The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia

Martí Català, Núria Mercadé-Besora, Raivo Kolde, Nhung T H Trinh, Elena Roel, Edward Burn, Trishna Rathod-Mistry, Kristin Kostka, Wai Yi Man, Antonella Delmestri, Hedvig M E Nordeng, Anneli Uusküla, Talita Duarte-Salles, Daniel Prieto-Alhambra, Annika M Jödicke**

Summary

Background Although vaccines have proved effective to prevent severe COVID-19, their effect on preventing long-term symptoms is not yet fully understood. We aimed to evaluate the overall effect of vaccination to prevent long COVID symptoms and assess comparative effectiveness of the most used vaccines (ChAdOx1 and BNT162b2).

Methods We conducted a staggered cohort study using primary care records from the UK (Clinical Practice Research Datalink [CPRD] GOLD and AURUM), Catalonia, Spain (Information System for Research in Primary Care [SIDIAP]), and national health insurance claims from Estonia (CORIVA database). All adults who were registered for at least 180 days as of Jan 4, 2021 (the UK), Feb 20, 2021 (Spain), and Jan 28, 2021 (Estonia) comprised the source population. Vaccination status was used as a time-varying exposure, staggered by vaccine rollout period. Vaccinated people were further classified by vaccine brand according to their first dose received. The primary outcome definition of long COVID was defined as having at least one of 25 WHO-listed symptoms between 90 and 365 days after the date of a PCR-positive test or clinical diagnosis of COVID-19, with no history of that symptom 180 days before SARS-Cov-2 infection. Propensity score overlap weighting was applied separately for each cohort to minimise confounding. Sub-distribution hazard ratios (sHRs) were calculated to estimate vaccine effectiveness against long COVID, and empirically calibrated using negative control outcomes. Random effects meta-analyses across staggered cohorts were conducted to pool overall effect estimates.

Findings A total of 1 618 395 (CPRD GOLD), 5 729 800 (CPRD AURUM), 2 744 821 (SIDIAP), and 77 603 (CORIVA) vaccinated people and 1 640 371 (CPRD GOLD), 5 860 564 (CPRD AURUM), 2 588 518 (SIDIAP), and 302 267 (CORIVA) unvaccinated people were included. Compared with unvaccinated people, overall HRs for long COVID symptoms in people vaccinated with a first dose of any COVID-19 vaccine were 0·54 (95% CI 0·44–0·67) in CPRD GOLD, 0·48 (0·34–0·68) in CPRD AURUM, 0·71 (0·55–0·91) in SIDIAP, and 0·59 (0·40–0·87) in CORIVA. A slightly stronger preventative effect was seen for the first dose of BNT162b2 than for ChAdOx1 (sHR 0·85 [0·60–1·20] in CPRD GOLD and 0·84 [0·74–0·94] in CPRD AURUM).

Interpretation Vaccination against COVID-19 consistently reduced the risk of long COVID symptoms, which highlights the importance of vaccination to prevent persistent COVID-19 symptoms, particularly in adults.

Funding National Institute for Health and Care Research.

Lancet Respir Med 2024; 12: 225–36

Published Online January 11, 2024 https://doi.org/10.1016/S2213-2600(23)00414-9

*Joint senior authors

Pharmaco- and Device Epidemiology Group, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (M Català PhD, E Burn PhD, T Rathod-Mistry PhD, K Kostka MSc, W Yi Man MSc, A Delmestri PhD, Prof D Prieto-Alhambra PhD, A M Jödicke PhD) **and Oxford National Institute for Health and Care Research Biomedical Research Centre** (Prof D Prieto-Alhambra), **University of Oxford, Oxford, UK; Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Barcelona, Spain** (N Mercadé-Besora BA, E Roel PhD, T Duarte-Salles PhD); **Institute of Computer Science** (R Kolde PhD) **and Department of Family Medicine and Public Health** (A Uusküla PhD), **University of Tartu, Tartu, Estonia; Pharmacoepidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway** (N T H Trinh PhD, Prof H M E Nordeng PhD); **Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway** (Prof H M E Nordeng); **Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands** (T Duarte-Salles, Prof D Prieto-Alhambra)

Introduction

In March 2020, WHO declared the outbreak of COVID-19, a disease caused by the novel SARS-CoV-2 virus, to be a global pandemic.[1] 3 years later, in March, 2023, more than 760 million COVID-19 cases were confirmed worldwide.[2] Vaccines against SARS-CoV-2 were rapidly developed to tackle the pandemic, with the first approved COVID-19 vaccine dose being administered in the UK in December, 2020. Since then, eight COVID-19 vaccines have received marketing authorisation from the European Medicines Agency,[3], with BNT162b2 (Pfizer–BioNTech), mRNA-1273 (Moderna), ChAdOx1 (Oxford–AstraZeneca), and Ad26.COV2.S (Janssen) COVID-19 vaccines being the most frequently used in Europe.[4] These vaccines proved to be highly effective in preventing severe COVID-19,[5–7] mortality,[8] and community transmission.[9] However, persistent symptoms and complications after recovering from acute COVID-19 disease have been increasingly reported, shifting the focus of clinical research from the acute phase of the infection to its long-term complications. Studies suggest that approximately one in ten people infected with SARS-CoV-2 have persisting symptoms,[10,11] with increased risk for developing long COVID associated with age,[10,12] female sex,[10,12,13] and comorbidities.[12,13]

In October, 2021, WHO characterised the post COVID-19 condition as probable or confirmed SARS-CoV-2 infection, with new or persisting symptoms 3 months after infection that cannot be explained by

Correspondence to:
Prof Daniel Prieto-Alhambra, Pharmaco- and Device Epidemiology Group, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK
daniel.prietoalhambra@ndorms.ox.ac.uk

Research in context

Evidence before this study

Although previous studies have provided insights on the risk of long COVID following breakthrough infections, most of them only included people with COVID-19. As recognised by a systematic review of the literature led by the UK Health Security Agency, these studies have probably underestimated the effectiveness of vaccines to prevent long COVID. As post-acute COVID-19 complications can only occur in people who were previously infected with SARS-CoV-2, the effect of vaccines to prevent SARS-CoV-2 infections is a crucial factor to include when estimating vaccine effectiveness to prevent long COVID.

Added value of this study

To our knowledge, this is the first multinational study to assess population-level vaccine effectiveness to prevent long COVID symptoms. Our study of more than 10 million vaccinated people and 10 million unvaccinated people, showed that COVID-19 vaccination reduces the risk of developing long COVID. Our findings were consistent across three different European countries and four databases, covering different health-care settings and national health-care policies. All vaccines reduced the risk of developing long COVID symptoms, with BNT162b2 showing slightly better effectiveness than ChAdOx1.

Implications of all the available evidence

Vaccination against COVID-19 consistently reduced the risk of long COVID symptoms, highlighting yet another benefit of vaccination, particularly in adults who are less at risk of severe outcomes.

alternative causes.[14] A list of common symptoms was established, comprising 25 key symptoms such as fatigue, shortness of breath, and cognitive dysfunction.[14] Unfortunately, neither long COVID or the post COVID-19 condition were included as study outcomes in any of the pivotal trials that led to the approval of the vaccines mentioned above.

Although COVID-19 vaccines remain widely used, according to a recent review by the UK Health Security Agency (UKHSA), their effect on preventing long COVID is not yet fully understood.[15] Most previous studies assessing the association of pre-infection vaccination and long-term complications showed a reduction in risk for long COVID associated with vaccination.[15] However, some only reported reductions for selected symptoms, or no overall risk reduction. Systematic reviews showed that effect estimates vary greatly between studies, largely depending on the long COVID definition used, the study population included, and the health-care setting studied.[15–18] However, most of these studies only included people with COVID-19.[19–30] Therefore, these studies have probably underestimated the effectiveness of vaccines to prevent long COVID. Because post-acute COVID-19 complications can only occur in people who were previously infected with SARS-CoV-2, the effect of vaccines to prevent SARS-CoV-2 infections is a crucial factor to include when estimating vaccine effectiveness to prevent long COVID.

The aim of this study was to evaluate the overall effect of vaccination to prevent long COVID and assess comparative effectiveness of the most used vaccines (ChAdOx1 and BNT162b2).

