

1 Six-month sequelae of post-vaccination SARS-CoV-2  
2 infection: a retrospective cohort study of 10,024  
3 breakthrough infections  
4

5 Maxime Taquet, PhD,<sup>1,2</sup> Quentin Dercon, MSc,<sup>1,3</sup> Paul J Harrison, FRCPsych<sup>1,2</sup>

6  
7 <sup>1</sup>Department of Psychiatry, University of Oxford, Oxford, U.K.

8 <sup>2</sup>Oxford Health NHS Foundation Trust, Oxford, U.K.

9 <sup>3</sup>MRC Cognition & Brain Sciences Unit, University of Cambridge, U.K.

10  
11 Correspondence: Maxime Taquet, Department of Psychiatry, University of Oxford,  
12 Warneford Hospital, Oxford OX3 7JX, United Kingdom, [maxime.taquet@medsci.ox.ac.uk](mailto:maxime.taquet@medsci.ox.ac.uk)

13  
14  
15 3492 words | 1 Table | 3 Figures | 28 References | Appendix

## 16 **Abstract**

17 Vaccination has proven effective against infection with SARS-CoV-2, as well as death and  
18 hospitalisation following COVID-19 illness. However, little is known about the effect of  
19 vaccination on other acute and post-acute outcomes of COVID-19. Data were obtained from  
20 the TriNetX electronic health records network (over 81 million patients mostly in the USA).  
21 Using a retrospective cohort study and time-to-event analysis, we compared the incidences of  
22 COVID-19 outcomes between individuals who received a COVID-19 vaccine (approved for  
23 use in the USA) at least 2 weeks before SARS-CoV-2 infection and propensity score-  
24 matched individuals unvaccinated for COVID-19 but who had received an influenza vaccine.  
25 Outcomes were ICD-10 codes representing documented COVID-19 sequelae in the 6 months  
26 after a confirmed SARS-CoV-2 infection (recorded between January 1 and August 31, 2021).  
27 Associations with the number of vaccine doses (1 vs. 2) and age (< 60 vs.  $\geq$  60 years-old)  
28 were assessed. Among 10,024 vaccinated individuals with SARS-CoV-2 infection, 9479  
29 were matched to unvaccinated controls. Receiving at least one COVID-19 vaccine dose was  
30 associated with a significantly lower risk of respiratory failure, ICU admission,  
31 intubation/ventilation, hypoxaemia, oxygen requirement, hypercoagulopathy/venous  
32 thromboembolism, seizures, psychotic disorder, and hair loss (each as composite endpoints  
33 with death to account for competing risks; HR 0.70-0.83, Bonferroni-corrected  $p < .05$ ), but  
34 not other outcomes, including long-COVID features, renal disease, mood, anxiety, and sleep  
35 disorders. Receiving 2 vaccine doses was associated with lower risks for most outcomes.  
36 Associations between prior vaccination and outcomes of SARS-CoV-2 infection were  
37 marked in those < 60 years-old, whereas no robust associations were observed in those  $\geq$  60  
38 years-old. In summary, COVID-19 vaccination is associated with lower risk of several, but  
39 not all, COVID-19 sequelae in those with breakthrough SARS-CoV-2 infection. These  
40 benefits of vaccination were clear in younger people but not in the over-60s. The findings

- 41 may inform service planning, contribute to forecasting public health impacts of vaccination  
42 programmes, and highlight the need to identify additional interventions for COVID-19  
43 sequelae.

## 44 **Introduction**

45 The observation that individuals can be infected with SARS-CoV-2 after being vaccinated  
46 against COVID-19 (so-called breakthrough infections) has caused concerns (Nixon and  
47 Ndhlovu, 2021). These concerns are mitigated by abundant evidence that the risk of severe  
48 COVID-19 illness (as proxied by hospitalisation, admission to intensive care unit, and  
49 mortality) is lessened by vaccination (Agrawal et al., 2021; Antonelli et al., 2021; Bahl et al.,  
50 2021; Butt et al., 2021; Cabezas et al., 2021; Glatman-Freedman et al., 2021; Haas et al.,  
51 2021; Hyams et al., 2021; Mateo-Urdiales et al., 2021; Roest et al., 2021). One case-control  
52 study investigated the association between self-reported SARS-CoV-2 infection in 908 pairs  
53 of vaccinated and unvaccinated individuals and self-reported symptoms beyond 28 days  
54 (Antonelli et al., 2021). It found that compared to unvaccinated individuals, those with  
55 breakthrough infection were at a lower risk of symptoms beyond 28 days.

56  
57 However, how COVID-19 vaccination affects the broad spectrum of sequelae of SARS-CoV-  
58 2 infection remains elusive. In particular, it is unknown if vaccinated individuals are at the  
59 same risk as unvaccinated individuals of venous thromboembolisms, ischaemic strokes,  
60 neuropsychiatric complications, long-COVID presentations, and other post-acute sequelae  
61 following infection with SARS-CoV-2. In addition, unvaccinated people might have health  
62 behaviours related to vaccination hesitancy (Latkin et al., 2021), and this potential source of  
63 bias has not been addressed by previous studies.

64  
65 This cohort study based on electronic health records compares the 6-months outcomes of  
66 SARS-CoV-2 infection among individuals who were (vs. those who were not) vaccinated  
67 against COVID-19. We investigated a range of outcomes (both acute and post-acute) with  
68 documented associations with COVID-19 across multiple body systems.

## 69 **Methods**

### 70 **Data and study design**

71 The study used TriNetX Analytics, a federated network of linked EHRs recording  
72 anonymized data from 59 healthcare organizations (HCOs), primarily in the USA, totalling  
73 81 million patients. Available data include demographics, diagnoses (ICD-10 codes),  
74 procedures (Current Procedural Terminology [CPT] codes), and measurements (e.g. blood  
75 pressure). The HCOs consist in a mixture of primary care centres, hospitals, and specialist  
76 units. They provide data from uninsured as well as insured individuals. Data de-identification  
77 is attested to and receives a formal determination by a qualified expert as defined in Section  
78 §164.514(b)(1) of the HIPAA Privacy Rule. This formal determination supersedes TriNetX's  
79 waiver from the Western Institutional Review Board (IRB). Using the TriNetX user interface,  
80 cohorts are created based on inclusion and exclusion criteria, matched for confounding  
81 variables, and compared for outcomes of interest over specified time periods. For further  
82 details about TriNetX, see Appendix pp. 1-2.

83

### 84 **Cohorts**

85 Both the primary and control cohorts were defined as all patients who had, between January  
86 1, 2021 and August 31, 2021, a confirmed SARS-CoV-2 infection, which we defined as  
87 either a confirmed diagnosis of COVID-19 (ICD-10 code U07.1) or a first positive PCR test  
88 for SARS-CoV-2. In the primary cohort, patients were included only if their confirmed  
89 SARS-CoV-2 infection occurred at least 14 days after a recorded administration of a COVID-  
90 19 vaccine approved for use in the USA (i.e. BNT162b2 'Pfizer/BioNTech', mRNA-1273  
91 'Moderna', or Ad26.COV2.S 'Janssen'). In the control cohort, patients were included only if  
92 no recorded vaccine against COVID-19 was recorded before their SARS-CoV-2 infection  
93 and if they had received a vaccine against influenza at any time. In the USA, the Centers for

94 Disease Control and Prevention (CDC) recommend yearly influenza vaccination to everyone  
95 over the age of 6 months. This inclusion criterion thus excludes patients with obvious vaccine  
96 hesitancy (so-called ‘anti-vaxxers’) as this is correlated with other health-related behaviours  
97 that might confound associations with COVID-19 outcomes (Latkin et al., 2021). More  
98 details are provided in the Appendix pp. 2-4.

