spanning from the Delta tot the Omicron variants.

6. Research question and objectives

Table 1 Primary and secondary research questions and objective

A. Primary research question and objective: Time to infection

Objective:	To study the effect of different monovalent COVID-19 vaccine strategies in preventing lab-confirmed SARS-CoV-2 infection.
Hypothesis:	H_0 There is no difference in the hazard of COVID-19 infection over time between individuals assigned to one of the three vaccination strategies and those assigned to remain unvaccinated, H_1 There is a difference in the hazard of COVID-19 infection over time between individuals assigned to one of the three vaccination strategies compared to those that were assigned to remain unvaccinated.
Population (mention key inclusion-exclusion criteria):	The study population selection criteria are similar to our previous TNCC study. Participants are adults in the northern regions of the Netherlands. Key inclusion criteria are: - Adult respondents to the Lifelines COVID-19 questionnaire 15b or successive COVID-19 questionnaires. We selected this population, as from assessment 15b onwards (November 2, 2020), the question "were the tests performed by the GGD (municipal health service, Gemeentelijke gezondheidsdienst)" has been included in the questionnaire. - Respondents should be at least 18 years of age at their time-zero. - They should not have had a registered COVID-19 infection prior to COVID-19 questionnaire 15b.
Exposure:	Monovalent COVID-19 vaccines, regardless of vaccine type. Bivalent COVID-19 vaccines are not included because they were authorized across the EU September 1, 2022 at earliest, while our data collection is only until October 19, 2022, therefore there is not enough follow-up. Participants will be categorized as: Partially vaccinated: if they received one dose of a two-dose vaccine (Spikevax [Moderna], Comirnaty [BioNTech/Pfizer], Vaxzevria [AstraZeneca]). Fully vaccinated: if they received both doses of a two-dose vaccine at least 14 days after receiving the first dose, or one dose of a single-dose vaccine (COVID-19 vaccine Janssen). Booster vaccinated: if they received an additional booster dose at least 3 months after receiving full vaccination. For those who have been vaccinated, the date of vaccination, the number of doses, the type of vaccines (Moderna, Pfizer/BioNTech, Janssen, AstraZeneca), and time since last dose at the index date will be primarily collected from Statistics Netherlands (CBS).
Comparator:	Unvaccinated: if they have not received any dose of a COVID-19 vaccine.
Outcome:	The primary outcome will be the first SARS-CoV-2 infection determined by a positive RT-PCR test at the GGD.
Time (when follow up begins and ends):	The index date of each participant will be the first occurrence in the cohort in the case the eligibility criteria have been met. For the vaccinated, the time-zero will be the date the date of their eligible vaccination (when a daily emulation starts). For the unvaccinated, the time-zero will be the date the sequential trial emulation starts if they remain unvaccinated on that date. Follow-up ends after first positive RT-PCR test, death, loss to follow-up, or end of data collection, whichever occurs first.
Setting:	Population based.

B. Secondary research question and objective: Odds of infection

Objective:	Same as Table 1A.
Hypothesis:	H_0 There is no difference in the odds of COVID-19 infection between individuals, assigned to one of the three vaccination strategies and those assigned to remain unvaccinated. H_1 The odds of infection are lower among individuals assigned to one of the three vaccination strategies compared to those that were assigned to remain unvaccinated.
Population (mention key inclusion-exclusion criteria):	Same as Table 1A.
Exposure:	Same as Table 1A.
Comparator:	Same as Table 1A.
Outcome:	Same as Table 1A.
Time (when follow up begins and ends):	Same as Table 1A.
Setting:	Same as Table 1A.

C. Secondary research question and objective: Variant-specific odds of infection

Objective:	To study the effect of different monovalent COVID-19 vaccine strategies in preventing SARS-CoV-2 infection caused by different viral variants.
Hypothesis:	H_0 There is no difference in the odds of variant-specific COVID-19 infection between individuals assigned to one of the three vaccination strategies and those assigned to remain unvaccinated, H_1 The odds of variant-specific COVID-19 infection are lower among individuals assigned to one of the three vaccination strategies compared to those that were assigned to remain unvaccinated.
Population (mention key inclusion-exclusion criteria):	Same as Table 1A.
Exposure:	Same as Table 1A.
Comparator:	Same as Table 1A.
Outcome:	The primary outcome will be the first SARS-CoV-2 infection determined by a positive RT-PCR test at the GGD, where infections will be further categorized according to variant as ¹⁶ ▪ delta infections if the positive test date was between July 5, 2021 and December 19, 2021 (more than 80% cases were delta infections) ▪ omicron BA.1 or BA.2 infections if the positive test date between January 3, 2022 and June 6, 2022 (more than 85% cases were BA.1 or BA.2 infections) ▪ omicron BA.4 or BA.5 infections if the positive test date between June 20, 2022 and October 19, 2022 (more than 85% cases were BA.4 or BA.5 infections) For positive test dates falling outside the boundaries of the intervals described above (e.g., February 4 th , 2021) no clear dominant strain could be identified. These incidents will be categorized as n/a, see 8. Limitations 8.1.
Time (when follow up begins and ends):	Same as Table 1A.
Setting:	Same as Table 1A.

¹⁶ Rijksinstituut voor volksgezondheid en milieu (RIVM). Varianten van het coronavirus SARS-CoV-2. Published 2022. Accessed October 5, 2023. <https://www.rivm.nl/coronavirus-covid-19/virus/varianten>

D. Secondary research question and objective: Odds of infection for the booster vaccinated with fully vaccinated as control

