## Online methods

We used data from the ComPaRe long COVID cohort to emulate a target trial evaluating the effect of a first COVID-19 vaccine injection among patients with long COVID on the severity and impact of their symptoms 1.

### Data sources

The ComPaRe long COVID cohort is an ongoing nationwide e-cohort of patients with long COVID, in France, nested in the ComPaRe research program ([www.compare.aphp.fr](http://www.compare.aphp.fr)), an umbrella e-cohort of patients with chronic conditions 2. Participation opened in November 2020 and is ongoing. The cohort includes adult patients who have reported a COVID-19 infection (whether or not laboratory-confirmed by a positive SARS-CoV2 test by PCR swab and/or serological assay) and with symptoms persisting more than three weeks past the initial infection. Recruitment took place through 1) a social and general media campaign, 2) calls for participation from partner patient associations and on the official French contact tracing app “TousAntiCOVID”, and 3) a "snowball" sampling method where participants were encouraged to invite people who had a COVID-19 infection and persisting symptoms to enrol.3 All patients provided online consent before participating in the cohort. The Institutional Review Board of Hôtel-Dieu Hospital, Paris, approved the study (IRB: 0008367).

Participants in the ComPaRe long COVID cohort are contacted for follow-up every 60 days by e-mails with links to an online questionnaire. At each observation point (T0: cohort enrollment, T1, T2 etc.), patients are first asked if they still have symptoms related to COVID-19. Those reporting persisting symptoms complete the long COVID symptom tool (ST) and impact tool (IT), a pair of validated patient-reported instruments assessing respectively 53 long COVID symptoms and 6 dimensions of patients’ lives that can be affected by the disease.4 Those reporting that they no longer had any symptoms are asked to specify the date when they first noticed the absence of symptoms.

In addition, since May 11, 2021, and every 45 days since, in a different online questionnaire, patients have been self-reporting their COVID-19 vaccination status, and, if vaccinated, the vaccine received, the date(s) of vaccination, and possible adverse effects experienced. In February 2022, all patients who had not previously reported vaccination were contacted by e-mail to confirm their vaccination status.

### Eligibility criteria

Our analyses used data from patients enrolled in the ComPaRe long COVID cohort before May 1, 2021. We considered eligible for the emulated trial adult patients (≥ 18 years old) with a confirmed or suspected COVID-19 infection; with symptoms persisting more than three months past the initial infection and who still reported at least one symptom attributable to long COVID. We excluded patients reporting a history of severe allergy in ComPaRe, because they are also likely to have a history of anaphylaxis, considered a contraindication to vaccination at the time of the study.

### Interventions

We compared the efficacy of receiving a first COVID-19 vaccine injection with the ChAdOx1 (Astra Zeneca), BNT162b2 mRNA (Pfizer-BioNTech), Ad26.COV2. S (Johnson & Johnson) or mRNA-1273 (Moderna) vaccines to no such vaccination. Of note no other COVID-19 vaccine was approved in France at the time of the study.

### Outcomes

The primary outcome was the long COVID ST score. This score is the number of symptoms reported among 53 included in the questionnaire and can therefore range from 0 (i.e., disease remission) to 53.4 Second, we investigated the disease remission (i.e., complete disappearance of symptoms). Third, we evaluated the long covid IT score, which is the sum of the responses to the six items evaluating the impact of the disease on patients’ lives and which can range from 0 (no impact) to 60 (maximal impact). Finally, we analysed the long COVID IT score after dichotomisation according to its Patient Acceptable Symptom State (PASS), which represents the score below which 75% of patients considered their symptom state acceptable. In a previous study, we estimated the PASS for long COVID IT to be 30/60.4 All outcomes were assessed at 120 days after inclusion in the emulated trial (thereafter termed baseline).

Serious adverse events after vaccination were analysed by a single investigator (VTT) from participants’ open-text answers to the related questions in the online questionnaire.

### Treatment groups and follow-up

To emulate a trial evaluating the effectiveness of vaccination in a population where most patients were eventually vaccinated against COVID-19, we used a method previously described to define a sequence of three trials that we subsequently pooled.5

In the first trial, we identified all patients who met the eligibility criteria when they were enrolled in the ComPaRe long COVID cohort (i.e. their first observation point, T0). Patients who received their first COVID-19 vaccine injection between baseline and 60 days (i.e., their second observation point, T1), were classified in the vaccination group and propensity score-matched at a 1:1 ratio to patients who did not receive the vaccine in the same period (T0 to T1), classified in the control group. Patients were followed up for 120 days (i.e., to their third observation point, T2, and endpoint of the first trial). Unvaccinated controls who were vaccinated before T2 were censored at the date of vaccination.

We repeated this procedure by emulating two additional trials, by considering baseline at 60 days (i.e., T1) for the second trial, and T2 for the third; we applied a similar follow-up strategy (i.e., follow-up until T3 and T4, respectively). At the baseline of each of the three trials, patients' who no longer met the eligibility criteria, for example because they no longer reported symptoms, were not eligible to that trial. Control patients who had since received COVID-19 vaccination were eligible for inclusion in the vaccination group even though they had previously served as a control in the first (or second) trial, but a patient could only be selected once as a control and once as a vaccinated patient. The sequence of trials is defined in more detail in **Supplementary Material 1**.