99

## 100 **Covariates**

101 A set of established and suspected risk factors for COVID-19 and for more severe COVID-19  
102 illness was used (de Lusignan et al., 2020; Taquet et al., 2021c; Williamson et al., 2020): age,  
103 sex, race, ethnicity, obesity, hypertension, diabetes, chronic kidney disease, asthma, chronic  
104 lower respiratory diseases, nicotine dependence, substance misuse, ischaemic heart disease  
105 and other forms of heart disease, socioeconomic deprivation, cancer (and haematological  
106 cancer in particular), chronic liver disease, stroke, dementia, organ transplant, rheumatoid  
107 arthritis, lupus, psoriasis, and disorders involving an immune mechanism. To capture these  
108 risk factors in patients’ health records, 55 variables were used. More details including ICD-10  
109 codes are provided in the Appendix pp. 4-5. Cohorts were matched for all these variables, as  
110 described below. In addition, cohorts were stratified by the date of the SARS-CoV-2  
111 infection in 2-monthly periods (January 1 to February 28, 2021, March 1 to April 30, 2021,  
112 May 1 to June 30, 2021, and July 1 to August 31, 2021) and matching was achieved  
113 independently within each period (guaranteeing that as many patients in the matched cohorts  
114 had their SARS-CoV-2 infection in every 2-month period).

115

## 116 **Outcomes**

117 We investigated the 6-month incidence of all acute and post-acute outcomes which have been  
118 shown to be significantly associated with COVID-19 in four large-scale studies based on

119 electronic health records (Al-Aly et al., 2021; Daugherty et al., 2021; Taquet et al., 2021b,  
120 2021a), namely:

- 121     ▪ Hospitalisation
- 122     ▪ Intensive Care Unit (ICU) admission
- 123     ▪ Death
- 124     ▪ Intubation/Ventilation
- 125     ▪ Respiratory failure
- 126     ▪ Hypoxaemia
- 127     ▪ Oxygen requirement
- 128     ▪ Long COVID feature (any and each of the following as defined in Ref. 18)
  - 129         - Abdominal symptoms
  - 130         - Abnormal breathing
  - 131         - Anxiety/Depression
  - 132         - Chest/Throat pain
  - 133         - Cognitive symptoms
  - 134         - Fatigue
  - 135         - Headache
  - 136         - Myalgia
  - 137         - Other pain
- 138     ▪ Hypertension
- 139     ▪ Arrhythmia
- 140     ▪ Cardiac failure
- 141     ▪ Cardiomyopathy
- 142     ▪ Myocarditis
- 143     ▪ Coronary disease

- 144       ▪   Hypercoagulopathy/Deep vein thrombosis (DVT)/Pulmonary embolism (PE)
- 145       ▪   Ischaemic stroke
- 146       ▪   Cerebral haemorrhage
- 147       ▪   Peripheral neuropathy
- 148       ▪   Seizures
- 149       ▪   Type 2 diabetes mellitus
- 150       ▪   Liver disease
- 151       ▪   Kidney disease
- 152       ▪   Interstitial lung disease
- 153       ▪   Urticaria
- 154       ▪   Sleep disorders
- 155       ▪   Gastro-oesophageal reflux disease (GORD)
- 156       ▪   Hair loss
- 157       ▪   Hyperlipidaemia
- 158       ▪   Joint pain
- 159       ▪   Obesity
- 160       ▪   Anosmia
- 161       ▪   Nerve/nerve root/plexus disorder
- 162       ▪   Myoneural junction/muscle disease
- 163       ▪   Psychotic disorder
- 164       ▪   Mood disorder
- 165       ▪   Anxiety disorder
- 166
- 167   Each outcome was defined as the set of corresponding ICD-10, CPT, and VA Formulary
- 168   codes as specified in the original study which showed their association with COVID-19. To

169 account for death as a competing risk and thus address survivorship bias, each outcome was  
170 analysed as part of a composite outcome with death as the other component (Manja et al.,  
171 2017). While necessary to address survivorship bias, this approach is unlikely to provide  
172 insight into the risk of outcomes which are much rarer than death itself. The analysis was  
173 performed on October 12, 2021. More details, and ICD-10/CPT/VA Formulary codes, are  
174 provided in the Appendix pp. 5-6.

175

### 176 **Secondary analyses**

177 We assessed whether the associations between prior vaccinations and outcomes of SARS-  
178 CoV-2 infection were moderated by the number of vaccine doses received and by age at the  
179 time of infection. This was achieved by restricting the primary cohort to (i) those who had  
180 received only one vaccine dose at least 14 days before infection, (ii) those who had received  
181 two vaccine doses 14 days before infection, (iii) those <60 years-old, and (iv) those  $\geq 60$   
182 years-old. For the latter two subgroup analyses, the control cohorts were also restricted to  
183 those <60 and  $\geq 60$  years-old respectively.

184

### 185 **Statistical analyses**

186 Propensity score matching (carried out within the TriNetX network) was used to create  
187 cohorts with matched baseline characteristics (Austin, 2011). Propensity score 1:1 matching  
188 used a greedy nearest neighbour approach with a caliper distance of 0.1 pooled standard  
189 deviations of the logit of the propensity score. Any characteristic with a standardized mean  
190 difference (SMD) between cohorts lower than 0.1 is considered well matched (Haukoos and  
191 Lewis, 2015). The Kaplan-Meier estimator was used to estimate the incidence of each  
192 outcome. Hazard ratios (HRs) with 95% confidence intervals were calculated using the Cox  
193 model and the null hypothesis of no difference between cohorts was tested using log-rank

194 tests. The proportional hazard assumption was tested using the generalized Schoenfeld  
195 approach. When the assumption was violated, a time-varying HR was estimated using natural  
196 cubic splines fitted to the log-cumulative hazard (Royston and Parmar, 2002). The  
197 contribution of the individual outcomes of interest within the composite endpoint (with death  
198 as the other component) was reported as the number of events of interest over the total  
199 number of events (e.g. the number of respiratory failures over the number of respiratory  
200 failures or deaths).

201  
202 We statistically tested whether the associations between prior COVID-19 vaccine and  
203 outcomes of SARS-CoV-2 infection was moderated by age. This was achieved for each  
204 outcome by testing whether the ratio between the HR in those  $\geq 60$  and the HR in those  $< 60$   
205 years-old was statistically significantly different from 1. Similarly, moderation by the number  
206 of vaccine doses was achieved by testing whether the ratio between the HR in those who had  
207 received 1 dose and the HR in those who had received 2 doses at the time of infection was  
208 statistically significantly different from 1.

209  
210 Further details are provided in the appendix p. 7. Statistical analyses were conducted in R  
211 version 3.6.3 except for the log-rank tests which were performed within TriNetX. Statistical  
212 significance was set at two-sided p-values  $< 0.05$ . Bonferroni correction for multiple  
213 comparisons was applied to correct for the simultaneous assessments of 45 outcomes in the  
214 primary analysis. A REporting of studies Conducted using Observational Routinely-collected  
215 health Data (RECORD) statement was completed (see Appendix).