Methods

Data sources

We used routinely collected, de-identified health data from three European countries: primary care electronic health records from two large UK data sources (the Clinical Practice Research Datalink [CPRD] GOLD and CPRD AURUM), primary care records linked to hospital admission data from Catalonia, Spain (Information System for Research in Primary Care [SIDIAP]), and national health insurance claims from Estonia (CORIVA).

For more on **CPRD** see https://cprd.com/

For more on **SIDIAP** see https://www.sidiap.org/index.php/en/

CPRD GOLD[31] currently comprises 3·1 million active participants and CPRD AURUM[32] comprises 13·3 million active participants, with CPRD GOLD predominantly covering practices in Scotland and Wales, and CPRD AURUM mainly covering practices in England. SIDIAP[33] represents around 80% of the population living in Catalonia and was linked to hospital admission data for this study, including all hospitals in the region's universal health-care system (Conjunt Mínim Bàsic de Dades d'Alta Hospitalària). The CORIVA database covers approximately 440 000 people, a random sample from the Estonian population plus all COVID-19 cases from the first year of the pandemic (Feb 28, 2020 to Feb 28, 2021).

CPRD (GOLD and AURUM) and SIDIAP include information on patient demographics, comorbidities, medicine use, laboratory measurements, clinical measurements, lifestyle factors, and referrals to secondary care. Vaccination status was obtained from linked national or regional registry data, including vaccines administered in any health-care or public health setting, and vaccination centres. CORIVA includes information on patient demographics, diagnoses and conditions, drug prescriptions, and procedures from insurance claims, with linked COVID-19 testing and vaccination status from the national health information system and information from the causes of death registry (all causes of death were included).

All databases were mapped to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM)[34,35] to enable federated analytics.

Study design and populations

As people became eligible to receive their first vaccine dose at different times, this study was designed as a staggered cohort study following the vaccine rollout guidelines in the UK, Spain, and Estonia. For each database and country, we created four study cohorts, with each of them representing a specific stage of the national vaccination campaign rollout. Details of the country-specific enrolment periods and priority groups for vaccination, and the study design are provided in the appendix (pp 8–10).

Here, we describe the study design using CPRD and the UK vaccine rollout as an illustrative example. All adults registered in CPRD GOLD or CPRD AURUM for at least 180 days as of the start of the first staggered cohort (Jan 4, 2021) comprised the source population. From this source population, people were included to the respective study cohorts and eligibility was assessed separately for each cohort: cohort one (enrolment period from Jan 4 to Jan 27, 2021, corresponding to priority groups 2–3 in the UK vaccination plan) included all people aged 75 years and older who did not receive a COVID-19 vaccination and had no history of COVID-19 before enrolment. Among this group, people were included to the vaccinated or unvaccinated group based on whether they received their first vaccine dose during the enrolment period. Index date for vaccinated people was the date of vaccine administration. For unvaccinated people, index date was randomly assigned during the enrolment period following the distribution of the index dates for the vaccinated subpopulation. Cohort two (enrolment from Jan 28 to Feb 28, 2021, priority groups 4–6) included all people aged 65 years and older and clinically extremely vulnerable people, and those with underlying health conditions aged 18 years and older.[36,37] Cohort three (enrolment from March 1 to April 13, 2021, priority groups 7–9) included people aged 50 years and older. Finally, cohort four (enrolment from April 14 to July 31, 2021, priority group 10) included people aged 18 years and older. People who were vaccinated before the start of enrolment were excluded for each of these four staggered cohorts. In addition, all people with a positive test or clinical diagnosis of COVID-19 before their assigned index date were excluded. Unvaccinated people from a previous cohort were eligible for inclusion to the subsequent cohort if they met the respective eligibility criteria. Therefore, the study design allowed unvaccinated people who were censored when they received a first vaccine dose to contribute vaccinated person-time from the time they change exposure status.

In each cohort, people were followed-up from the index date until the earliest end of their observation—ie, date of data extraction (December, 2021 for CPRD GOLD; January, 2022 for CPRD AURUM; June, 2022 for SIDIAP; and December, 2022 for CORIVA), death from any cause, leaving the general practitioner (GP) practice or the practice stopping contribution of data to database (for CPRD), or the next vaccination (date of first vaccination for the unvaccinated group, second vaccination for the vaccinated group; appendix pp 11–12). Sensitivity analyses were conducted without censoring at the second vaccine dose for vaccinated people.

Records of vaccine doses received in the UK, Spain, and Estonia were available for all people in all databases, including the vaccination date and vaccines approved between January and July, 2021 (ChAdOx1, BNT162b2, Janssen (Ad26.COV2.S), and mRNA-1273). All four vaccine brands were included in the effectiveness analyses.

See Online for appendix

This study was approved by the CPRD's Research Data Governance Process (21_000557), the Clinical Research Ethics committee of Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (4R22/133), and the Research Ethics Committee of the University of Tartu (330/T-10).

Long COVID

Our primary outcome definition of long COVID was pre-defined as the presence of at least one symptom included in the WHO clinical case definition.[14] We identified 25 long COVID symptoms based on Systemized Nomenclature of Medicine (SNOMED) codes, and defined long COVID as having at least one record of any of the pre-defined symptoms between 90 and 365 days after the date of a PCR-positive test or clinical diagnosis of COVID-19, with no record of that symptom 180 days before SARS-Cov-2 infection.[38] Sensitivity analyses were conducted using alternative definitions of long COVID, including symptoms recorded between 28 and 365 days after COVID-19, and using the SNOMED code for a GP diagnosis of post-acute COVID-19 in CPRD AURUM and SIDIAP (recorded alone, without requiring a previous record of COVID-19 or between 90 and 365 days after a SARS-CoV-2 infection). The effectiveness of vaccination against COVID-19 was estimated separately post-hoc.

All the operational definitions for each individual symptom and for the composite outcomes were evaluated using Cohort Diagnostics (version 2.2.4), an R package that provides information on face validity, including aggregate estimates of affected patient characteristics, descriptive epidemiology, and clinical presentation. These definitions were reviewed by at least one senior clinical epidemiologist and one senior pharmacoepidemiologist (DP-A and AMJ reviewed the definitions, with input from other clinicians in the team) and improved in an iterative manner as per standard procedure to reach the final definition.

For more on **Cohort Diagnostics** see https://ohdsi.github.io/CohortDiagnostics/

Statistical analysis

A common analytical script was developed, which was subsequently adapted to mimic the country-specific vaccination rollouts (eg, dates and priority groups). The resulting code was then run locally—ie, in a federated

manner on locally available OMOP CDM-mapped data. All results were obtained separately from CPRD GOLD, CPRD AURUM, SIDIAP, and CORIVA. Comparative effectiveness analyses were only conducted in the UK data due to small sample sizes and restrictions for the use of ChAdOx1 in younger and older age groups in Estonia and Spain.

For the **code lists and algorithms** see https://github.com/oxford-pharmacoepi/LongcovidVaccineEffectiveness

Large-scale propensity scores were used to minimise confounding by indication based on recorded covariates. Propensity scores represent the probability of vaccination on the basis of sociodemographic and clinical characteristics at the time of enrolment in each of the staggered cohorts. Covariates to be included in the large-scale propensity scores were selected from comedications and comorbidities recorded before the cohort-specific index dates. Covariates with a prevalence of less than 0·5% in the study population were omitted. Logistic regression with LASSO regularisation was then used for variable selection, and the list of selected covariates was reviewed (by DP-A and AMJ) to exclude instrumental variables. In addition, to the covariates identified using LASSO, key pre-specified confounders identified based on previous literature and clinical knowledge were forced into the propensity scores equation, namely age (5-year age bands for people aged 18 years and older); location (primary care practice [for CPRD], health-care region [for SIDIAP], and nation [for CORIVA]); index date; previous observation years (the time a person was registered in the database before index date); number of previous GP visits; number of previous SARS-CoV-2 PCR tests; and regional vaccination, testing, and COVID-19 incidence rates (for CPRD). Previous work comparing methods of vaccine effectiveness research showed that overlap weighting adequately accounted for confounding.[39] Therefore, we applied overlap weighting based on the estimated propensity scores to estimate the average treatment effect in the overlap group. We assessed covariate balance using absolute standardised mean differences smaller than 0·1 indicating adequate balance based on previous literature.[40]

For the **interactive web application** see https://dpa-pde-oxford.shinyapps.io/LongcovidVaccineEffectiveness/

As previous studies have showed increased mortality after acute COVID-19, we used Fine-Gray models to estimate vaccine effectiveness while accounting for death as a competing risk. Sub-distribution hazard ratios (sHRs) were calculated separately for each staggered cohort. 43 prespecified and clinically reviewed negative control outcomes (NCOs)[39,41] were used to empirically calibrate sHR to account for residual confounding and systematic error.[42,43] Proportionality of hazards was tested by visual inspection of Kaplan-Meier plots and log(–log) plots. Cox proportional hazards regression was used as a cause-specific approach in addition to Fine-Gray models as sensitivity analyses for the primary outcome. Random effect meta-analysis was used to obtain an overall estimate (meta-analytic calibrated sHR) across all four cohorts for each database.