Objective:	To study the effect of monovalent COVID-19 booster vaccine strategies in preventing SARS-CoV-2 infection.
Hypothesis:	H_0 There is no difference in the odds of COVID-19 infection between booster vaccinated individuals and fully vaccinated individuals, H_1 The odds of infection are lower among booster vaccinated individuals as compared to the fully vaccinated individuals.
Population (mention key inclusion-exclusion criteria):	. Participants are vaccinated adults in the northern regions of the Netherlands. Key inclusion criteria are: ▪ Adult respondents to the Lifelines COVID-19 questionnaire 15b or successive COVID-19 questionnaires. We selected this population, as from assessment 15b onwards (November 2, 2020), the question “were the tests performed by the GGD (municipal health service, Gemeentelijke gezondheidsdienst)” has been included in the questionnaire. ▪ Respondents should be at least 18 years of age during their time-zero. ▪ They should not have had a registered COVID-19 infection prior to COVID-19 questionnaire 15b. ▪ Completed their primary series (received one dose of a single-dose vaccine or the second dose of a two-dose vaccine).
Exposure:	Monovalent COVID-19 vaccines, regardless of vaccine type. Bivalent COVID-19 vaccines are not included because they were authorized across the EU September 1, 2022 at earliest, while our data collection is only until October 19, 2022, therefore there is not enough follow-up. Participants will be categorized as: ▪ Booster vaccinated: if they received an additional booster dose at least 3 months after receiving full vaccination. For those who have been vaccinated, the date of vaccination, the number of doses, the type of vaccines (Moderna, Pfizer/BioNTech, Janssen, AstraZeneca), and time since last dose at the index date will be primarily collected from Statistics Netherlands (CBS).
Comparator:	Monovalent COVID-19 vaccines, regardless of vaccine type. Bivalent COVID-19 vaccines are not included because they were authorized across the EU September 1, 2022 at earliest, while our data collection is only until October 19, 2022, therefore there is not enough follow-up. Participants will be categorized as: ▪ Fully vaccinated: if they received both doses of a two-dose vaccine at least 14 days after receiving the first dose, or one dose of a single-dose vaccine (COVID-19 vaccine Janssen). For those who have been vaccinated, the date of vaccination, the number of doses, the type of vaccines (Moderna, Pfizer/BioNTech, Janssen, AstraZeneca), and time since last dose at the index date will be primarily collected from Statistics Netherlands (CBS).
Outcome:	Same as Table 1A.
Time (when follow up begins and ends):	Same as Table 1A.
Setting:	Same as Table 1A.

E. Meta research question and objective

Objective:	To what extent can the estimates obtained under this target trial emulation (protocol described here) be meaningfully compared to estimates obtained from the previous test-negative case-control (TNCC) study performed similar in all aspects except the study design and its design-specific derivatives?
Hypothesis:	H_0 Estimates from the retrospective cohort study using the sequential TTE approach are similar to the estimates from the counterpart study adapting the TNCC design when it comes to statistical metrics (i.e., statistical significance, estimate agreement and standardized mean difference agreement). H_1 Estimates from the retrospective cohort study using the sequential TTE approach differ from the estimates obtained from the counterpart study adapting the TNCC design on the statistical measures (i.e., statistical significance, estimate agreement or standardized mean difference agreement).

7. Research methods

7.1. Target Trial specifications

The target trial components as given in table 2 below are independent of the calendar period. Any trial component affected by calendar time is not included: The COVID-19-related VE in the real-world is strongly influenced by time-varying factors such as the dominant circulating viral variant, which is not only impossible to predict and capture in experimental trial scenarios (e.g., it is near impossible to prospectively predict a future dominantly circulating COVID-19 variant), although assumptions have to be made, but also unethical for the control group (either placebo or unvaccinated) to remain unvaccinated when vaccine efficacy for the original variant has already been determined and vaccination offered.

Table 2 Components of the three target trials

Target trial component	Definition
Eligibility	Adult participants, at least 18 years of age.
Treatment strategy	Monovalent COVID-19 vaccines, regardless of vaccine type. Participants will be randomized to: <i>Rationale of RCT1:</i> A target trial is designed that compares partially vaccinated, fully vaccinated, and booster vaccinated against a single control (reference) group: placebo (akin to receiving the first dose of a two-dose vaccine). We acknowledge that comparing fully and booster vaccinated to this placebo group would lead to complexities in the treatment assignment & randomization process (as those assigned to the booster group should have completed the primary series before trial start, whereas those assigned to placebo are eligible right away), and would clinically be less relevant. However, RCT1 is designed to allow comparison of results to our previous TNCC study. ▪ Arm1 Partially vaccinated: receiving one dose of a two-dose vaccine on t0, ▪ Arm2 Fully vaccinated: receiving one dose of a single-dose vaccine or the second dose of a two-dose vaccine on t0 at least 14 days after receiving the first dose, ▪ Arm3 Booster vaccinated: receiving the first dose of a booster vaccine on t0 at least 3 months after completing the primary series, ▪ Control (reference): Receiving placebo (akin to first dose of a two-dose vaccine) on t0. <i>RCT2: Booster vaccinated</i> <i>Rationale of RCT2:</i> As acknowledged in target trial RCT1, the placebo group of RCT1 may lead to biased outcome estimates for the fully and booster vaccinated. For example, those eligible assigned to the booster group should have survived long enough to complete the primary series and may differ in (healthcare seeking) behavior. Therefore, RCT2 is designed to compare booster vaccinated against a properly defined control reference to quantify how the effectiveness of the booster differs from RCT1. ▪ Booster vaccinated: receiving the first booster dose on t0, after completing the primary series at least 3 months before t0. ▪ Control (reference): receiving full vaccination at least 3 months before t0 and placebo on t0.
Treatment assignment	Participants will be assigned to either exposure or control (reference) control group in a 1:1 ratio
Follow-up	Participants will be followed from date to randomization until: ▪ A positive RT-PCR test result, ▪ Death, ▪ Protocol deviation (receiving vaccination), ▪ Loss to follow up, ▪ Withdrawing consent, whichever occurs first.
Outcomes	The primary endpoint will be the first symptomatic SARS-CoV-2 infection determined by a positive RT-PCR test.
Causal contrast	Hazard Ratios, PP
Statistical analysis	Generalized additive model – similar to the estimator used in our TNCC study – using smoothing splines to examine the variation over time of vaccine protection. The parametric part will include linear effects from the following covariates: age at baseline, sex, ethnicity, smoking status, alcohol consumption, body mass index (BMI), health-cautiousness behaviour, marital status, educational level, comorbidities and the usage of repurposed drug and drugs related to high-risk comorbidities. Time passed since vaccination will be modelled non-parametrically using penalized thin plate regression splines.

7.2. Estimand

Table 3 Components of the estimand

Population (who and at what time)	Contrasted treatments (what, when, and how)	Endpoint (what, when, and how)	Intercurrent events (what and corresponding strategy)	Population-level summary measure	Estimand started as a research question	Example of a treatment decision to be informed by the estimand
<i>Time to infection (Table 1A)</i>						
Adult participants, at least 18 years of age.	Unvaccinated participants have not received any dose of a COVID-19 vaccine before baseline; Partially vaccinated participants received their first dose of a two-dose vaccine at t0; Fully vaccinated participants received the second dose of a two-dose vaccine or a single dose of a single-dose vaccine at t0; Booster vaccinated participants received their first additional booster dose at t0;	Time to SARS-CoV-2 infection, where the infection is determined by a positive RT-PCR test.	- New COVID-19 vaccination: <i>hypothetical</i> - Dispensing of non-vaccine medicinal products <i>treatment policy</i>	Time-varying hazard ratios describing the relative hazard of COVID-19 infection over time while being vaccinated (distinguishing between the three vaccination groups), versus while being unvaccinated.	What is the difference the hazard if all adults had been vaccinated (distinguishing between the three exposure groups) compared to if they all remained unvaccinated.	Advising if timely vaccination prolongs the time to infection;
<i>Probability of infection (Table 1B)</i>						
Adult participants, at least 18 years of age.	Unvaccinated participants have not received any dose of a COVID-19 vaccine before baseline; Partially vaccinated participants received their first dose of a two-dose vaccine at t0;	SARS-CoV-2 infection determined by a positive RT-PCR test.	- New COVID-19 vaccination: <i>hypothetical</i> - Dispensing of non-vaccine medicinal products <i>treatment policy</i>	Odds ratios indicating the odds of COVID-19 infection while being vaccinated (distinguishing between the three exposure groups), versus the odds	What are the odds of COVID-19 infection if all adults had been vaccinated (distinguishing between the three vaccination groups) compared to if they all remained unvaccinated	Advising if vaccination reduced the odds of COVID-19 infection.