### Statistical analysis

Within each of the three trials, each vaccinated patient was matched to an unvaccinated control according to their probability of getting vaccinated against COVID-19 given their baseline covariates (i.e., the propensity score). The propensity score was calculated with a non-parsimonious multivariable logistic regression model including variables planned and prespecified before outcome analyses: 1) sex; 2) age; 3) educational level (two years or more post-secondary education, ie, high education versus lower), 4) number of comorbidities (self-reported by using the International Classification of Primary Care-Version 2)6; 5) laboratory-confirmed SARS-CoV-2 infection (yes, for patients reporting a positive result for SARS-CoV2 by PCR swab and/or serological assay, and otherwise no); 6) time since COVID-19 symptom onset, 7) history of hospitalisation for COVID-19 during its acute phase; 8) long COVID ST score at baseline; and 9) long COVID IT score at baseline. Standardised differences were examined to assess balance, with values below 10% considered as indicating successful balance.7 Propensity score matching used a calliper width of 0.2 of the pooled standard deviation of the logit of the propensity score.8

For analysis, we pooled the vaccination and the control groups from the three trials and estimated the effect of treatment by using paired t-tests for continuous outcomes,9 marginal Poisson models for dichotomous outcomes, and marginal Cox proportional hazard models for time-to-event outcomes. The proportional hazards assumption was tested using the Grambsch and Therneau test 10. We took into account the artificial censoring of patients in the unvaccinated group by using multiple imputation by chained equations for continuous outcomes and using the inverse probability of censoring weighting technique for time-to-event outcomes.11 Further, in the survival analyses, to account for immortal time bias, baseline was considered as the vaccination date for patients in the vaccination group and the vaccination date of their matched patient for those in the control group.

All outcomes were studied in the total population and in a subgroup restricted to participants with a confirmed SARS-CoV-2 infection. A post-hoc subgroup analysis of patients whose time since symptom onset was ≤ and > 12 months evaluated whether the effect of vaccination on disease severity varied by time since initial infection.

To examine the potential correlation induced by including the same patient in several trials, we performed a sensitivity analysis where the study population was limited to patients who had been included in only one of the three trials (i.e., excluding patients included twice in the study, once as an unvaccinated patient and then as a vaccinated patient). Finally, the design of the study included a grace period of 60 days after baseline during which patients receive vaccination. To account for the variation in the length of follow-up of patients, we explored how time since vaccination affected the study outcomes. To that end, we described patients’ long COVID ST and IT scores as a function of the delay between vaccination and outcome measurement (120 days), in the vaccination group.

Missing baseline and outcome variables were handled by multiple imputation by chained equations that used the other variables available. All statistical analyses were performed with the R statistical package version 4.0.3 (The R Foundation for Statistical Computing, <https://www.R-project.org/>).

1 Hernan, M. A. & Robins, J. M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American journal of epidemiology* **183**, 758-764, doi:10.1093/aje/kwv254 (2016).

2 Tran, V. T. & Ravaud, P. COllaborative open platform E-cohorts for research acceleration in trials and epidemiology. *Journal of clinical epidemiology* **124**, 139-148, doi:10.1016/j.jclinepi.2020.04.021 (2020).

3 Denscombe, M. (Open University Press, Buckingham, 1997).

4 Tran, V. T. *et al.* Development and validation of the long covid symptom and impact tools, a set of patient-reported instruments constructed from patients' lived experience. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, doi:10.1093/cid/ciab352 (2021).

5 García-Albéniz, X., Hsu, J., Bretthauer, M. & Hernán, M. A. Effectiveness of Screening Colonoscopy to Prevent Colorectal Cancer Among Medicare Beneficiaries Aged 70 to 79 Years: A Prospective Observational Study. *Annals of internal medicine* **166**, 18-26, doi:10.7326/m16-0758 (2017).

6 World Health Organization. *International Classification of Primary Care, Second edition (ICPC-2)*, <<https://www.who.int/classifications/icd/adaptations/icpc2/en/>> (2003).

7 Austin, P. C. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* **28**, 3083-3107, doi:10.1002/sim.3697 (2009).

8 Austin, P. C. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* **27**, 2037-2049, doi:10.1002/sim.3150 (2008).

9 Austin, P. C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate behavioral research* **46**, 399-424, doi:10.1080/00273171.2011.568786 (2011).

10 Grambsch, P. & Therneau, T. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* **81**, 515-526 (1994).

11 Rotnitzky, A. & Robins, J. Analysis of semi-parametric regression models with non-ignorable non-response. *Stat Med* **16**, 81-102, doi:10.1002/(sici)1097-0258(19970115)16:1<81::aid-sim473>3.0.co;2-0 (1997).