All analyses were conducted using R (version 4.2.2). All analytical code, code lists for vaccines, COVID-19 tests and diagnoses, long COVID symptoms, and NCOs are publicly available at GitHub.

Role of the funding source

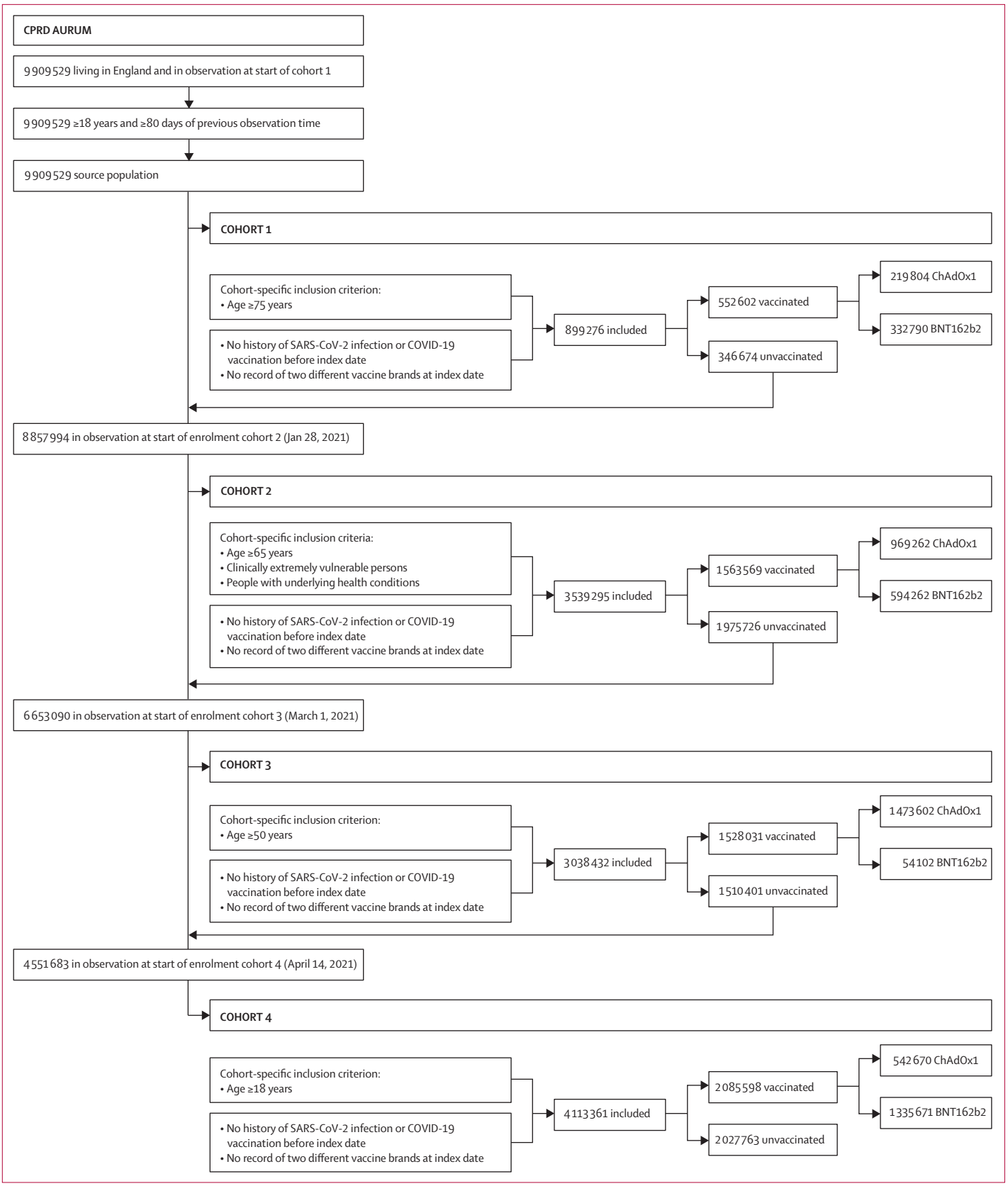
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

A total of 1 618 395 (CPRD GOLD), 5 729 800 (CPRD AURUM), 2 744 821 (SIDIAP), and 77 603 (CORIVA) vaccinated people and 1 640 371 (CPRD GOLD), 5 860 564 (CPRD AURUM), 2 588 518 (SIDIAP), and 302 267 (CORIVA) unvaccinated people were included. Among those, the proportion of people included in cohort one ranged from 6–21% across databases, 9–33% cohort two, 26–34% in cohort three, and 31–39% in cohort four (figure 1, appendix pp 13–15). Among vaccinated people, 957 025 (59%) of 1 618 395 (CPRD GOLD), 3 205 338 (56%) of 5 729 915 (CPRD AURUM), 408 486 (15%) of 2 744 821 (SIDIAP), and 4947 (6%) of 77 603 (CORIVA) were vaccinated with ChAdOx1, and 615 269 (38%) of 1 618 395 (CPRD GOLD), 2 316 825 (40%) of 5 729 915 (CPRD AURUM), 1 821 323 (66%) of 2 744 821 (SIDIAP), and 58 081 (75%) of 77 603 (CORIVA) were vaccinated with BNT162b2. The study inclusion process is shown for CPRD AURUM (figure 1) and for CPRD GOLD, SIDIAP, and CORIVA (appendix pp 13–15).

For each cohort in each database comparing vaccinated with unvaccinated people, covariate balance was calculated before and after propensity score weighting, with adequate balance achieved for all covariates after weighting, except for GP practice in CPRD AURUM and CPRD GOLD (for cohorts two and three only; results are accessible in the interactive web application). Propensity scores and overlap weight distributions for the primary outcome are included in the appendix (pp 132–133). Baseline characteristics for cohorts one to four are shown in table 1 for CPRD AURUM, and in the appendix for CPRD GOLD, SIDIAP, and CORIVA (pp 16–17, 26–29, 38–41, and 48–51), with all tables showing comparability of study participants when vaccinated and unvaccinated groups were compared after weighting. Negative control outcome analyses suggested some residual confounding after weighting; therefore, in the subsequent sections we report empirically calibrated estimates and uncalibrated estimates were reported only in the web application.

Figure 1: **Study inclusion flowchart for CPRD AURUM**
Individual vaccine numbers add to less than the total vaccinated due to other vaccines being given to a small number of patients. CPRD=Clinical Practice Research Datalink.