	Fully vaccinated participants received the second dose of a two-dose vaccine or a single dose of a single-dose vaccine at t0; Booster vaccinated participants received their first additional booster dose at t0;			of COVID-19 infection while being unvaccinated; representing the average treatment effect (ATE).		
Probability of variant-specific infection (Table 1C)						
See the estimand for <i>probability of infection</i>						
Probability of infection for the booster vaccinated (Table 1D)						
Fully and booster vaccinated adult participants, at least 18 years of age.	(Exposure) Booster vaccinated participants received their first additional booster dose at t0, and had received the second dose of a two-dose vaccine or a single dose of a single-dose vaccine at least 14 days before t0. (Control) Fully vaccinated participants received the second dose of a two-dose vaccine or a single dose of a single-dose vaccine at least 14 days before t0;	SARS-CoV-2 infection determined by a positive RT-PCR test.	- New COVID-19 vaccination: <i>hypothetical</i> - Dispensing of non-vaccine medicinal products <i>treatment policy</i>	Odds ratios indicating the odds of COVID-19 infection while being booster vaccinated versus the odds of COVID-19 infection while remaining fully vaccinated; representing the average treatment effect (ATE).	What are the odds of COVID-19 infection if all adults had been booster vaccinated compared to if they all remained fully vaccinated	Advising if booster vaccinations reduce the odds of COVID-19 infection.

7.3. Study design

Research design (e.g. cohort, case-control, etc.): Retrospective cohort study using sequential trial design.

Rationale for study design choice: This study design reduces risk of bias from unmeasured time varying confounding by indication, i.e., depletion of susceptibles and immortal time bias by emulating daily sequential target trials over the calendar period. Patients will be followed from t0 until censoring (death, new vaccination, end of follow-up or end of data collection) or outcome.

7.4. Study design diagram

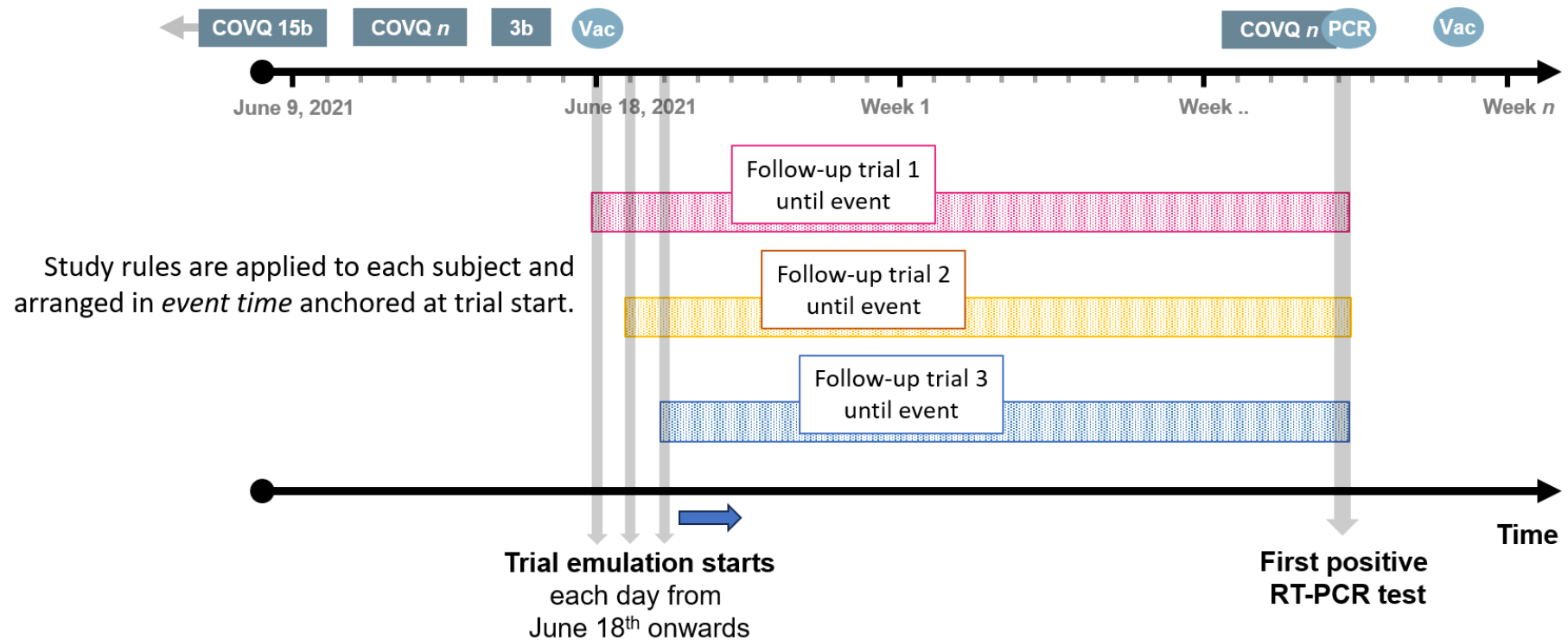


Figure 1. Individual-patient data is retrieved from Lifelines and CBS and documented as encounters from various sources: Lab-confirmed SARS-CoV-2 infection (PCR), vaccination with COVID-19 vaccine (Vac), or filled in Lifelines assessment form 1a-3b (1a-3b) or Lifelines COVQ assessment form 15b-29 (COVQ n). Timelines were arranged in days. Example for the first three trial emulations in this sequential approach given. Follow-up until first RT-PCR, new vaccination, L2FU, death or end of data collection, whichever comes first.

7.5. Setting

7.5.1 Context and rationale for definition of time 0 (and other primary time anchors) for entry to the study population

Time 0

In this sequential trial emulation approach, daily trial emulations will start from June 18, 2021.