## 216 **Results**

217 A total of 10,024 individuals with SARS-CoV-2 infections recorded at least 2 weeks after a  
218 first dose of vaccine against COVID-19 were identified (mean [SD] age at infection: 57.0  
219 [17.9] years-old, 59.4% female). Among them, 65.1% were vaccinated with BNT162b2  
220 ‘Pfizer/BioNTech’, 9.0% with mRNA-1273 ‘Moderna’, 1.6% with Ad26.COV2.S ‘Janssen’,  
221 and 24.4% with unspecified subtype. 9479 of these individuals were matched to 9479  
222 individuals with SARS-CoV-2 infections who did not have a COVID-19 vaccine before  
223 SARS-CoV-2 infection. The main demographic features and comorbidities of both cohorts  
224 are summarised in table 1 (additional baseline characteristics presented in the appendix pp.  
225 18-19). Adequate propensity-score matching (standardised mean difference < 0.1) was  
226 achieved for all comparisons and baseline characteristics and all subgroups (appendix pp. 18-  
227 27).

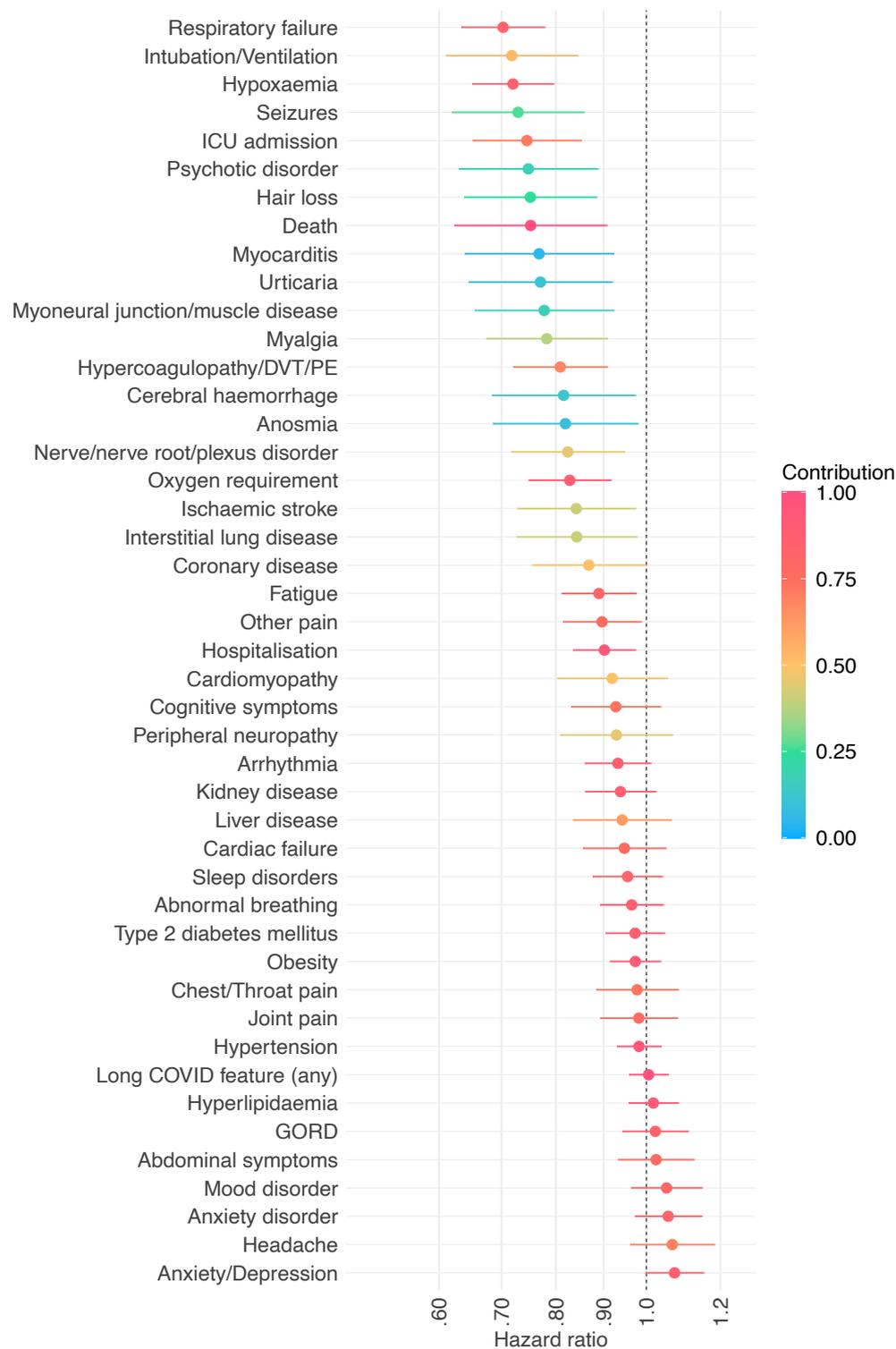
228  
229 We estimated the HRs for the occurrence within 6 months of infection of a range of health  
230 events (combined with death in composite endpoints) previously documented to occur at an  
231 increased rate after SARS-CoV-2 infections. As seen in Fig. 1 and Fig. 2 (see Appendix pp.  
232 9-12 for all Kaplan-Meier curves and pp. 28-29 for a summary table), compared to  
233 unvaccinated individuals, those who were vaccinated at the time of SARS-CoV-2 infection  
234 were at a significantly lower risk of composite outcome of death and respiratory failure (HR  
235 0.70, 95% CI 0.63-0.78, Bonferroni-corrected  $p < 0.0001$ ), intubation/ventilation (HR 0.72,  
236 95% CI 0.61-0.84, Bonferroni-corrected  $p = 0.0024$ ), hypoxaemia (HR 0.72, 95% CI 0.65-  
237 0.80, Bonferroni-corrected  $p < 0.0001$ ), seizures (HR 0.73, 95% CI 0.62-0.86, Bonferroni-  
238 corrected  $p = 0.0057$ ), ICU admission (HR 0.75, 95% CI 0.65-0.85, Bonferroni-corrected  
239  $p < 0.0001$ ), psychotic disorder (HR 0.75, 95% CI 0.63-0.89, Bonferroni-corrected  $p = 0.036$ ),  
240 hair loss (HR 0.75, 95% CI 0.64-0.88, Bonferroni-corrected  $p = 0.024$ ), hypercoagulopathy or

241 venous thromboembolism (HR 0.81, 95% CI 0.72-0.91, Bonferroni-corrected  $p=0.014$ ), and  
242 oxygen requirement (HR 0.83, 95% CI 0.75-0.92, Bonferroni-corrected  $p=0.011$ ). In contrast,  
243 there was no significant difference in the risk of many other outcomes including composite of  
244 death and any long-COVID feature (HR 1.01, 95% CI 0.96-1.05,  $p=0.83$ , Bonferroni-  
245 corrected  $p=1.0$ ), Type 2 diabetes mellitus (HR 0.97, 95% CI 0.90-1.05,  $p=0.45$ , Bonferroni-  
246 corrected  $p=1.0$ ), mood disorder (HR 1.05, 95% CI 0.96-1.15,  $p=0.27$ , Bonferroni-corrected  
247  $p=1.0$ ), and anxiety disorder (HR 1.06, 95% CI 0.97-1.15,  $p=0.20$ , Bonferroni-corrected  
248  $p=1.0$ ), among others. There was no violation of the proportionality assumption for 35 out of  
249 45 outcomes (Appendix pp. 30-31 and p. 13 for the time-varying HR for the other 10  
250 outcomes). For most of the other 10 outcomes (including death), the time-varying HRs were  
251 significantly lower than 1 (indicating a stronger association with vaccination status) earlier in  
252 the follow-up window.