CPRD AURUM
9 909 529 living in England and in observation at start of cohort 1
9 909 529 ≥18 years and ≥80 days of previous observation time
9 909 529 source population
COHORT 1
Cohort-specific inclusion criterion:
• Age ≥75 years
• No history of SARS-CoV-2 infection or COVID-19 vaccination before index date
• No record of two different vaccine brands at index date
899 276 included
552 602 vaccinated
346 674 unvaccinated
219 804 ChAdOx1
332 790 BNT162b2
8 857 994 in observation at start of enrolment cohort 2 (Jan 28, 2021)
COHORT 2
Cohort-specific inclusion criteria:
• Age ≥65 years
• Clinically extremely vulnerable persons
• People with underlying health conditions
• No history of SARS-CoV-2 infection or COVID-19 vaccination before index date
• No record of two different vaccine brands at index date
3 539 295 included
1 563 569 vaccinated
1 975 726 unvaccinated
969 262 ChAdOx1
594 262 BNT162b2
6 653 090 in observation at start of enrolment cohort 3 (March 1, 2021)
COHORT 3
Cohort-specific inclusion criterion:
• Age ≥50 years
• No history of SARS-CoV-2 infection or COVID-19 vaccination before index date
• No record of two different vaccine brands at index date
3 038 432 included
1 528 031 vaccinated
1 510 401 unvaccinated
1 473 602 ChAdOx1
54 102 BNT162b2
4 551 683 in observation at start of enrolment cohort 4 (April 14, 2021)
COHORT 4
Cohort-specific inclusion criterion:
• Age ≥18 years
• No history of SARS-CoV-2 infection or COVID-19 vaccination before index date
• No record of two different vaccine brands at index date
4 113 361 included
2 085 598 vaccinated
2 027 763 unvaccinated
542 670 ChAdOx1
1 335 671 BNT162b2

	Cohort 1			Cohort 2			Cohort 3			Cohort 4		
	Unvaccinated (n=154 864)	Vaccinated (n=154 245)	SMD	Unvaccinated (n=420 707)	Vaccinated (n=420 931)	SMD	Unvaccinated (n=463 495)	Vaccinated (n=462 463)	SMD	Unvaccinated (n=818 917)	Vaccinated (n=827 124)	SMD
Age, years	80 (76–84)	80 (76–84)	0·000	58 (44–67)	58 (44–67)	0·005	50 (41–58)	52 (40–58)	0·003	34 (26–42)	34 (26–42)	0·004
Sex												
Female	88 349 (57%)	87 639 (57%)	0·005	248 156 (59%)	249 561 (59%)	0·006	245 248 (53%)	245 600 (53%)	0·004	351 435 (43%)	358 688 (43%)	0·009
Male	66 515 (43%)	66 606 (43%)	0·005	172 551 (41%)	171 370 (41%)	0·006	218 247 (47%)	216 863 (47%)	0·004	467 482 (57%)	468 436 (57%)	0·009
Previous observation years*	24 (10·4–35·4)	24·4 (10·4–35·7)	0·006	17·6 (7·9–28·5)	17·6 (7·9–28·5)	0·003	13·9 (6·46–23·7)	14·3 (6·5–23·9)	0·008	7·8 (3·8–16·9)	7·4 (3·2–17·7)	0·001
Number of GP visits	10 (5–18)	10 (6–17)	..	8 (3–15)	8 (5–14)	..	4 (1–11)	6 (3–11)	..	2 (0–6)	2 (0–6)	..
Number of PCR tests	0 (0–0)	0 (0–0)	..	0 (0–0)	0 (0–0)	..	0 (0–0)	0 (0–0)	..	0 (0–0)	0 (0–0)	..
Comorbidities†												
Anxiety	23 200 (15%)	22 789 (15%)	0·006	94 390 (22%)	91 644 (22%)	0·016	92 820 (20%)	90 807 (20%)	0·010	123 055 (15%)	125 202 (15%)	0·003
Asthma	16 978 (11%)	16 663 (11%)	0·005	95 770 (23%)	94 550 (22%)	0·007	79 642 (17%)	78 266 (17%)	0·007	63 687 (8%)	61 472 (7%)	0·013
Chronic kidney disease	36 149 (23%)	36 046 (23%)	0·001	28 181 (7%)	29 756 (7%)	0·015	10 283 (2%)	10 577 (2%)	0·005	3840 (<1%)	3572 (<1%)	0·006
Chronic obstructive pulmonary disease	13 385 (9%)	13 181 (9%)	0·003	17 447 (4%)	17 999 (4%)	0·006	6062 (1%)	5754 (1%)	0·006	1901 (<1%)	1918 (<1%)	0·000
Dementia	9483 (6%)	8517 (6%)	0·026	4182 (1%)	3879 (1%)	0·007	1361 (<1%)	1392 (<1%)	0·001	276 (<1%)	495 (<1%)	0·012
Depressive disorder	18 632 (12%)	18 547 (12%)	0·000	85 280 (20%)	81 945 (19%)	0·020	81 891 (18%)	79 804 (17%)	0·011	94 373 (12%)	97 053 (12%)	0·007
Diabetes (type 1 and type 2)	29 365 (19%)	28 831 (19%)	0·007	49 408 (12%)	48 562 (12%)	0·006	26 616 (6%)	28 628 (6%)	0·019	12 787 (2%)	12 539 (2%)	0·004
Gastro-oesophagial reflux disease	8718 (6%)	8515 (6%)	0·005	19 907 (5%)	18 924 (4%)	0·011	15 646 (3%)	14 982 (3%)	0·008	13 882 (2%)	13 893 (2%)	0·001
Heart failure	9349 (6%)	8851 (6%)	0·013	7284 (2%)	6502 (2%)	0·015	2660 (1%)	2470 (1%)	0·005	930 (<1%)	816 (<1%)	0·005
Hypertension	81 563 (53%)	80 806 (52%)	0·006	97 707 (23%)	98 193 (23%)	0·002	54 649 (12%)	55 798 (12%)	0·008	22 925 (3%)	24 450 (3%)	0·009
Hypothyroidism	15 125 (10%)	15 098 (10%)	0·001	25 579 (6%)	25 962 (6%)	0·004	17 162 (4%)	17 580 (4%)	0·005	12 427 (2%)	12 641 (2%)	0·001
Malignant neoplastic disease	33 467 (22%)	33 024 (21%)	0·005	30 194 (7%)	35 085 (8%)	0·043	14 815 (3%)	14 140 (3%)	0·008	6447 (1%)	5766 (1%)	0·011
Myocardial infarction	7824 (5%)	7731 (5%)	0·002	9964 (2%)	11 319 (3%)	0·020	3787 (1%)	3664 (1%)	0·003	1315 (<1%)	1069 (<1%)	0·008
Osteoporosis	15 275 (10%)	15 373 (10%)	0·003	10 626 (3%)	10 718 (3%)	0·001	4113 (1%)	4131 (1%)	0·001	1376 (<1%)	1472 (<1%)	0·002
Pneumonia	8573 (6%)	7621 (5%)	0·027	11 355 (3%)	10 691 (3%)	0·010	6651 (1%)	6545 (1%)	0·002	5144 (1%)	5151 (1%)	0·001
Rheumatoid arthritis	3066 (2%)	3092 (2%)	0·002	6198 (1%)	6570 (2%)	0·007	2355 (1%)	3111 (1%)	0·021	1201 (<1%)	859 (<1%)	0·012
Stroke	7667 (5%)	7047 (5%)	0·018	8041 (2%)	8794 (2%)	0·013	3518 (1%)	3293 (1%)	0·006	1496 (<1%)	1305 (<1%)	0·006
Venous thromboembolism	9589 (6%)	9241 (6%)	0·008	11 836 (3%)	12 475 (3%)	0·009	6503 (1%)	8075 (2%)	0·028	4661 (1%)	2441 (<1%)	0·042

Data are n (%) or median (IQR). The four cohorts represent vaccine rollout periods. CPRD=Clinical Practice Research Datalink. SMD=standardised mean difference. *Calculated as the days of previous observation in the database before index date per 365·25 days. †Assessed any time before the index date.

Table 1: **Baseline characteristics after propensity score weighting stratified by staggered cohort and exposure status (any COVID-19 vaccine) in CPRD AURUM**

We summarised the number of people included in each of the analyses, the number of COVID-19 cases, and the number of people who developed long COVID across cohorts and databases stratified by vaccination status (table 2). Outcome counts from sensitivity analyses, stratified for vaccine and without censoring at the second vaccine dose for vaccinated people, are available in the appendix (pp 64–90).

Vaccination with any COVID-19 first vaccine dose (ChAdOx1, BNT162b2, Janssen (Ad26.COV2.S), and mRNA-1273) was associated with a reduced risk of developing long COVID across all databases (figure 2), with meta-analytic calibrated sHRs of 0·54 (95% CI 0·44–0·67) in CPRD GOLD, 0·48 (0·34–0·68) in CPRD AURUM, 0·71 (0·55–0·91) in SIDIAP, and 0·59 (0·40–0·87) in CORIVA. Kaplan-Meier and log(–log) plots are shown in the appendix (pp 108–109, 134–135).