Limitations around time 0

Due to limitations of the data, when questionnaire responses related to eligibility criteria and/or baseline characteristics are spread out over multiple different questionnaires (see 7.4.4 Context and rationale for covariates), we assume baseline covariates remain constant over time (see 8. Limitations 8.2). Participants are followed until loss to follow-up (i.e., , the outcome, end of data collection or a censoring event (e.g., failure to adhere to treatment strategy).

Due to the gradual introduction of vaccinates based on participants their birth-year, the study period will start from June 18, 2021 onwards. This avoids violation of positivity, but as a drawback introduces selection bias, see Limitations 8.3.

Table 4. Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Care Setting	Code Type	Incident with respect to...	Source of algorithm
The study population are adults in the northern regions of the Netherlands.	First day of the calendar month starting from June 18, 2021.	Single entry for each vaccinated exposure group according to the sequential TTE definition. Multiple entries for the unvaccinated.	Incident (vaccination)	Population based	Anonymized citizen service number (BSN) in Lifelines and CBS, where applicable also IADB	Remaining unvaccinated.	n/a

7.5.2 Context and rationale for study inclusion criteria:

We will include adult respondents to the COVID-19 questionnaires. As the COVID-19 questionnaires were only sent to Lifelines participants aged > 18 years, only adults are included. From assessment 15b onwards (November 2, 2020), the question “were the tests performed by the GGD (municipal health service, Gemeentelijke gezondheidsdienst)” has been included in the questionnaire, and is considered the most reliable test among all other tests. Our study will therefore only include individuals who have answered to COVQ 15b or later questionnaires, also to improve comparability against the TNCC study, which selected participants on tests from COVQ 15b onwards as well.

Table 5. Operational Definitions of Inclusion Criteria

Criterion	Details	Order of application	Assessment window	Care Settings	Source for algorithm
Response to COVID-19 questionnaire	Each participant should have at least replied to COVQ 15B or later questionnaires. See Table 11 and 8. Limitations 8.4.	Before selection of index date	[COVQ15b, COVQ 29]	Any (inpatient, outpatient etc.)	Investigator confirmed answer to RT-PCR question in COVQ 15b – 29

7.5.3 Context and rationale for study exclusion criteria

See 7.3.2. To get an estimate of the vaccine effectiveness in preventing SARS-CoV-2 infection of the general adult population in the northern part of the Netherlands, few exclusion criteria will be applied, as this will lead to a study population that more closely represents the real-world population.

Table 6. Operational Definitions of Exclusion Criteria

Criterion	Details	Order of application	Assessment window	Applied to study populations:	Source for algorithm
Prior infection	Participants who have reported SARS-CoV-2 infection prior to t0 will be excluded. See Table 11 and 8. Limitations 8.5.	Before selection of index date	[COVQ1, t0]	All	PCR confirmed, answer to test/infection questions in COVQ1-15

7.6. Variables

7.6.1 Context and rationale for exposure(s) of interest

The exposure of interest is being (partially) vaccinated with a two-dose COVID-19 vaccine (Spikevax-Moderna, Comirnaty-BioNTech/Pfizer, Vaczevria-AstraZeneca), or a single dose COVID-19 vaccine (Jcovden, also known as COVID-19 vaccine Janssen), or receiving the booster COVID-19 vaccine after completing the primary series.

Algorithm to define duration of exposure effect:

The exposure effect is defined as having received a COVID-19 vaccine dose with one of the previously mentioned COVID-19 vaccines, registered by CBS, as described in detail in Table 7 below.

Table 7. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting ¹	Applied to study populations:	Incident with respect to...	Source of algorithm
Booster vaccinated	Date of the first received booster vaccination, similar to t0.	n/a	[0, 0]	n/a	All as defined in 7.3.1, Table 3	...being an additional booster dose at least 3 months after receiving the second dose of a two-dose COVID-19 vaccine [‡] or one dose of a single dose COVID-19 vaccine [†] .	Investigator confirmed incident with respect to previous vaccinations. See “Incident with respect to...”
Fully vaccinated	Date of receiving the second-dose of a two-dose vaccine [‡] , or the sole dose of a single-dose vaccine [†] , similar to t0.	n/a	[0, 0]	n/a	All as defined in 7.3.1, Table 3	...being the second dose of a two-dose COVID-19 vaccine [‡] at least 14 days after receiving the first dose, or one dose of a single dose COVID-19 vaccine [†] .	Investigator confirmed incident with respect to previous vaccinations. See “Incident with respect to...”
Partially vaccinated	Date of receiving the first dose of a two-dose vaccine [‡] , similar to t0	n/a	[0, 0]	n/a	All as defined in 7.3.1, Table 3	...being the first dose of a two-dose COVID-19 vaccine [‡] .	n/a
Unvaccinated	Any date where the eligibility criteria are met.	n/a	[0, 0]	n/a	All as defined in 7.3.1, Table 3	...not having received any dose of COVID-19 vaccines ^{†,‡} .	n/a

[†] Single-dose COVID-19 vaccine being COVID-19 vaccine Janssen.

[‡] Two-dose COVID-19 vaccines being Spikevax (Moderna), Comirnaty (BioNTech/Pfizer), Vaczevria (AstraZeneca).

7.6.2 Context and rationale for outcome(s) of interest

The primary outcome will be the first laboratory-confirmed SARS-CoV-2 infection determined by a positive RT-PCR test at the GGD.

Table 8. Operational Definitions of Outcome

Outcome name	Details	Primary outcome?	Type of outcome	Care Settings ¹	Diagnosis Position ²	Applied to study populations:	Source of algorithm
SARS-CoV-2 infection	laboratory-confirmed SARS-CoV-2 infection determined by a positive RT-PCR test at the GGD.	Yes	Time-to-event, binary	Any	Primary only	All as defined in 7.3.1, Table 3	Investigator confirmed answer to RT-PCR question in COVQ 15b – 29.

¹ IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

² Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

7.6.3 Context and rationale for follow up

Follow-up starts at time0 (the index date of the specific trial emulation), as described in 7.5.1 Table 4, and ends at the first registered RT-PCR test, death, administrative or artificial censoring, whichever comes first.

Table 9. Operational Definitions of Follow Up

Follow up start	Day 1	
Follow up end ¹	Select all that apply	Specify
Date of outcome	Yes	See 7.4.2
Date of death	Yes	Registry recorded death
End of observation in data	Yes	Censor, L2FU or end of data collection
Day X following index date (specify day)	No	n/a
End of study period (specify date)	Yes	19 October 2022
End of exposure (specify operational details, e.g. stockpiling algorithm, grace period)	No	n/a
Date of add to/switch from exposure (specify algorithm)	Yes	Considered as censoring event, see the intercurrent events in Table 3.
Other date (specify)	No	n/a

¹ Follow up ends at the first occurrence of any of the selected criteria that end follow up.