253  
254 Besides the outcomes shown to be significantly associated with vaccination in the primary  
255 analysis, those who had received two vaccine doses at the time of SARS-CoV-2 infection  
256 were also at a significantly lower risk of myalgia, myocarditis, cerebral haemorrhage,  
257 interstitial lung disease, urticaria, anosmia, and myoneural junction/muscle disease (each as  
258 composite endpoints with death), and at a significantly lower risk of death (Fig. 3 and  
259 appendix pp. 14, 32-33). In those who had received only one vaccine dose, HRs followed a  
260 similar pattern, though they were generally closer to 1 (Fig. 3 and appendix pp. 15, 34-35).  
261 For each of the outcome, the HR in those who had received 2 doses was not statistically  
262 significantly different from the HR in those who had received 1 vaccine dose (appendix pp.  
263 36-37).

264

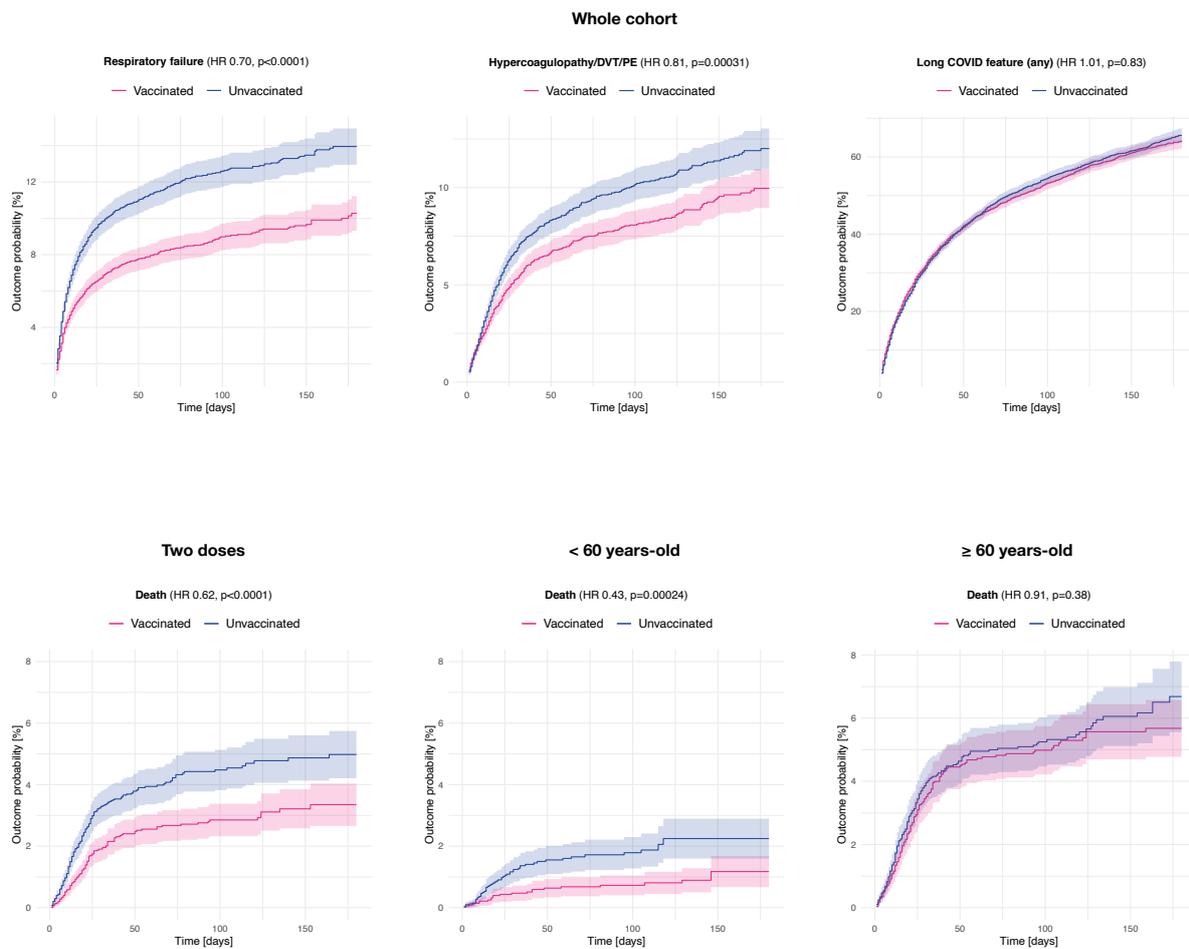
265 We found a substantial effect of age on the results. Many HRs in younger individuals (< 60  
266 years-old) were in general lower (i.e. favouring vaccination even more) for outcomes  
267 significantly associated with vaccination (Fig. 3 and appendix pp. 16, 38-39). These lower  
268 HRs were observed against the backdrop of lower *absolute* risks for most outcomes among  
269 younger individuals (appendix pp. 38-39). In contrast, in individuals  $\geq 60$  years-old, although  
270 most trends were similar to those in the younger group, none of the HRs were statistically  
271 significantly different from 1 after Bonferroni correction (Fig. 3 and appendix pp. 17, 40-41).



272

273 **Fig. 1** – Hazard ratios for the outcome within 6 months of infection with SARS-CoV-2 between individuals  
 274 vaccinated vs. unvaccinated against COVID-19. HR lower than 1 indicate outcomes less common among  
 275 vaccinated individuals. Horizontal bars represent 95% confidence intervals. Each outcome is a composite  
 276 endpoint with death as a component to address competing risks. The contribution of the outcome of interest to  
 277 the overall incidence of the composite endpoint is encoded by the colour.