Results for sensitivity analyses showed similar findings to the main results, with meta-analytic calibrated sHRs of 0·59 (0·50–0·70) in CPRD GOLD, 0·50 (0·35–0·74) in CPRD AURUM, 0·74 (0·57–0·97) in SIDIAP, and 0·60 (0·43–0·85) in CORIVA when using a definition of long COVID symptoms recorded from 28 days (instead of from 90 days). Similarly, meta-analytic calibrated sHRs for the association with post-acute COVID-19 as coded in SNOMED were 0·30 (0·21–0·44) for CPRD AURUM and 0·34 (0·22–0·56) for SIDIAP. Sensitivity analyses without censoring at the second vaccine dose for vaccinated people are shown in the appendix (pp 94–100, 108–131).

Follow-up time for vaccinated and unvaccinated people, censoring proportions and reasons for censoring, and vaccine effectiveness against COVID-19 are presented in the appendix (pp 91–92). Estimates from Cox regression are presented in the appendix (p 93) and were very similar to Fine-Gray estimates. Estimates of vaccine effectiveness to prevent each of the individual WHO-listed long COVID symptoms are provided in the web application.

Analyses stratified by vaccine brand were conducted for BNT162b2 and ChAdOx1. No analyses were done for mRNA-1273 and Ad26.COV2.S, or for ChAdOx1 in cohorts one and four (SIDIAP) and cohorts two, three, and four (CORIVA), as pre-specified study diagnostics showed insufficient sample size and strong evidence of unresolved confounding despite propensity score weighting. Baseline characteristics before and after weighting are included in the appendix (pp 18–25, 30–37, 42–47, and 52–55).

The effectiveness of each of the studied COVID-19 vaccines (BNT162b2 and ChAdOx1) against the risk of developing long COVID is shown in figure 2. Meta-analytic calibrated sHRs for ChAdOx1 were 0·59 (0·37–0·93) in CPRD GOLD, 0·56 (0·38–0·83) in CPRD AURUM, and 0·76 (0·43–1·35) in SIDIAP. For BNT162b2, the sHRs were 0·53 (0·39–0·72) in CPRD GOLD, 0·42 (0·29–0·62) in CPRD AURUM, 0·65 (0·51–0·83) in SIDIAP, and 0·65 (0·44–0·96) in CORIVA.

	Vaccinated			Unvaccinated		
	Number of individuals	With COVID-19	With long COVID	Number of individuals	With COVID-19	With long COVID
CPRD AURUM						
Cohort 1	552 602	3620	904 (25·0%)	346 674	4064	734 (18·1%)
Cohort 2	1 563 569	8803	1435 (16·3%)	1 975 726	68 744	4949 (7·2%)
Cohort 3	1 528 031	8011	923 (11·5%)	1 510 401	72 127	3509 (4·9%)
Cohort 4	2 085 598	79 203	3899 (4·9%)	2 027 763	153 898	4762 (3·1%)
CORIVA						
Cohort 1	25 780	453	140 (30·9%)	22 390	3529	881 (25·0%)
Cohort 2	4208	39	11 (28·2%)	31 574	4937	1236 (25·0%)
Cohort 3	20 800	472	101 (21·4%)	87 466	15 529	3545 (22·8%)
Cohort 4	18 233	1555	293 (18·8%)	132 693	27 493	5742 (21·0%)
CPRD GOLD						
Cohort 1	118 507	392	86 (21·9%)	169 100	617	82 (13·3%)
Cohort 2	486 619	1071	178 (16·6%)	583 399	10 751	578 (5·4%)
Cohort 3	462 832	1414	103 (7·3%)	417 996	9968	314 (3·2%)
Cohort 4	550 437	12 067	289 (2·4%)	469 876	20 316	343 (1·7%)
SIDIAP						
Cohort 1	89 941	416	114 (27·4%)	223 947	7288	997 (13·7%)
Cohort 2	819 590	6199	1593 (25·7%)	433 161	30 694	3766 (12·3%)
Cohort 3	954 232	9609	1625 (16·9%)	869 456	89 301	9454 (10·6%)
Cohort 4	880 950	46 574	6624 (14·2%)	1 061 913	182 965	20 795 (11·4%)

Data are N or n (%). Number of vaccinated and unvaccinated participants, individuals with COVID-19, and individuals with at least one long COVID symptom between 90 and 365 days after SARS-CoV-2 infection across cohorts and databases, stratified by exposure status (any COVID-19 vaccination) and before overlap weighting was applied. Unvaccinated people can be included in different cohorts and contribute to event counts multiple times. CPRD=Clinical Practice Research Datalink. SIDIAP=Information System for Research in Primary Care.

Table 2: **Primary study outcomes**

Baseline characteristics for recipients of BNT162b2 and ChAdOx1 before and after propensity score weighting are reported in the appendix (pp 56–63), and confirm the comparability (ie, they are similar enough for a comparison) of the included participants for comparative effectiveness analyses. Across all cohorts, a slightly stronger preventative effect was seen for the first dose of BNT162b2 than for ChAdOx1, with meta-analytic calibrated sHR of 0·84 (0·74–0·94) in CPRD AURUM (this was non-significant in CPRD GOLD at 0·85 [0·60–1·20] figure 3). Results for sensitivity analyses were consistent with these findings and are detailed in the appendix (pp 101–107).

Discussion

Our analyses of more than 20 million vaccinated and unvaccinated people show the clinical effectiveness of COVID-19 vaccines to prevent the development of long COVID in three European countries (the UK, Estonia, and Spain), with overall vaccine effectiveness ranging from 29% to 52%. These findings were robust to multiple sensitivity analyses and various definitions of long COVID, including different duration of symptoms and clinically diagnosed long COVID in a secondary analysis. A slightly stronger preventive effect was seen for BNT162b2 than ChAdOx1 in CPRD AURUM.

