nature research

Corresponding author(s):   Laith J Abu-Raddad

Last updated by author(s):   Jun 8, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were available to authors through .csv files
Data analysis	Analyses were conducted in STATA/SE 16.1. The commands/code are accessible using URL: https://github.com/IDEGWCMQ/Vaccine-effectiveness-code

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The dataset of this study is a property of the Qatar Ministry of Public Health that was provided to the researchers through a restricted-access agreement that prevents sharing the dataset with a third party or publicly. Access to this dataset at any time can be considered through a direct application for data access to her Excellency the Minister of Public Health (https://www.moph.gov.qa/english/Pages/default.aspx). Aggregate data are available within the manuscript and its Supplementary information.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	COVID-19 laboratory testing, vaccination, and clinical infection data were extracted from the national, federated SARS-CoV-2 databases that captured all SARS-CoV-2-related data since epidemic onset. The data is based on a national cohort that includes every single individual tested using PCR in Qatar. Sample size varied depending on the definition used for Cases (PCR-positive infected with B.1.1.7 variant, or infected with B.1.351 variant, or developing a severe, critical or fatal infection) and controls (PCR-negative) were matched one-to-one by sex, age, nationality, and reason for SARS-CoV-2 polymerase chain reaction (PCR) testing. Only matched pairs were included in the analysis. Given that the sample sizes were based on national cohorts with only individuals that do not fit the eligibility criteria excluded, the sample size for each sub-study can be considered sufficient. Detailed sample sizes can be found in Extended Data 1-4.
Data exclusions	Exclusion criteria were specified a priori. For the case control study, these included having a BNT162b2 vaccination record or being tested using PCR outside the study period. For the cohort study, these included having a BNT162b2 vaccination record, being tested using PCR outside the study period, not completing 14 days after the second vaccine dose before the start of the start of the study follow-up, or contracting the infection or dying before the start of the follow-up. Cases (PCR-positive infected with B.1.1.7 variant, or infected with B.1.351 variant, or developing a severe, critical or fatal infection) and controls (PCR-negative) were matched one-to-one by sex, age, nationality, and reason for SARS-CoV-2 PCR testing. Only matched pairs were included in the analysis and individuals not matched were excluded.
Replication	Sensitivity analyses matching by PCR testing date in addition to age, sex, nationality, and reason for PCR testing; adjusting for calendar week in logistic regression; or additionally adjusting for matching factors, that is sex, age, nationality, and reason for PCR testing, all confirmed results
Randomization	Cases and controls were matched one-to-one by sex, age, nationality, and reason for SARS-CoV-2 polymerase chain reaction (PCR) testing. Additionally, to ensure that vaccine effectiveness estimates were not biased by epidemic phase and the gradual roll-out of vaccination during the study, two sensitivity analyses were conducted, first matching by the exact PCR testing date and second by logistic regression to adjust for calendar week[16,30]. To further ensure control for confounding[31,32], a third sensitivity analysis was conducted adjusting additionally for the matching factors in logistic regression, that is sex, age, reason for PCR testing, and nationality.
Blinding	Not applicable as this is an observational study and those vaccinated are aware of their vaccination status

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The demographic characteristics of the different study populations can be found in Table 1 and Supplementary Table 2.
Recruitment	This is a retrospective study where COVID-19 laboratory testing, vaccination, clinical infection data, and related demographic details were extracted from the integrated nationwide digital-health information platform that hosts the national, federated SARS-CoV-2 databases. These databases are complete and have captured all SARS-CoV-2-related data since epidemic onset.

Cases and controls were defined based on analysis for these data. For reasons that remain unclear, among persons 7-13 days after the first dose, risk of infection with B.1.351 was higher compared to those who remained unvaccinated. This might reflect a higher underlying risk of infection, bias due to uncontrolled confounding such as differences in social behavior at or following vaccination, an immunological effect, or an artifact of the estimation method, possibly because the first vaccine dose coincided often with the peak of the B.1.351 wave.

Ethics oversight

The study was approved by the Hamad Medical Corporation and Weill Cornell Medicine-Qatar Institutional Review Boards.

Note that full information on the approval of the study protocol must also be provided in the manuscript.