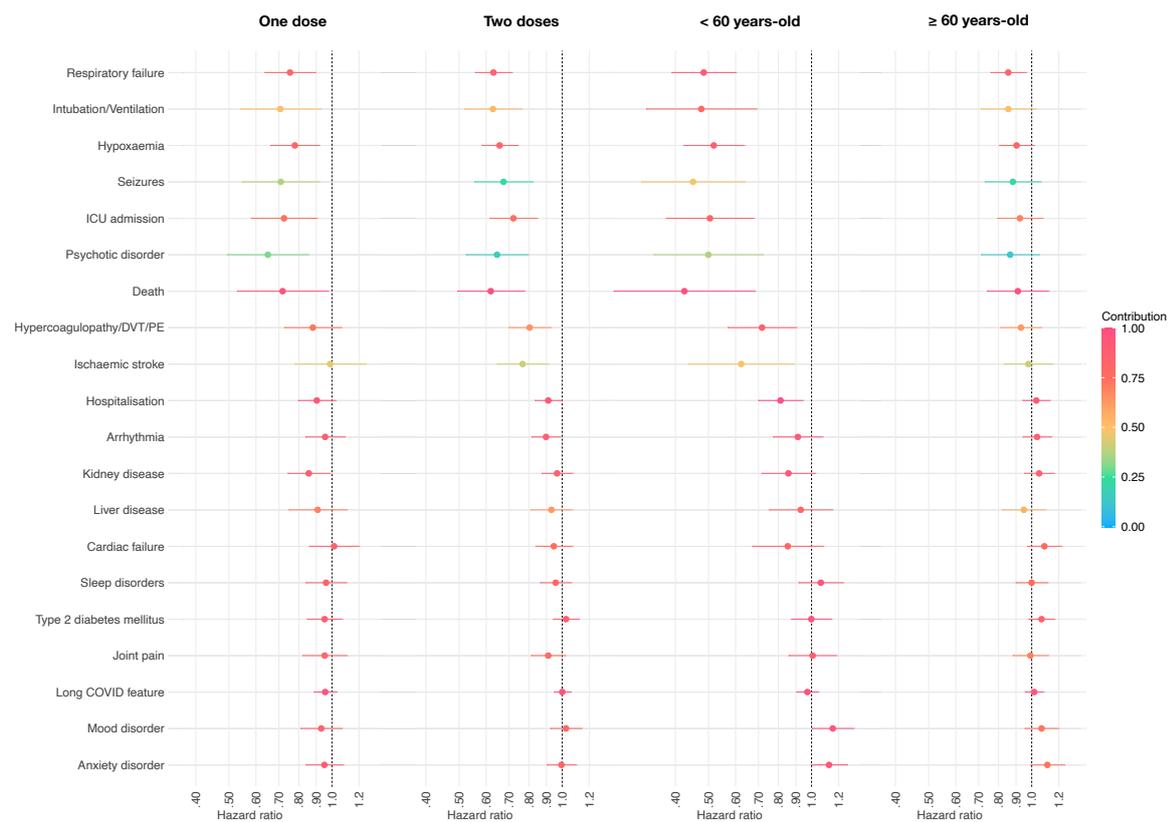
278



279

280 **Fig. 2** – Kaplan-Meier estimates for the incidence of outcomes of SARS-CoV-2 infection between vaccinated  
281 and unvaccinated individuals in the whole cohort (top) and for the incidence of death in different subgroups  
282 (bottom). 95% confidence intervals are shaded. For Kaplan-Meier curves of all other outcomes in the whole  
283 cohort, see appendix pp. 9-12. DVT=Deep vein thrombosis. PE=Pulmonary embolism.

284



285

286 **Fig. 3** - Hazard ratios for the outcomes within 6 months of infection with SARS-CoV-2 between individuals  
 287 who received one dose of the vaccine (vs. unvaccinated individuals), those who received two doses of the  
 288 vaccine (vs. unvaccinated individuals), vaccinated vs. unvaccinated individuals under the age of 60, and  
 289 vaccinated vs. unvaccinated individuals over the age of 60. HR lower than 1 indicate outcomes less common  
 290 among vaccinated individuals. Horizontal bars represent 95% confidence intervals. Each outcome is a  
 291 composite endpoint with death as a component to address competing risks. The contribution of the outcome of  
 292 interest to the overall incidence of the composite endpoint is encoded by the colour. Only a subset of  
 293 representative outcomes is displayed. The same figures with all outcomes are presented in the appendix pp. 14-  
 294 17.

295

296 For all but one outcomes associated with vaccination status in the primary analysis, the HR in  
 297 the older group was significantly higher (i.e. favouring vaccination less) than the HR in the  
 298 younger group (Appendix pp. 42-43). For instance, the HR for the composite outcome of  
 299 death or respiratory failure was 0.48 (95% CI 0.39-0.60) in individuals <60 years-old and  
 300 0.85 (95% CI 0.76-0.96) in those ≥60 years-old (ratio between the two: 1.77, 95% CI 1.38-  
 301 2.26,  $p < 0.0001$ ). The only exception was the composite of death and  
 302 hypercoagulopathy/DVT/PE (HR in the older group 0.93, HR in the younger group 0.72,  
 303 ratio between the two 1.30, 95% CI 1.00-1.70,  $p = 0.054$ ).

## 304 **Discussion**

305 It is established that vaccination protects against hospitalisation, ICU admission, and death  
306 from COVID-19 (Agrawal et al., 2021; Antonelli et al., 2021; Bahl et al., 2021; Butt et al.,  
307 2021; Cabezas et al., 2021; Glatman-Freedman et al., 2021; Haas et al., 2021; Hyams et al.,  
308 2021; Mateo-Urdiales et al., 2021; Roest et al., 2021) but little was known about other  
309 outcomes of breakthrough SARS-CoV-2 infections. The data presented in this study, from a  
310 large-scale electronic health records network, confirm that vaccination protects against death  
311 and ICU admission following breakthrough SARS-CoV-2 infection and provide estimates of  
312 HR for these outcomes in a general population. Our study also shows that vaccination against  
313 COVID-19 is associated with lower risk of additional outcomes that had not been assessed in  
314 previous studies, namely respiratory failure, hypoxaemia, oxygen requirement,  
315 hypercoagulopathy or venous thromboembolism, seizures, psychotic disorder, and hair loss.

316  
317 On the other hand, previous vaccination does not appear to be protective against several  
318 previously documented outcomes of COVID-19 such as long-COVID features, arrhythmia,  
319 joint pain, type 2 diabetes, liver disease, sleep disorders, and mood and anxiety disorders. The  
320 narrow confidence intervals (related to the high incidence of these outcomes post-COVID)  
321 rules out the possibility that these negative findings are merely a result of lack of statistical  
322 power. The inclusion of death in a composite endpoint with these outcomes rules out  
323 survivorship bias as an explanation. Instead, these negative findings might indicate that these  
324 outcomes arise through different pathophysiological mechanisms than outcomes which are  
325 affected by prior vaccination. For example, for anxiety and depression, it might be that  
326 antagonistic forces are at play with the protective effect of the vaccine being counteracted by  
327 the additional stressor of being infected despite being vaccinated.

328

329 The absence of a protective effect against long-COVID features is concerning given the high  
330 incidence and burden of these sequelae of COVID-19 (Taquet et al., 2021a). However, the  
331 risk of several individual long-COVID features were significantly associated with prior  
332 vaccination (but did not survive correction for multiple comparisons): myalgia (HR 0.78,  
333 95% CI 0.67-0.91), fatigue (HR 0.89, 95% CI 0.81-0.97), and pain (HR 0.90, 95% CI 0.81-  
334 0.99), with potentially additional protection after a second dose of the vaccine against  
335 abnormal breathing (HR 0.89, 95% CI 0.81-0.98) and cognitive symptoms (HR 0.87, 95% CI  
336 0.76-0.99). Relative differences in the incidence of individual long-COVID features might  
337 explain why our findings differ from those of an app-based survey suggesting that  
338 vaccination is overall protective against long-COVID symptoms (Antonelli et al., 2021).  
339 Other reasons might also explain this difference. First, the present study used data routinely  
340 collected from a general population rather than self-selected individuals. Second, this study  
341 uses a large sample size and matches cohorts on a broad range of recorded comorbidities. To  
342 further decrease selection bias, the control cohort of this study was selected among those  
343 having received an influenza vaccine thus helping control for the confounding effect of  
344 vaccine hesitancy and related health behaviours (Latkin et al., 2021). Third, by using  
345 symptoms and diagnoses extracted from health records rather than self-reported, our study  
346 might capture the most severe presentation of symptoms and these might be less prone to  
347 demand characteristics and other forms of detection bias. Finally, there might be differences  
348 in the type of vaccines used between the two study. In particular, no ChAdOx1 nCov-19  
349 ('Oxford/AstraZeneca') vaccine was used in our study (as this vaccine is not used in the  
350 USA).