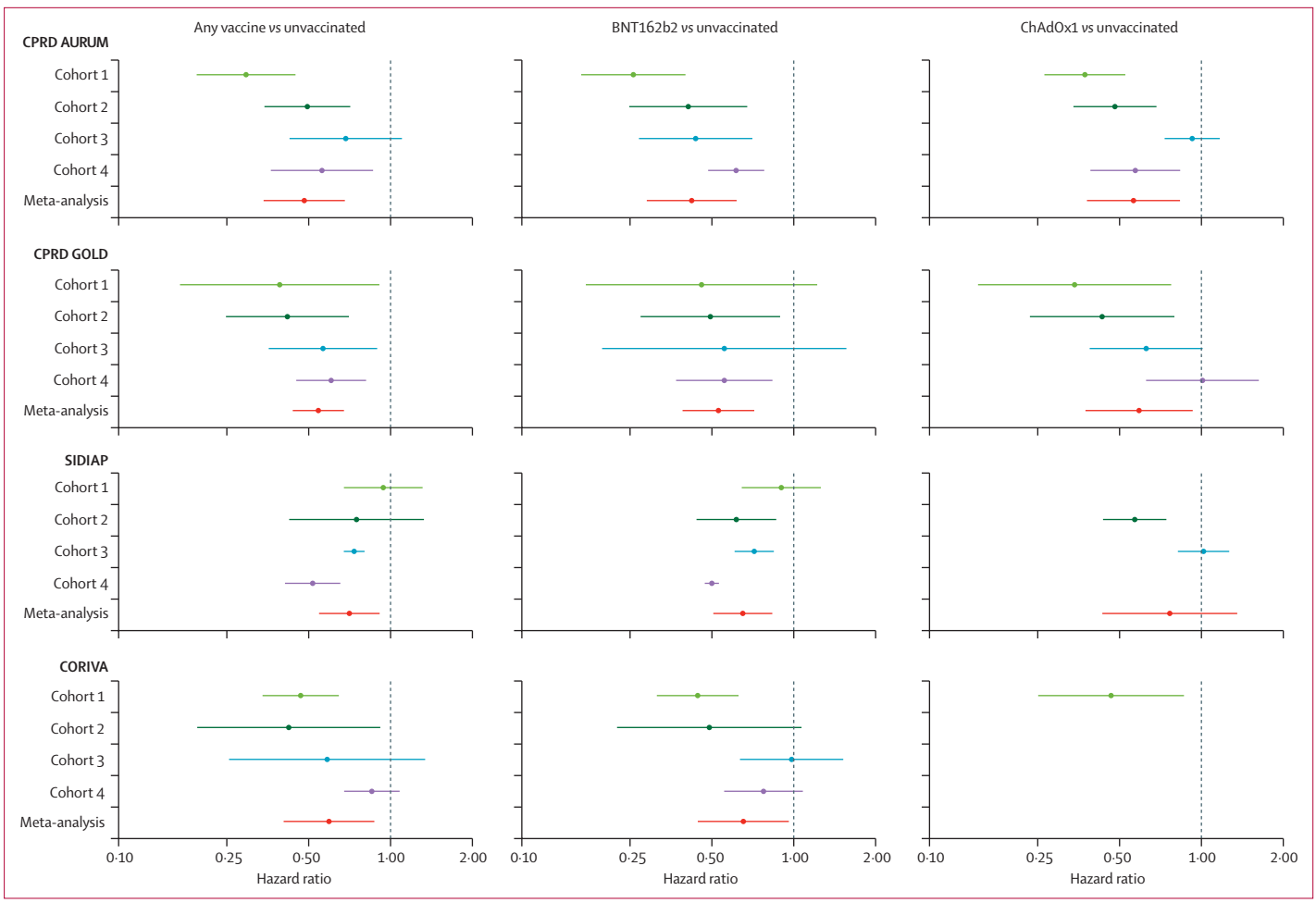


Figure 2: **Forest plots of vaccine effectiveness against long COVID**
Calibrated subdistribution hazard ratios from CPRD GOLD, CPRD AURUM, SIDIAP, and CORIVA for cohorts one to four and meta-analyses. Comparative effectiveness analyses for ChAdOx1 in SIDIAP and CORIVA were not fully conducted due to small sample sizes and restrictions for the use of ChAdOx1 in Estonia and Spain. CPRD=Clinical Practice Research Datalink. SIDIAP=Information System for Research in Primary Care.

To our knowledge, this is the first multinational study assessing population-level vaccine effectiveness to prevent long COVID symptoms. Most previous studies assessed pre-infection vaccination and long COVID only among people with COVID-19,[19–30] thus overlooking the effect of vaccines to prevent SARS-CoV-2 infection as part of the pathway to developing post-COVID-19 complications. Our results showed vaccination to protect against COVID-19 and long COVID. However, previous studies that included people with and without SARS-CoV-2 infection[44–46] did not aim to estimate population-level vaccine effectiveness to prevent long COVID. Therefore, our work addresses this gap in the current literature.

In line with our results, a recent meta-analysis showed lower risk of developing the post-COVID-19 condition for vaccinated versus unvaccinated people. Only small differences in vaccine effectiveness were reported with one dose (0·60 [0·43–0·83]) or two doses (0·64 [0·45–0·92]).[18] Two large-scale studies from the USA based on the Veteran Affairs national health-care database and the National COVID Cohort Collaborative showed reduced risk for post-acute COVID-19 complications (HR 0·85 [0·82–0·89])[20] and reduced risk[47] associated with pre-COVID-19 vaccination in people with COVID-19 (HR 0·67 [0·56–0·79]). Similarly, the ZOE app study[19] reported lower odds (OR 0·51 [0·32–0·82]) of persistent symptoms (≥28 days) after full vaccination (with two doses) in 2370 vaccinated case and control participants. However, these studies were not representative of the general population, and only included people with COVID-19 infection, not those without.

Our study also compared different vaccines, with a slightly higher vaccine effectiveness to prevent persisting COVID-19 symptoms seen for BNT162b2 compared with ChAdOx1. Higher efficacy for preventing COVID-19 infection has been shown with BNT162b2 (95% [95 % CI

90·3–97·6%])[5] versus ChAdOx1 (70·4% [54·8–80·6%]) in interim analyses of randomised controlled trials, with similar findings in routinely collected data.[6,39] This difference in vaccine efficacy could possibly explain our observation of slightly stronger protective effects for BNT162b2, given that people might have been less likely to be infected with SARS-CoV-2 in the first place.

Although previous studies highlighted several risk factors for developing long COVID, many of those are not modifiable (eg, age and sex). A recent meta-analysis showed that people with severe COVID-19, including previous hospitalisation or intensive care admission, were at higher risk of developing post-COVID-19 complications.[48] Although our study did not take the severity of COVID-19 into account, effectiveness of vaccines to prevent severe COVID-19 could have contributed to reduced long COVID risk in vaccinated people. The effectiveness of vaccines varied throughout the pandemic, with waning immunity over time and reduced effectiveness against new variants being reported.[49] Persistent COVID-19 symptoms were more common in people infected with pre-delta[50] and delta[51] variants compared with omicron. However, a recent study showed that these differences were no longer significant after accounting for vaccination status.[50] Our study period covered the alpha and delta waves and the beginning of the omicron wave, with only SIDIAP and CORIVA including the full omicron wave. This difference in included waves of predominant variants could contribute to a higher number of people with COVID-19 infections due to omicron,[51,52] and larger proportions of people with long COVID symptoms in SIDIAP and CORIVA compared with CPRD, particularly for cohorts three and four, probably due to longer follow-up available to assess persisting symptoms.

Our study had limitations. First, although health-care workers and carers were prioritised for vaccination, we could not include them for analyses in cohorts one and two as participants' professions were not accurately recorded in the databases. We do not expect this limitation to affect our overall results, as no consistent evidence of heterogeneity in effectiveness was noted across the four proposed study periods. Secondly, we expect some degree of outcome misclassification due to changes in the definition of long COVID over time. Although our main definition of long COVID requiring one recorded symptom is considered very broad and probably overestimates cases, under-reporting of some symptoms in health-care records might lead to underestimated long COVID rates. Moreover, under-reporting of COVID-19 due to missed tests or diagnoses in asymptomatic cases is expected, although the rise in self-administered screening tests might possibly mitigate this issue as asymptomatic cases might have been picked up during routine testing for contact people in workplaces, schools, and for travel. We addressed possible outcome misclassification by adding secondary outcome definitions, including the reportedly underused[53] but very specific diagnosis code for post-acute COVID-19. Reassuringly, results were very similar to our main findings using the symptom-based definition. Given the observational nature of our data, we cannot guarantee the absence of confounding, which could partly account for our findings. Lastly, vaccine waning probably leads to lesser effects over time, and research on the protective effects of vaccination against long COVID in the long term remains necessary.

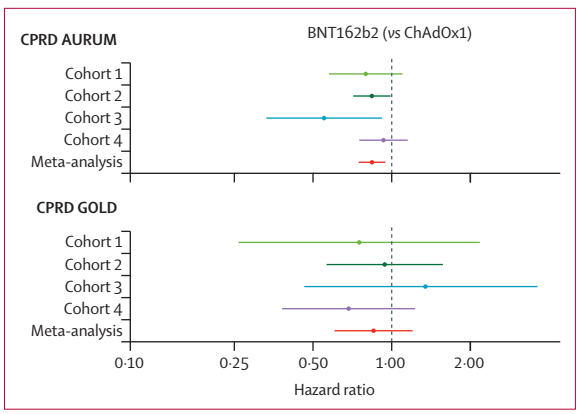


Figure 3: **Forest plots of comparative effectiveness against long COVID**
Empirically calibrated subdistribution hazard ratios obtained from CPRD AURUM and CPRD GOLD. CPRD=Clinical Practice Research Datalink.