7.6.4 Context and rationale for covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)

We select the same covariates as used in the TNCC study. Covariates in the TNCC study were selected based on available literature during the time of TNCC study design combined with available variables as provided in the data dictionaries of Lifelines, IADB.nl and Statistics Netherlands.

Table 10. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings	Code Type	Source for algorithm
Sex	Female, Male – at baseline.	Binary	[0,0]	Any		n/a
Age	Age at baseline	Continuous	[0,0]	Any		n/a
Ethnicity	European, African/Afro-American, Asian, Other	Categorical	[Most recent survey response, 0]	Any		Investigator confirmed answer to Lifelines 1B
Educational level	Low (no education; primary education; lower or secondary vocational education; junior general secondary education) Middle (secondary vocational education or work-based learning pathway; senior general secondary education or pre-university secondary education) High (higher vocational education; university education) Other	Categorical	[Most recent survey response, 0]	Any		Investigator confirmed answer to COVQ, Lifelines 2A, and/or Lifelines 3A
Marital status	Single, registered partnership, married or other	Categorical	[Most recent survey response, 0]	Any		Investigator confirmed answer to COVQ, Lifelines 2A, and/or Lifelines 2B
BMI	Underweight=BMI < 18.5, Normal=BMI 18.5-24.9, Overweight=BMI 25-29.9, Obese=BMI ≥ 30,	Continuous, Categorical	[Most recent survey response, 0]	Any	cm/kg	Investigator confirmed answer to Lifelines 2A, Lifelines 3A (cm) and COVQ (kg). Conversion done by investigator
Currently smoking	Smoking at baseline (1=yes, 0=no)	Binary	[Most recent survey response, 0]	Any		Investigator confirmed answer to Lifelines 1A, 1B, 1C, 2A, 2B, 3A
Alcohol usage	Low, moderate or high.	Categorical	[Most recent survey response, 0]	Any		Investigator confirmed answer to COVQ and/or Lifelines 3A
Comorbidities	Hypertension, high cholesterol, other CVD; Asthma/COPD, other lung diseases; Depression, other psychological diseases; Neurological diseases; Diabetes	Binary for each comorbidity	[Most recent survey response, 0]	Any	ATC-codes and patient-reported	Investigator confirmed answer to COVQ (and n/a for IADB data)
drug related to high-risk comorbidities and repurposed drug usage ³	Number of drugs used at baseline, divided into chronic or acute drugs.	Continuous	[Most recent survey response, 0]	Any	ATC-codes	n/a

Characteristic	Details	Type of variable	Assessment window	Care Settings	Code Type	Source for algorithm
Health-cautiousness behaviour	Cautious (go for test when mild symptoms occur) or not cautious (do not go for test when mild symptoms occur)	Binary	[0,0]	Any		Investigator confirmed answer to COVQ
Lockdown	Time-varying covariate indicating the stringency of COVID-19 measurements in the Netherlands in 2021 and 2022.	Time-varying index (0, 100)	[0, end]	Any		Oxford Coronavirus Government Response Tracker (OxCGRT)

7.7. Data analysis

7.7.1 Context and rationale for analysis plan

- We apply Inverse Probability of Censor (IPC) weights to adjust for the introduced informative censoring of getting vaccinated after time0 aka those who deviate from the assigned treatment strategy – by weighting the data. IPC upweights those that remain uncensored to get a population that better represent what would have happened without informative censoring.
- We apply Inverse Probability of Treatment (IPT) weights to balance the four exposure groups on their observed covariates at baseline.
- For studying the time-variation of vaccine protection generalized additive models (GAMs) with smoothing splines will be fitted to the data using the mgcv package in R. The parametric part of the model included linear effects from potential confounders while time elapsed since the index date will be modelled non-parametrically using penalized thin plate regression splines.
- As we would like to compare the results of the retrospective cohort study to the results of the test-negative case control study, which used a conditional logistic regression model yielding odds ratios across post-vaccination time-intervals (14-28 days since vaccination, 29-82 days, 83-181 days, 181-364 days and 365+ days), our primary analysis will make use of a pooled logistic regression model using the same discrete time intervals, since these yield odds ratios. These can then be converted to vaccine effectiveness percentages.
- To account for missing covariate data, the models described above will be fitted separately in each of the multiply imputed datasets.
 - o For the GAM, pooling spline coefficients for penalized smooths can be affected by the smoothing parameter estimation. To avoid this, we will generate model-based predictions and corresponding standard errors for each imputed dataset. At each time elapsed since the index date, the predicted linear-predictor contrasts and their within-imputation variances will be combined across imputations using Rubin's rules. The pooled effects are then exponentiated to obtain the final odds ratios. Pointwise 95% confidence intervals will be obtained using the pooled total variance from Rubin's rules.