351

352 The findings that vaccination against SARS-CoV-2 does not protect against some of the post-  
353 acute outcomes of COVID-19 should not obscure the fact that vaccination remains an

354 important protective factor against these outcomes at the population level, since the best way  
355 to prevent those outcomes is to prevent SARS-CoV-2 infection in the first place. This is also  
356 the case in those  $\geq 60$  years-old (Polack et al., 2020). However, our results highlight that  
357 some post-acute outcomes of SARS-CoV-2 (and notably long-COVID presentations) are  
358 likely to persist even after successful vaccination of the population, so long as breakthrough  
359 infections occur. These findings thus help in determining the necessary service provision.  
360 They also underline the urgency to identify other preventive or curative interventions to  
361 mitigate the impact of such COVID-19 sequelae.

362

363 Death was included in a composite endpoint with each outcome of interest to address  
364 competing risks (Manja et al., 2017). This implies that for outcomes much rarer than death  
365 (which is the minority of outcomes investigated), HRs might be driven by death rather than  
366 the outcome of interest. For instance, the 6-month incidence of psychotic disorders in the  
367 unvaccinated cohort was 1.11% whereas the death rate was 4.5% in this cohort. Hence the  
368 HR of 0.75 for the composite endpoint of death and psychotic disorder is in part driven by  
369 differences in death rate. Observing the HR for the outcome of interest in isolation (i.e. not  
370 composed with death) is informative so long as it is lower than 1. A HR lower than 1  
371 suggests a lower rate of the outcome of interest despite survivorship bias (which in the  
372 present study increases HRs). If the HR is higher than 1, it is impossible to know whether the  
373 outcome is more common among vaccinated individuals or if it is a result of survivorship  
374 bias. HRs for all outcomes in isolation are provided in Appendix pp. 44-53 for the primary  
375 and secondary analyses. For instance, the HR for psychotic disorder measured in isolation  
376 was 0.79, indicating that the incidence of this outcome might indeed be lower in vaccinated  
377 than unvaccinated individuals and its significant HR might not be merely driven by  
378 differences in death rates.

379

380 Importantly, the protective effects of the vaccine against acute severity of infection and some  
381 post-acute sequelae appears to affect primarily those <60 years-old. In this group, the effects  
382 were large and robust, whereas in the  $\geq 60$  year-old group, effects were smaller and not  
383 statistically robust. This finding regarding age cannot be explained by differences in  
384 statistical power since the younger and older subgroups have similar sample sizes (4633 and  
385 4657 respectively) and outcomes are in general more frequent in the older subgroup. Instead,  
386 it might be due to differences in contributions of pathophysiological pathways of  
387 breakthrough infections between younger and older individuals. In younger patients, effective  
388 B-cell response to vaccination might be followed by infection with variants against which  
389 antibodies have less neutralising activity (Wu et al., 2021). In older patients, the B-cell  
390 response to vaccination might itself be ineffective (Siegrist and Aspinall, 2009) so that the  
391 clinical presentation of breakthrough infection is similar to that of infection in unvaccinated  
392 individuals. Regardless of explanation, the seemingly age-limited benefits of vaccination on  
393 the severity and sequelae of SARS-CoV-2 infection has clear public health implications.

394

395 This study has several limitations beyond those inherent to research using EHRs (Casey et  
396 al., 2016; Taquet et al., 2021b) (summarised in the appendix p. 2) such as the unknown  
397 completeness of records, no validation of diagnoses, and sparse information on  
398 socioeconomic and lifestyle factors. First, we do not know which SARS-CoV-2 variant  
399 individual patients were infected with and this might affect the protective effect of vaccines.  
400 There is evidence that variants of concerns are overrepresented in breakthrough infections  
401 (McEwen et al., 2021). Since such variants tend to be associated with worse outcomes  
402 (Nyberg et al., 2021), their enrichment in breakthrough infections means that some HRs  
403 presented in this study might be conservative estimates. Second, it might be that vaccination

404 status affects the probability to seek or receive medical attention, particularly for less severe  
405 outcomes. Third, this study says nothing about the outcomes in patients infected with SARS-  
406 CoV-2 but who did not get tested nor diagnosed with COVID-19. Fourth, this study did not  
407 investigate whether the association between vaccination status and outcomes of SARS-CoV-  
408 2 infection was moderated by timing of the vaccine with respect to the infection. Fifth, we  
409 could not compare the different vaccines against each other since the majority had received  
410 the Pfizer/BioNTech vaccine. Finally, as an observational study, causation cannot be inferred  
411 (though the specificity of the association with some outcomes and not others and the dose-  
412 response relationship support causality as an explanation).

413

414 In summary, the present data show that prior vaccination against COVID-19, especially after  
415 two doses, is associated with significantly less risk of many but not all outcomes of COVID-  
416 19, in younger but not older individuals. These findings may inform service planning,  
417 contribute to forecasting public health impacts of vaccination programmes, and highlight the  
418 urgent need to identify or develop additional preventive and curative interventions for  
419 sequelae of COVID-19.

## 420 **Role of Funding Source**

421 The funding source had no role in study design; in the collection, analysis, and interpretation  
422 of data; in the writing of the report; and in the decision to submit the paper for publication.

## 423 **Contributors**

424 MT and PJH conceptualised the study, acquired funding, defined the methodology, and were  
425 involved in project administration and supervision. MT and QD curated the data, did the  
426 formal analyses, and validated the findings. MT created visualizations and wrote the original

427 draft. MT, QD, and PJH reviewed and edited the manuscript for style and content. PJH and  
428 MT were granted unrestricted access to the TriNetX Analytics network for the purposes of  
429 research, and with no constraints on the analyses performed nor the decision to publish. The  
430 corresponding author had full access to all the data in the study, conducted the analysis, and  
431 takes responsibility for the integrity of the data and the accuracy of the data analysis.

### 432 **Acknowledgments**

433 Work supported by the National Institute for Health Research (NIHR) Oxford Health  
434 Biomedical Research Centre (grant BRC-1215-20005). MT is an NIHR Academic Clinical  
435 Fellow and NIHR Oxford Health BRC Senior Research Fellow. QD is supported by the  
436 Medical Research Council (SUAG/077 G101400) and an AXA Research Fund Fellowship  
437 (G102329). The views expressed are those of the authors and not necessarily those of the UK  
438 National Health Service, NIHR, or the UK Department of Health.

### 439 **Declarations of interest**

440 The authors declare no conflict of interest.

### 441 **Data sharing**

442 The TriNetX system returned the results of these analyses as csv files which were  
443 downloaded and archived. Data presented in this paper and the Appendix can be freely  
444 accessed at [URL to be added on publication]. Additionally, TriNetX will grant access to  
445 researchers if they have a specific concern (via the third-party agreement option).