Our study also has several strengths. Although previous studies were conducted among people with COVID-19, we estimated the effectiveness of COVID-19 vaccination to prevent long COVID at the population level. This approach allowed us to assess vaccine effectiveness for long COVID prevention, including the vaccines effect on preventing SARS-CoV-2 infections, which is a crucial part of the pathway to prevent post-COVID-19 complications. Vaccination records were directly retrieved via linkage, allowing for completeness of records for vaccination received in all four databases. Therefore, misclassification of unvaccinated people is minimal in all contributing data sources. We used state-of-the art methods and diagnostics to identify and minimise residual confounding. Similar approaches were not previously used in the literature on this topic: propensity score weighting successfully minimised measured confounding, including balancing key confounders, such as age, comorbidities, calendar time (represented by index date), and location. We used empirical calibration based on NCO analyses to account for residual confounding. NCOs capture unmeasured confounding, similar to the types of confounding in the vaccination–outcome pathway. Although most of our NCOs were proxies for health-care seeking behaviour and thus, cover the predominant source of unmeasured confounding relevant to our study, remaining residual confounding cannot be ruled out. All diagnostics are available in the interactive web application. For

comparisons where confounding could not adequately be accounted for, effect estimation studies were not done.

The proposed staggered cohort design after the national vaccine rollouts is an additional strength of this study as it minimises time-varying confounding by indication. Our main results were consistent across three European countries, with different national health-care policies and vaccination strategies, highlighting the robustness of our findings.

Our study adds important information to the risk–benefit assessment of COVID-19 vaccines. Although the safety profile is now better understood, safety concerns and the lower risk of severe COVID-19 among young people trumped vaccination efforts in many parts of the world. Conversely, long COVID is known to affect young people at least as much as older populations. We hope that this study suggesting protection against long COVID in adults from cohort four (median age 34–40 years) will encourage a higher uptake of COVID-19 vaccines among previously hesitant adults. Future studies should further expand this work to assess the effect of SARS-CoV-2 variants, booster doses, and vaccine effectiveness among other subpopulations.

For more on **SIDIAP** see https://www.sidiap.org/index.php/en/solicituds-en

In conclusion, our study shows the clinical effectiveness of COVID-19 vaccines to prevent long COVID, highlighting yet another benefit of vaccination, particularly for adults. These findings were consistent across three European countries and different populations, and were robust to multiple definitions of long COVID and sensitivity analyses.

Contributors

DP-A and AMJ led the conceptualisation of the study with contributions from MC, EB, and TR-M. KK, DP-A, and AMJ led the phenotyping of long COVID symptoms. AMJ, TD-S, ER, AU, and NTTH adapted the study design with respect to the local vaccine rollouts. AD and WYM mapped and curated the Clinical Practice Research Datalink (CPRD) data. MC developed the code for statistical analyses with methodological advice from EB and TR-M. MC, NM-B, and RK executed study code in the respective databases. DP-A, MC, NTTH, TD-S, HMEN, and AMJ clinically interpreted the results. AMJ wrote the first draft of the manuscript. All authors read, contributed to, and approved the last version of the manuscript. DP-A and AMJ obtained the funding. NM-B and TD-S had access to and verified the SIDIAP data; MC, AD, and WYM had access to and verified the CPRD GOLD and CPRD AURUM data; and RK and AU had access to and verified the CORIVA data. MC, DP-A, and AMJ were responsible for the decision to submit for publication.

For more on **GitHub** see https://github.com/oxford-pharmacoepi/LongcovidVaccineEffectiveness. Additionally, we have reported all study diagnostics and results in full in an interactive web application (https://dpa-pde-oxford.shinyapps.io/LongcovidVaccineEffectiveness/)

Declaration of interests

DP-A's department has received grants from Amgen, Chiesi-Taylor, Lilly, Janssen, Novartis, UCB Biopharma, the European Medicines Agency, and the Innovative Medicines Initiative. DP-A's research group has received consultancy fees from Astra Zeneca and UCB Biopharma. DP-A's department has organised training programmes funded or supported by Amgen, Astellas, Janssen, Synapse Management Partners, and UCB Biopharma. RK's research group has received consultancy fees from AstraZeneca and the Estonian Ministry of Social Affairs through RITA CORIVA and RITA MAITT (Machine learning and AI powered public services) projects. AU reports funding from the European Regional Development Fund (RITA 1/02–120) for her institution (Department of Family Medicine and Public Health, University of Tartu). HMEN reports support from the European Health Data and Evidence Network project grant, Innovative Medicines Initiative 2 Joint Undertaking (grant agreement 806968) for harmonisation of the Norwegian registry data into the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). TD-S reports funding from the Innovative Medicines Initiative 2 Joint Undertaking (grant agreement 806968) for the institute to map the SIDIAP data to the OMOP CDM. This initiative received support from the EU Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations. All other authors declare no competing interests.

Data sharing

CPRD data were obtained under the CPRD multi-study license held by the University of Oxford after Research Data Governance approval. Direct data sharing is not allowed. In accordance with current European and national law, SIDIAP data used in this study are only available for researchers participating in this study. Thus, these data cannot be distributed or make publicly available to other parties. However, researchers from public institutions can request data from SIDIAP, if they comply with specific requirements. For further information contact SIDIAP at sidiap@idiapjgol.org. CORIVA data were obtained under the approval of the research ethics committee at the University of Tartu and patient-level data sharing is not allowed.

Acknowledgments

This manuscript reports on independent research funded by the National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre (BRC; COV-LT2–0006). DP-A is funded through a NIHR Senior Research Fellowship (grant number SRF-2018–11-ST2–004). Funding to perform the study in the SIDIAP database was provided by the Real-World Epidemiology research group at Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina. AU reports funding from the European Regional Development Fund (RITA 1/02–120). The views expressed in this publication are those of the authors and not necessarily those of NIHR, the Department of Health and Social Care, or the CPRD. We thank David Hockaday for contributing to this research providing patient perspectives from the study design to the review of our findings. This study is based in part on data from the CPRD obtained under license from the UK Medicines and Healthcare products Regulatory Agency. We thank the patients who provided these data, and the National Health Service who collected the data as part of their care and support. We also thank the health-care professionals in the Catalan health-care system involved in the management of COVID-19 during these challenging times, from primary care to intensive care units, and the Institut de Català de la Salut and the Program d'Analítica de Dades per a la Recerca i la Innovació en Salut for providing access to the different data sources accessible through SIDIAP. The lead authors affirm that the manuscript is an honest, accurate, and transparent account of the study being reported, and that no important aspects of the study have been omitted. To increase transparency and reproducibility, we have made all our analytical code available on GitHub.

References

1 UK Health Security Agency. Guidance: COVID-19: epidemiology, virology and clinical features. 2022. https://www.gov.uk/government/publications/wuhan-novel-coronavirus-background-information/wuhan-novel-coronavirus-epidemiology-virology-and-clinical-features (accessed July 19, 2023).

2 WHO. Weekly epidemiological update on COVID-19. 2023. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-september-2023 (accessed July 19, 2023).

3 European Medicines Agency. COVID-19 medicines: authorised COVID-19 vaccines. 2023. https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/covid-19-medicines (accessed July 19, 2023).

4 European Centre for Disease Prevention and Control. COVID-19 vaccine tracker. 2021. https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#distribution-tab (accessed May 22, 2022).

5 Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N Engl J Med* 2020; **383:** 2603–15.

6 Voysey M, Clemens SAC, Madhi SA, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 2021; **397:** 99–111.

7 Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021; **384:** 403–16.

8 Lopez Bernal J, Andrews N, Gower C, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *BMJ* 2021; **373:** n1088.

9 Eyre DW, Taylor D, Purver M, et al. Effect of COVID-19 vaccination on transmission of alpha and delta variants. *N Engl J Med* 2022; **386:** 744–56.

10 Office for National Statistics. Statistical bulletin. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK. 2021. https://www.ons.gov.uk/peoplepopulationand community/healthandsocialcare/conditionsanddiseases/bulletins/ prevalenceofongoingsymptomsfollowingcoronavirus covid19infectionintheuk/1april2021 (accessed April 4, 2021).

11 Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute COVID-19 in primary care. *BMJ* 2020; **370:** m3026.

12 Subramanian A, Nirantharakumar K, Hughes S, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat Med* 2022; **28:** 1706–14.

13 Notarte KI, de Oliveira MHS, Peligro PJ, et al. Age, sex and previous comorbidities as risk factors not associated with SARS-CoV-2 infection for long COVID-19: a systematic review and meta-analysis. *J Clin Med* 2022; **11:** 7314.