Table 11. Primary, secondary, and subgroup analysis specification

A. Primary analysis (Time to infection)

Exposure contrast:	The different vaccination strategies (partial, full and booster) as compared to unvaccinated participants.
Outcome:	The outcome of interest is time to SARS-CoV-2 infection, defined by the date of a positive RT-PCR result taken by the Dutch GGD.
Analytic software:	R
Model(s): (provide details or code)	The individual patient data for each daily trial emulation will be pooled into a single dataset in long format. Each row representing a patient at time t (a continuous variable) since time 0 for trial k, a categorical column representing the vaccination status, a binary column for the outcome (infected=0;1) and the covariates. For studying the time-variation of vaccine protection, generalized additive models (GAMs) with smoothing splines will be fitted to the data using the mgcv package in R. The parametric part of the model included linear effects from potential confounders while time elapsed since the index date will be modelled non-parametrically using penalized thin plate regression splines. Pooling spline coefficients for penalized smooths can be affected by the smoothing parameter estimation. To avoid this, we will generate model-based predictions and corresponding standard errors for each imputed dataset. At each time elapsed since the index date, the predicted linear-predictor contrasts and their within-imputation variances will be combined across imputations using Rubin's rules. The pooled effects are then exponentiated to obtain the final hazard ratios. Pointwise 95% confidence intervals will be obtained using the pooled total variance from Rubin's rules.
Confounding adjustment method	Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other. Same confounders will be used as specified in the previous TNCC study. <u>IPC weights</u> are introduced for those who get vaccinated after baseline and thereby deviate from their assigned exposure strategy (see the ICE in Table 3). IPCW thereby corrects for the informative censoring (those receiving a new vaccination are assumed to be systematically different and would not have had the same future event experience as those not receiving a new vaccination). IPCW will be applied using logistic regression, considering the following variables as covariates: age at baseline, sex, ethnicity, smoking status, alcohol consumption, body mass index (BMI), health-cautiousness behaviour, marital status, educational level, comorbidities and the usage of repurposed drug and drugs related to high-risk comorbidities. <u>IPT weights</u> will be applied to create a pseudo-population in which treatment assignment is independent of measured baseline confounders. IPTW will be applied using logistic regression, considering the following variables as covariates: age at baseline, sex, ethnicity, smoking status, alcohol consumption, body mass index (BMI), health-cautiousness behaviour, marital status, educational level, comorbidities and the usage of repurposed drug and drugs related to high-risk comorbidities. <u>Sandwich Variance Estimator</u> will be applied to account for the correlation between observations from participants contributing to multiple emulated trials (i.e., the same participant ID being eligible for both trial emulation 1, 2 and n, thereby occurring in multiple rows of the analysis dataset). The sandwich estimator will replace the naïve covariance matrix with a cluster-robust covariance matrix, clustered on participant IDs.
Missing data methods	Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other. If crucial data is missing, such as the anonymized identifier, the data for this participant will be removed. For missing data of the other variables, first the missingness pattern will be assessed using expert knowledge. The pattern can be missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). MAR data will be imputed using <u>multivariate imputation by chained equations (MICE)</u> in R. MCAR data is deemed unrealistic and therefore distinguishing between the two is not necessary, thus both MCAR and MAR data will be imputed using MICE. Exposures and all complete covariates will be used as predictors, and polytomous regression will be applied to estimate missing values for categorical variables. Due to the type of respondents to the Lifelines questionnaires, it can be expected that some data is MNAR, due to participants with severe COVID-19 disease (unobserved) being less likely to respond, and vaccines have been observed in literature to have a protective effect. We therefore expect that risk of infection differs between the exposure groups (i.e., we would expect more missingness for the more severe cases in the unvaccinated group). To tackle this case of MNAR, we will perform sensitivity analysis to assess robustness, see Table 12 Sensitivity analysis.
Subgroup Analyses	List all subgroups n/a

B. Secondary analysis (odds of infection)

Exposure contrast:	Same as primary analysis
Outcome:	The outcome of interest is SARS-CoV-2 infection, defined by a positive RT-PCR result taken by the Dutch GGD.
Analytic software:	Same as primary analysis
Model(s): (provide details or code)	The individual patient data for each daily trial emulation will be pooled into a single dataset in long format. Each row representing a patient at time t (a categorical variable) since time 0 for trial k, a categorical column representing the vaccination status, a binary column for the outcome (infected=0;1) and the covariates. As we would like to compare the results of the retrospective cohort study to the results of the test-negative case control study, which used a conditional logistic regression model, our primary analysis for each trial emulation will make use of a pooled logistic regression model on this single dataset using the same discrete time intervals (0-14, 14-28 days since vaccination, 29-82 days, 83-181 days, 181-364 days and 365+ days), since these yield similarly interpretable odds ratios. For the outcome model we consider the following variables as covariates: age at baseline, sex, ethnicity, smoking status, alcohol consumption, body mass index (BMI), health-cautiousness behaviour, marital status, educational level, comorbidities and the usage of drugs related to high-risk comorbidities.
Confounding adjustment method	<i>Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.</i>
	Same as primary analysis
Missing data methods	<i>Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.</i>
	Same approach as primary analysis
Subgroup Analyses	<i>List all subgroups</i>
	See Table 11C.

C. Secondary analysis (variant-specific infection)

Exposure contrast:	Same as primary analysis
Outcome:	Same as primary analysis
Analytic software:	Same as primary analysis
Model(s): (provide details or code)	Same as primary analysis, however the sequential trials will be grouped on their time0 according to the circulating dominant SARS-CoV-2 strains, and only followed till the end of the circulating strain: ○ Delta (July 5, 2021 - December 19, 2021) ○ Omicron BA.1/BA.2 (January 3, 2022 - June 6, 2022) ○ Omicron BA.4/BA.5 (June 20, 2022 - October 19, 2022) This analysis will be performed by restricting the follow-up window of each subgroup to the time window of the particular strain (see Table 1C). For example, a trial emulation starting on September 1, 2021 will have its follow-up only until the end of the Delta period (December 19, 2021).
Confounding adjustment method	<i>Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.</i> Same approach as primary analysis
Missing data methods	<i>Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.</i> Same approach as primary analysis
Subgroup Analyses	List all subgroups See Model(s) above.

D. Secondary analysis (Odds of infection for the booster vaccinated with fully vaccinated as control)

Exposure contrast:	Those who get booster vaccinated at t0 as compared to those who remain fully vaccinated during the enrolment window.
Outcome:	Same as primary analysis
Analytic software:	Same as primary analysis
Model(s): (provide details or code)	Same approach as primary analysis
Confounding adjustment method	<i>Name method and provide relevant details, e.g. bivariate, multivariable, propensity score matching (specify matching algorithm ratio and caliper), propensity score weighting (specify weight formula, trimming, truncation), propensity score stratification (specify strata definition), other.</i> Same approach as primary analysis, but time from completing to primary series to t0 is also considered as covariate.
Missing data methods	<i>Name method and provide relevant details, e.g. missing indicators, complete case, last value carried forward, multiple imputation (specify model/variables), other.</i> Same approach as primary analysis
Subgroup Analyses	List all subgroups n/a

E. Meta research question and objective

- For the meta research question (see Table 2.B), estimates between the previous TNCC study and this sequential TTE study will be compared using three statistical measures, as described by Franklin et al. (2019)¹⁷.
 - **Statistical significance agreement:** point estimates and confidence intervals (CIs) of both the TNCC and sequential TTE studies are on the same side of the null.
 - **Estimate agreement:** The sequential TTE study point estimate falls within the CI of the TNCC study point estimate, and vice versa.
 - **Standardized difference agreement:** standardized difference between the two point estimates is less than 1.96 standard errors.