## References

- Agrawal, U., Katikireddi, S.V., McCowan, C., Mulholland, R.H., Azcoaga-Lorenzo, A., Amele, S., Fagbamigbe, A.F., Vasileiou, E., Grange, Z., Shi, T., Kerr, S., Moore, E., Murray, J.L.K., Shah, S.A., Ritchie, L., O'Reilly, D., Stock, S.J., Beggs, J., Chuter, A., Torabi, F., Akbari, A., Bedston, S., McMenamin, J., Wood, R., Tang, R.S.M., de Lusignan, S., Hobbs, F.D.R., Woolhouse, M., Simpson, C.R., Robertson, C., Sheikh, A., 2021. COVID-19 hospital admissions and deaths after BNT162b2 and ChAdOx1 nCoV-19 vaccinations in 2·57 million people in Scotland (EAVE II): a prospective cohort study. *Lancet Respir. Med.* [https://doi.org/10.1016/S2213-2600\(21\)00380-5](https://doi.org/10.1016/S2213-2600(21)00380-5)
- Al-Aly, Z., Xie, Y., Bowe, B., 2021. High-dimensional characterization of post-acute sequelae of COVID-19. *Nature.* <https://doi.org/10.1038/s41586-021-03553-9>
- Antonelli, M., Penfold, R.S., Merino, J., Sudre, C.H., Molteni, E., Berry, S., Canas, L.S., Graham, M.S., Klaser, K., Modat, M., Murray, B., Kerfoot, E., Chen, L., Deng, J., Österdahl, M.F., Cheetham, N.J., Drew, D.A., Nguyen, L.H., Pujol, J.C., Hu, C., Selvachandran, S., Polidori, L., May, A., Wolf, J., Chan, A.T., Hammers, A., Duncan, E.L., Spector, T.D., Ourselin, S., Steves, C.J., 2021. 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.* [https://doi.org/10.1016/S1473-3099\(21\)00460-6](https://doi.org/10.1016/S1473-3099(21)00460-6)
- Austin, P.C., 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav. Res.* 46, 399–424.
- Bahl, A., Johnson, S., Maine, G., Garcia, M.H., Nimmagadda, S., Qu, L., Chen, N.-W., 2021. Vaccination reduces need for emergency care in breakthrough COVID-19 infections: A multicenter cohort study. *Lancet Reg Health Am* 100065.
- Butt, A.A., Nafady-Hego, H., Chemaitelly, H., Abou-Samra, A.-B., Khal, A.A., Coyle, P.V., Kanaani, Z.A., Kaleeckal, A.H., Latif, A.N., Masalmani, Y.A., Bertollini, R., Raddad, L.J.A., 2021. Outcomes among patients with breakthrough SARS-CoV-2 infection after vaccination. *Int. J. Infect. Dis.* 110, 353–358.
- Cabezas, C., Coma, E., Mora-Fernandez, N., Li, X., Martinez-Marcos, M., Fina, F., Fabregas, M., Hermosilla, E., Jover, A., Contel, J.C., Lejardi, Y., Enfedaque, B., Argimon, J.M., Medina-Peralta, M., Prieto-Alhambra, D., 2021. Associations of BNT162b2 vaccination with SARS-CoV-2 infection and hospital admission and death with covid-19 in nursing homes and healthcare workers in Catalonia: prospective cohort study. *BMJ* 374, n1868.
- Casey, J.A., Schwartz, B.S., Stewart, W.F., Adler, N.E., 2016. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annu. Rev. Public Health* 37, 61–81.

- Daugherty, S.E., Guo, Y., Heath, K., Dasmariñas, M.C., Jubilo, K.G., Samranvedhya, J., Lipsitch, M., Cohen, K., 2021. Risk of clinical sequelae after the acute phase of SARS-CoV-2 infection: retrospective cohort study. *BMJ* 373, n1098.
- de Lusignan, S., Dorward, J., Correa, A., Jones, N., Akinyemi, O., Amirthalingam, G., Andrews, N., Byford, R., Dabrera, G., Elliot, A., Ellis, J., Ferreira, F., Lopez Bernal, J., Okusi, C., Ramsay, M., Sherlock, J., Smith, G., Williams, J., Howsam, G., Zambon, M., Joy, M., Hobbs, F.D.R., 2020. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30371-6](https://doi.org/10.1016/S1473-3099(20)30371-6)
- Glatman-Freedman, A., Bromberg, M., Dichtiar, R., Hershkovitz, Y., Keinan-Boker, L., 2021. The BNT162b2 vaccine effectiveness against new COVID-19 cases and complications of breakthrough cases: A nation-wide retrospective longitudinal multiple cohort analysis using individualised data. *EBioMedicine* 72, 103574.
- Haas, E.J., Angulo, F.J., McLaughlin, J.M., Anis, E., Singer, S.R., Khan, F., Brooks, N., Smaja, M., Mircus, G., Pan, K., Southern, J., Swerdlow, D.L., Jodar, L., Levy, Y., Alroy-Preis, S., 2021. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* 397, 1819–1829.
- Haukoos, J.S., Lewis, R.J., 2015. The propensity score. *JAMA* 314, 1637–1638.
- Hyams, C., Marlow, R., Maseko, Z., King, J., Ward, L., Fox, K., Heath, R., Tuner, A., Friedrich, Z., Morrison, L., Ruffino, G., Antico, R., Adegbite, D., Szasz-Benczur, Z., Garcia Gonzalez, M., Oliver, J., Danon, L., Finn, A., 2021. Effectiveness of BNT162b2 and ChAdOx1 nCoV-19 COVID-19 vaccination at preventing hospitalisations in people aged at least 80 years: a test-negative, case-control study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(21\)00330-3](https://doi.org/10.1016/S1473-3099(21)00330-3)
- Latkin, C.A., Dayton, L., Yi, G., Colon, B., Kong, X., 2021. Mask usage, social distancing, racial, and gender correlates of COVID-19 vaccine intentions among adults in the US. *PLoS One* 16, e0246970.
- Manja, V., AlBashir, S., Guyatt, G., 2017. Criteria for use of composite end points for competing risks—a systematic survey of the literature with recommendations. *J. Clin. Epidemiol.* 82, 4–11.
- Mateo-Urdiales, A., Spila Alegiani, S., Fabiani, M., Pezzotti, P., Filia, A., Massari, M., Riccardo, F., Tallon, M., Proietti, V., Del Manso, M., Puopolo, M., Spuri, M., Morciano, C., D’Ancona, F.P., Da Cas, R., Battilomo, S., Bella, A., Menniti-Ippolito, F., Italian Integrated Surveillance of COVID-19 study group, on behalf of the Italian COVID-19 vaccines registry, 2021. Risk of SARS-CoV-2 infection and subsequent hospital admission and death at different time intervals since first dose of COVID-19 vaccine administration, Italy, 27 December 2020 to mid-April 2021. *Euro Surveill.* 26. <https://doi.org/10.2807/1560-7917.ES.2021.26.25.2100507>