14 WHO. A clinical case definition of post COVID-19 condition by a Delphi consensus. 2021. https://www.who.int/publications/i/item/ WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_ definition-2021.1. (accessed Oct 6, 2021).

15 Harrison S, Walters B, Simmons Z, Cook M, Clark R. The effectiveness of vaccination against long COVID: a rapid evidence briefing UKHSA. 2022. https://ukhsa.koha-ptfs.co.uk/cgi-bin/ koha/opac-retrieve-file.pl?id=fe4f10cd3cd509fe045ad4f72ae0dfff (accessed Feb 20, 2022).

16 Byambasuren O, Stehlik P, Clark J, Alcorn K, Glasziou P. Effect of COVID-19 vaccination on long covid: systematic review. *BMJ Med* 2023; **2:** e000385.

17 Notarte KI, Catahay JA, Velasco JV, et al. Impact of COVID-19 vaccination on the risk of developing long-COVID and on existing long-COVID symptoms: a systematic review. *EClinicalMedicine* 2022; **53:** 101624.

18 Watanabe A, Iwagami M, Yasuhara J, Takagi H, Kuno T. Protective effect of COVID-19 vaccination against long COVID syndrome: a systematic review and meta-analysis. *Vaccine* 2023; **41:** 1783–90.

19 Antonelli M, Penfold RS, Merino J, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis* 2022; **22:** 43–55.

20 Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. *Nat Med* 2022; **28:** 1461–67.

21 Ayoubkhani D, Bosworth ML, King S, et al. Risk of long COVID in people infected with severe acute respiratory syndrome coronavirus 2 after 2 doses of a coronavirus disease 2019 vaccine: community-based, matched cohort study. *Open Forum Infect Dis* 2022; **9:** ofac464.

22 Azzolini E, Levi R, Sarti R, et al. Association between BNT162b2 vaccination and long covid after infections not requiring hospitalization in health care workers. *JAMA* 2022; **328:** 676–78.

23 Arjun MC, Singh AK, Pal D, et al. Characteristics and predictors of long COVID among diagnosed cases of COVID-19. *PLoS One* 2022; **17:** e0278825.

24 Emecen AN, Keskin S, Turunc O, et al. The presence of symptoms within 6 months after COVID-19: a single-center longitudinal study. *Ir J Med Sci* 2023; **192:** 741–50.

25 Ioannou GN, Baraff A, Fox A, et al. Rates and factors associated with documentation of diagnostic codes for long COVID in the National Veterans Affairs Health Care System. *JAMA Netw Open* 2022; **5:** e2224359.

26 Mohr NM, Plumb ID, Harland KK, et al. Presence of symptoms 6 weeks after COVID-19 among vaccinated and unvaccinated US healthcare personnel: a prospective cohort study. *BMJ Open* 2023; **13:** e063141.

27 Tannous J, Pan AP, Potter T, et al. Real-world effectiveness of COVID-19 vaccines and anti-SARS-CoV-2 monoclonal antibodies against postacute sequelae of SARS-CoV-2: analysis of a COVID-19 observational registry for a diverse US metropolitan population. *BMJ Open* 2023; **13:** e067611.

28 Taquet M, Dercon Q, Harrison PJ. Six-month sequelae of post-vaccination SARS-CoV-2 infection: a retrospective cohort study of 10,024 breakthrough infections. *Brain Behav Immun* 2022; **103:** 154–62.

29 Zisis SN, Durieux JC, Mouchati C, Perez JA, McComsey GA. The protective effect of coronavirus disease 2019 (COVID-19) vaccination on postacute sequelae of COVID-19: a multicenter study from a large national health research network. *Open Forum Infect Dis* 2022; **9:** ofac228.

30 Brannock MD, Chew RF, Preiss AJ, et al. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. *Nat Commun* 2023; **14:** 2914.

31 Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research datalink (CPRD). *Int J Epidemiol* 2015; **44:** 827–36.

32 Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019; **48:** 1740-1740g.

33 Recalde M, Rodríguez C, Burn E, et al. Data resource profile: the information system for research in primary care (SIDIAP). *Int J Epidemiol* 2022; **51:** e324–36.

34 Ryan, P et al. The book of OHDSI: chapter 4—the common data model. 2021,=. https://ohdsi.github.io/TheBookOfOhdsi/ CommonDataModel.html (accessed June 12, 2023).

35 Raventós B, Fernández-Bertolín S, Aragón M, et al. Transforming the information system for research in primary care (SIDIAP) in Catalonia to the OMOP common data model and its use for COVID-19 research. *Clin Epidemiol* 2023; **15:** 969–86.

36 UK Government Department of Health & Social Care. Guidance on shielding and protecting people who are clinically extremely vulnerable from COVID-19. 2021. https://www.gov.uk/government/ publications/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19/guidance-on-shielding-and-protecting-extremely-vulnerable-persons-from-covid-19 (accessed April 30, 2021).

37 UK Government Department of Health & Social Care. Joint committee on vaccination and immunisation: advice on priority groups for COVID-19 vaccination, 30 December 2020. 2021. https:// www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020/joint-committee-on-vaccination-and-immunisation-advice-on-priority-groups-for-covid-19-vaccination-30-december-2020 (accessed April 30, 2021).

38 Kostka K, Roel E, Trinh N, Delmestri A, Mateu L, Paredes R, et al. The burden of long COVID: a multinational cohort analysis of Spanish and UK data including SARS-CoV-2 infections, reinfections, and matched contemporaneous test negative controls. *Res Square* 2023; published online March 14. https://doi.org/10.21203/rs.3.rs-2642600/v1 (preprint).

39 Català M, Burn E, Rathod-Mistry T, et al. Observational methods for COVID-19 vaccine effectiveness research: a trial emulation and empirical evaluation. *Int J Epidemiol* 2023; published online Oct 13. https://doi.org/10.1093/ije/dyad138.

40 Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Med Res Methodol* 2017; **17:** 78.

41 Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010; **21:** 383–88.

42 Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci USA* 2018; **115:** 2571–77.

43 Shi X, Miao W, Tchetgen ET. A Selective review of negative control methods in epidemiology. *Curr Epidemiol Rep* 2020; **7:** 190–202.

44 Hastie CE, Lowe DJ, McAuley A, et al. Outcomes among confirmed cases and a matched comparison group in the long-COVID in Scotland study. *Nat Commun* 2022; **13:** 5663.

45 van der Maaden T, Mutubuki EN, de Bruijn S, et al. Prevalence and severity of symptoms 3 months after infection with SARS-CoV-2 compared to test-negative and population controls in the Netherlands. *J Infect Dis* 2023; **227:** 1059–67.

46 Kuodi P, Gorelik Y, Zayyad H, et al. Association between BNT162b2 vaccination and reported incidence of post-COVID-19 symptoms: cross-sectional study 2020-21, Israel. *NPJ Vaccines* 2022; **7:** 101.

47 Brannock MD, Chew RF, Preiss AJ, et al. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. *Nat Commun* 2023; **14:** 2914.

48 Tsampasian V, Elghazaly H, Chattopadhyay R, et al. Risk factors associated with post-COVID-19 condition: a systematic review and meta-analysis. *JAMA Intern Med* 2023; **183:** 566–80.

49 Andrews N, Stowe J, Kirsebom F, et al. COVID-19 vaccine effectiveness against the omicron (B.1·1·529) variant. *N Engl J Med* 2022; **386:** 1532–46.

50 Gottlieb M, Wang RC, Yu H, et al. Severe fatigue and persistent symptoms at 3 months following severe acute respiratory syndrome coronavirus 2 infections during the pre-delta, delta, and omicron time periods: a multicenter prospective cohort study. *Clin Infect Dis* 2023; **76:** 1930–41.

51 Antonelli M, Pujol JC, Spector TD, Ourselin S, Steves CJ. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. *Lancet* 2022; **399:** 2263–64.

52 Fernández-de-Las-Peñas C, Notarte KI, Peligro PJ, et al. Long-COVID symptoms in individuals infected with different SARS-CoV-2 variants of concern: a systematic review of the literature. *Viruses* 2022; **14:** 2629.

53 Walker AJ, MacKenna B, Inglesby P, et al. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records *in situ* using OpenSAFELY. *Br J Gen Pract* 2021; **71:** e806–14.