¹⁷ Franklin, J.M., et al., *Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project*. Clinical Pharmacology & Therapeutics, 2020. **107**(4): p. 817-826

Table 12. Sensitivity analyses – rationale, strengths and limitations

	What is being varied? How?	Why? (What do you expect to learn?)	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Survivorship bias 1	Participants who have not responded to COVQ 15b are included, and the first recorded infection in the questionnaires is used as outcome instead of lab-confirmed RT-PCR.	Avoid selection bias (survivorship bias).	Avoid selection bias (survivorship bias).	More uncertainty about trustworthiness of reported infection.
Survivorship bias 2	The exclusion criteria of not having had an infection prior to COVQ 15b is dropped, and the first recorded infection in the questionnaires is used as outcome instead of lab-confirmed RT-PCR.	Avoid selection bias (survivorship bias).	Avoid selection bias (survivorship bias).	More uncertainty about trustworthiness of reported infection.
Survivorship bias 3	Stratification of early-pandemic trial emulations vs. late-pandemic trial emulations.	Survivorship bias, especially for the unvaccinated, may play a major role for trial emulations starting near the end of the pandemic (still unvaccinated and never had a prior infection).	Gives an idea about the magnitude of selection bias.	Different in effects observed may also attributable to other factors (changing testing behaviour etc.).
MNAR patterns	To assess MNAR, we will perform tipping point analysis to identify when the results from the primary analysis are overturned. If similar risks are observed, MNAR does not seem to play a significant role.	Identify MNAR.	Assesses the effect of MNAR in estimating the risk of infection.	n.a.
Different frequency of sequential emulations (biweekly instead of daily)	Biweekly trial emulations will be started as opposed to daily (Figure 2). These biweekly emulations contain a 14-day pre-index date enrolment window to include all eligible vaccinated participants within trial starts. See study diagram below this table.	If the primary daily trial emulations suffer from insufficient power or positivity violations due to small sample sizes, a biweekly emulation will be conducted as a sensitivity analysis. For convenience, the first enrolment window will be the only exception being a 13-day enrolment window, which starts from the date the whole adult population in the Netherlands became eligible for vaccination (June 18, 2021; when those born in 2003 became eligible to get vaccinated). The influence of this 1-day difference for this first trial emulation (starting on July 1 st) as compared to the usual 14-day enrolment windows of the other trial emulations on estimation of the pooled outcomes is deemed negligible. In these enrolment windows, people who get vaccinated become eligible for the vaccinated arms of the trial emulation. Those that remain unvaccinated at the end of the enrolment window are assigned to the unvaccinated arm. The enrolment window of the trial emulation ends on each 1 st and middle day of the month starting from July 1, 2021. Enrolment windows are introduced to avoid the introduction of immortal time bias, but do alter the interpretation of the estimand (see Limitations 8.6). Daily trial emulations are not considered, since the sample sizes in daily trials are expected to lead to insufficient power.	The risk of violation of positivity is avoided.	The addition of the 14-day enrolment window changes the estimand of being vaccinated at time 0 (the index date) into being vaccinated within the 14 days prior to time 0

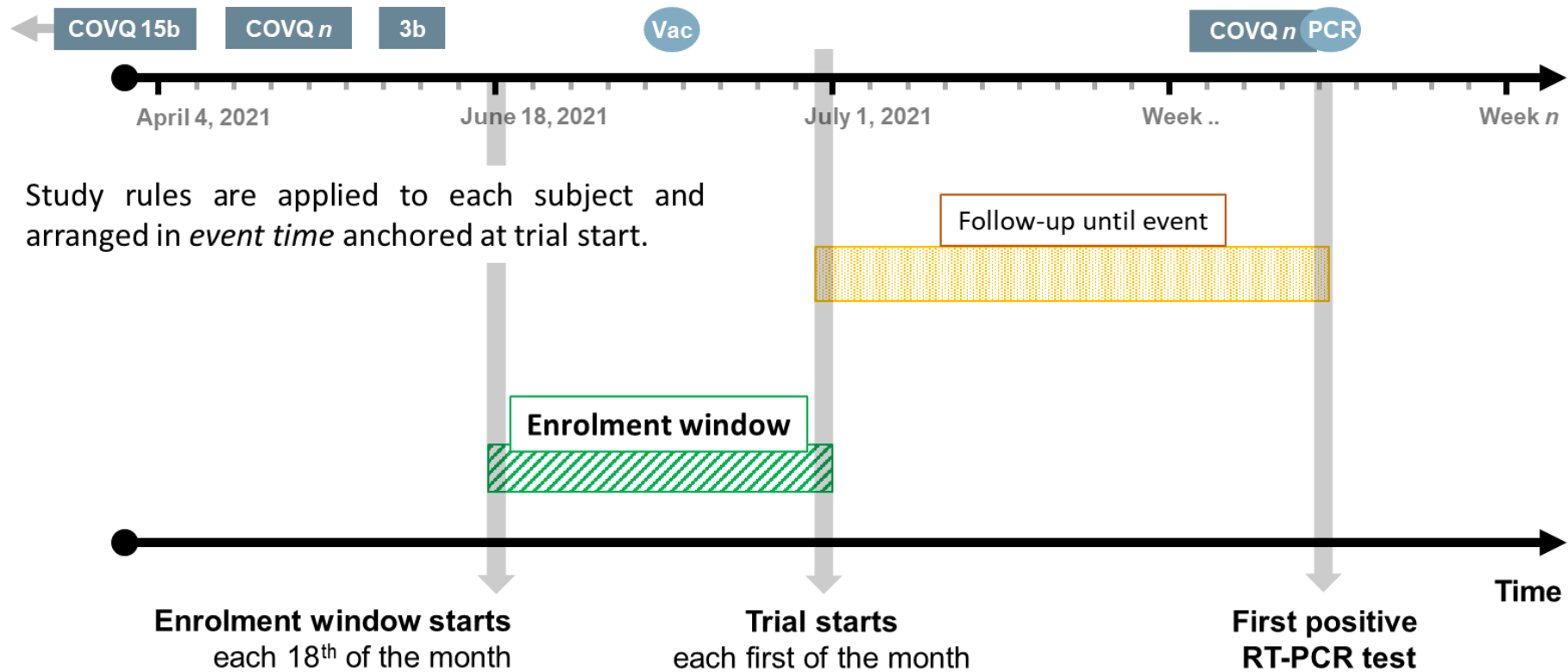


Figure 2. Study diagram of the sensitivity analysis with less frequent sequential trial starts (biweekly instead of daily). Individual-patient data is retrieved from Lifelines and CBS and documented as encounters from various sources: Lab-confirmed SARS-CoV-2 infection (PCR), vaccination with COVID-19 vaccine (Vac), or filled in Lifelines assessment form 1a-3b (1a-3b) or Lifelines COVQ assessment form 15b-29 (COVQ n). Timelines were arranged in days.

7.8. Data sources

The Lifelines cohort (1A, 2A, 2B, 3A, 3C and COVQ15b-29) will be linked to the IADB.nl pharmacy dispensing database, CBS microdata (mortality) and CBS cims (COVID-19 vaccination).