- McEwen, A.E., Cohen, S., Bryson-Cahn, C., Liu, C., Pergam, S.A., Lynch, J., Schippers, A., Strand, K., Whimbey, E., Mani, N.S., Zelikoff, A.J., Makarewicz, V.A., Brown, E.R., Bakhsh, S.A.M., Baker, N.R., Castor, J., Livingston, R.J., Huang, M.-L., Jerome, K.R., Greninger, A.L., Roychoudhury, P., 2021. Variants of concern are overrepresented among post-vaccination breakthrough infections of SARS-CoV-2 in Washington State. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciab581>
- Nixon, D.F., Ndhlovu, L.C., 2021. Vaccine breakthrough infections with SARS-CoV-2 variants. *N. Engl. J. Med.*
- Nyberg, T., Twohig, K.A., Harris, R.J., Seaman, S.R., Flannagan, J., Allen, H., Charlett, A., De Angelis, D., Dabrera, G., Presanis, A.M., 2021. Risk of hospital admission for patients with SARS-CoV-2 variant B.1.1.7: cohort analysis. *BMJ* 373, n1412.
- Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., Bailey, R., Swanson, K.A., Roychoudhury, S., Koury, K., Li, P., Kalina, W.V., Cooper, D., Frenck, R.W., Jr, Hammitt, L.L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D.B., Mather, S., Dormitzer, P.R., Şahin, U., Jansen, K.U., Gruber, W.C., C4591001 Clinical Trial Group, 2020. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* 383, 2603–2615.
- Roest, S., Hoek, R.A.S., Manintveld, O.C., 2021. BNT162b2 mRNA covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.*
- Royston, P., Parmar, M.K.B., 2002. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine.* <https://doi.org/10.1002/sim.1203>
- Siegrist, C.-A., Aspinall, R., 2009. B-cell responses to vaccination at the extremes of age. *Nat. Rev. Immunol.* 9, 185–194.
- Taquet, M., Dercon, Q., Luciano, S., Geddes, J.R., Husain, M., Harrison, P.J., 2021a. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLoS Med.* 18, e1003773.
- Taquet, M., Geddes, J.R., Husain, M., Luciano, S., Harrison, P.J., 2021b. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* 8, 416–427.
- Taquet, M., Luciano, S., Geddes, J.R., Harrison, P.J., 2021c. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* 8, 130–140.
- Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H.I., MacKenna, B., Tomlinson, L., Douglas, I.J., Rentsch, C.T., Mathur, R., Wong, A.Y.S., Grieve,

R., Harrison, D., Forbes, H., Schultze, A., Croker, R., Parry, J., Hester, F., Harper, S., Perera, R., Evans, S.J.W., Smeeth, L., Goldacre, B., 2020. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature*.  
<https://doi.org/10.1038/s41586-020-2521-4>

Wu, K., Werner, A.P., Koch, M., Choi, A., Narayanan, E., Stewart-Jones, G.B.E., Colpitts, T., Bennett, H., Boyoglu-Barnum, S., Shi, W., Moliva, J.I., Sullivan, N.J., Graham, B.S., Carfi, A., Corbett, K.S., Seder, R.A., Edwards, D.K., 2021. Serum neutralizing activity elicited by mRNA-1273 vaccine. *N. Engl. J. Med.* 384, 1468–1470.

**Table 1** – Baseline characteristics and major outcomes for the vaccinated and unvaccinated cohorts, before and after propensity score matching. Only characteristics with an overall prevalence above 5% after matching are presented here; for additional baseline characteristics see appendix pp. 18-19

	<b>Vaccinated (unmatched)</b>	<b>Unvaccinated (unmatched)</b>	<b>Vaccinated (matched)</b>	<b>Unvaccinated (matched)</b>
<b>COHORT SIZE</b>	10024	83957	9479	9479
<b>DEMOGRAPHICS</b>				
<b>Age; mean (SD)</b>	57.0 (17.9)	51.9 (23.1)	56.5 (18.0)	57.6 (20.6)
<b>Sex; n (%) female</b>	5950 (59.4)	48968 (58.3)	5676 (59.9)	5761 (60.8)
<b>Race; n (%)</b>				
White	7208 (71.9)	57310 (68.3)	6783 (71.6)	6873 (72.5)
Black or African American	1584 (15.8)	16296 (19.4)	1540 (16.2)	1514 (16.0)
Unknown	802 (8.0)	7843 (9.3)	783 (8.3)	756 (8.0)
<b>Ethnicity; n (%)</b>				
Hispanic or Latino	1097 (10.9)	7539 (9.0)	1015 (10.7)	974 (10.3)
Not Hispanic of Latino	7199 (71.8)	63015 (75.1)	6780 (71.5)	6793 (71.7)
Unknown	1728 (17.2)	13403 (16.0)	1684 (17.8)	1712 (18.1)
<b>COMORBIDITIES</b>				
<b>BMI; mean (SD)</b>	30.0 (7.4)	29.7 (8.3)	30.1 (7.4)	30.2 (7.5)
<b>Overweight and obesity; n (%)</b>	3175 (31.7)	30363 (36.2)	3086 (32.6)	3263 (34.4)
<b>Blood pressure; mean (SD)</b>				
Diastolic	74.7 (11.5)	73.0 (12.1)	74.6 (11.5)	73.1 (11.6)
Systolic	127.6 (19.0)	124.2 (19.6)	127.5 (18.8)	126.0 (19.4)
<b>Hypertensive diseases; n (%)</b>	5227 (52.1)	46211 (55.0)	5060 (53.4)	5197 (54.8)
<b>Diabetes mellitus; n (%)</b>				
Type 2 diabetes mellitus	2508 (25.0)	24073 (28.7)	2426 (25.6)	2551 (26.9)
<b>Chronic lower respiratory diseases; n (%)</b>				
Asthma	1526 (15.2)	17665 (21.0)	1502 (15.8)	1491 (15.7)
Bronchitis, not specified as acute or chronic	887 (8.8)	10121 (12.1)	870 (9.2)	867 (9.1)
Other chronic obstructive pulmonary disease	884 (8.8)	10493 (12.5)	871 (9.2)	958 (10.1)
<b>Nicotine dependence; n (%)</b>	929 (9.3)	11509 (13.7)	917 (9.7)	935 (9.9)
<b>Psychiatric comorbidities; n (%)</b>				
Anxiety disorders	3187 (31.8)	31966 (38.1)	3098 (32.7)	3159 (33.3)
Substance misuse	1351 (13.5)	16299 (19.4)	1334 (14.1)	1373 (14.5)
Mood disorders	2453 (24.5)	27245 (32.5)	2408 (25.4)	2458 (25.9)
<b>Heart disease; n (%)</b>				
Ischemic heart diseases	2074 (20.7)	18880 (22.5)	2013 (21.2)	2114 (22.3)
Other forms of heart disease	3573 (35.6)	34359 (40.9)	3485 (36.8)	3607 (38.1)
<b>Chronic kidney diseases; n (%)</b>				
Chronic kidney disease (CKD)	1501 (15.0)	15137 (18.0)	1471 (15.5)	1561 (16.5)
Hypertensive chronic kidney disease	783 (7.8)	8795 (10.5)	761 (8.0)	812 (8.6)
<b>Chronic liver diseases; n (%)</b>				
Fatty (change of) liver, not elsewhere classified	727 (7.3)	6530 (7.8)	705 (7.4)	730 (7.7)
<b>Cerebral infarction; n (%)</b>	497 (5.0)	5384 (6.4)	487 (5.1)	508 (5.4)
<b>Neoplasms (any); n (%)</b>	4184 (41.7)	34046 (40.6)	4038 (42.6)	4208 (44.4)
<b>Disorders involving the immune mechanism; n (%)</b>	810 (8.1)	5844 (7.0)	778 (8.2)	771 (8.1)