7.6.1 Context and rationale for data sources

Reason for selection: The Lifelines cohort contains data from participants of the Northern part of the Netherlands, including medical history, psychosocial characteristics, lifestyle and more. The Lifelines COVID-19 questionnaires (COVQ) contain data specifically related to COVID-19. The IADB.nl pharmacy dispensing database contains data on pharmacy dispensing for the Northern part of the Netherlands. CBS microdata contains mortality data of the Dutch population and CBS cims contains COVID-19 vaccination data of the Dutch population.

Strengths of data source(s): Through linkage we combine the strengths of the detailed cohort and questionnaires with administrative data related to vaccinations administered in the Netherlands as well as mortality data.

Limitations of data source(s): n.a.

Table 13. Metadata about data sources and software

	Data 1	Data 2	Data 3	Data 4
Data Source(s):	Lifelines	IADB.nl	CBS microdata	CBS cims
Study Period:	30-03-2020 to 19-10-2022	Linked to Lifelines	Linked to Lifelines	Linked to Lifelines
Type(s) of data:	Questionnaire responses	Pharmacy dispensing data	Mortality data	COVID-19 vaccination data
Eligible cohort entry period	Nov. 2, 2020 – Oct. 19, 2022	Nov. 2, 2020 – Oct. 19, 2022	Nov. 2, 2020 – Oct. 19, 2022	Nov. 2, 2020 – Oct. 19, 2022
Data version	Data retrieved on May 1, 2024.	Data retrieved on May 1, 2024.	Data retrieved on May 1, 2024.	Data retrieved on May 1, 2024.
Data type	Cohort questionnaire data	Pharmacy dispensings data	Demographic data	National COVID-19 vaccination data
Data sampling	All participants	Those that are present in the Lifelines cohort.	Those that are present in the Lifelines cohort.	Those that are present in the Lifelines cohort.
Data linkage:	yes	yes	yes	yes
Conversion to CDM*:	no	no	no	no
Software for data management:	Linked (see CBS microdata)	Linked (see CBS microdata)	SQL, R	Linked (see CBS microdata)

*CDM = Common Data Model

8. Limitation of the methods

- 8.1. Infections falling outside the time intervals acting as proxy for dominant variant were not considered. While ignoring this informative data, it probably will not affect outcomes as both negative as well as positive tests are equally excluded. Variant categorization relies on calendar-period proxies and may introduce misclassification due to transitional variant periods. This may bias variant-specific VE estimates if vaccination status is associated with calendar time
- 8.2. Due to limitations of the dataset (e.g., height was only requested in Lifelines assessments 2a and 3a, while weight only in the COVQ assessment forms), baseline characteristics are assumed to remain constant over time (over the consecutive Lifelines questionnaires).
- 8.3. Violation of positivity was avoided by setting the date of the first trial emulation to June 18, 2021, as not all participant were yet able to get the vaccination in the Netherlands prior to that particular date, i.e., the eligibility to get vaccinated depended on the birth-year of the participant, where the first vaccines became available for those born before 1931 (January 26, 2021) and the last vaccines became available for those born in 2003 (June 18, 2021). This means that our eligible population may be prone to selection bias, as those that did not yet get vaccinated by June 18 may have had specific reasons not generalizable to the whole population.
- 8.4. Selection bias (survivorship bias) may have been introduced by excluding participants without a response to questionnaire COVQ 15b, as they may have reported an infection in earlier COVQ questionnaires not considered due to no clarity if these infections have been identified through RT-PCR taken by the GGD. See Table 12 sensitivity analysis (survivorship bias 1).
- 8.5. Selection bias (survivorship bias) may have been introduced by excluding participants with an infection prior to questionnaire COVQ 15b. This survivorship bias is not mitigable by using the sequential TTE approach, as 15b is considered part of our outcome of interest and pre-15b periods/risk sets are ignored. See Table 12 sensitivity analysis (survivorship bias 2)
- 8.6. In our sensitivity analysis of a different sequential frequency, biweekly trials are introduced with a two-week enrolment window. This changes the estimand from 'effect of getting vaccinated on t0' to 'effect of becoming vaccinated in the first 14 days and surviving event-free until day 14.'

9. Protection of human subjects

The Lifelines cohort was approved by the Medical Ethics Committee of the University Medical Center Groningen (ID: METc 2007/152), and all participants signed informed consent forms confirming their permission for their (anonymized) data and material to be used for scientific purposes. IADB.nl data are collected in accordance with the national and European guidelines on privacy requirements for handling human data. The Dutch citizen service number (BSN) is anonymized.

Appendix A

List of repurposed drugs and drugs related to high-risk comorbidities

Medication	ATC code	Medication	ATC code
Chronic drugs			
<i>Drugs for cardiovascular system</i>			
Angiotensin-converting-enzyme inhibitors (ACEi)	C09AA	Angiotensin receptor blockers (ARBs)	C09CA
Beta-blocking agents	C07A	Calcium channel blockers (CCB)	C08
Diuretics	C03	Statin	C10AA
<i>Antivirals</i>			
Lopinavir/Ritonavir (LPV/r)	J05AR10	tenofovir disoproxil/emtricitabine (TDF/FTC)	J05AR03
Ribavirin	J05AP01		
<i>Antidiabetic drugs</i>			
Metformin	A10BA02	Sulfonylureas	A10BB
Thiazolidinediones	A10BG	DPP-4 inhibitors	A10BH
GLP-1 analogues	A10BJ	SGLT2 inhibitors	A10BK
Insulin	A10A		
<i>Antithrombotic agents</i>			
Aspirin as antithrombotic agent	B01AC06	Warfarin	B01AA03
Heparin group	B01AB		
<i>Antipsychotics and antidepressants</i>			
Non-selective monoamine reuptake inhibitors	N06AA	Selective serotonin reuptake inhibitors (SSRIs)	N06AB
Monoamine oxidase inhibitors (MAOIs), non-selective	N06AF	Monoamine oxidase inhibitors (MAOIs)	N06AG
Other antidepressants	N06AX	Antipsychotics	N05A
Fluoxetine	N06AB03	Fluvoxamine	N06AB08
<i>Others</i>			
Ciclesonide (inhalation)	R03BA08	Ciclesonide (nasal)	R01AD13
Colchicine	M04AC01	Non-steroidal anti-inflammatory drugs (NSAIDs)	M01A
Famotidine	A02BA03	Proton pump inhibitors (PPIs)	A02BC
Camostat	B02AB04	Azelastine (nasal)	R01AC03
Acute drugs			
<i>Antimalarials and antiparasitic</i>			
Hydroxychloroquine	P01BA02	Chloroquine	P01BA01
Ivermectin	P02CF01		
<i>Antivirals and antibacterial</i>			
Azithromycin	J01FA10	Umifenovir	J05AX13
Doxycycline	J01AA02	Favipiravir	J05AX27
<i>Others</i>			
Melatonin	N05CH01		