Observational methods for COVID-19 vaccine effectiveness research: an empirical evaluation and target trial emulation

Martí Català Sabaté

Health Data Sciences group, University of Oxford

11th September 2024, CIBER BEST

Disclosure

- This study was funded by the National Institute for Health and Care Research (NIHR) (COV-LT2 0006)

Introduction

Martí Català Sabaté
Senior Data Scientist
Health Data Sciences division
University of Oxford

- BSc Physics Engineering – Universitat Politecnica de Catalunya (Barcelona)
- MSc Advanced Physics – Universitat de Barcelona (Barcelona)
- MSc Bioinformatics and Biostatistics – Universitat Oberta de Catalunya (Barcelona)
- PhD in Computational and Applied Physics – Universitat Politecnica de Catalunya (Barcelona)
  Mathematical modelling to study infectious diseases: from understanding to prediction

Senior data scientist at University of Oxford:

- Co-leading the Oxinfer group focused on:
  - Epidemiological studies
  - Epidemiological methods research
  - Development of standardised tools (DARWIN-EU, HDRKUK)

Outline

- Background
- Objective
- Methods research: *Observational methods for COVID-19 vaccine effectiveness research: an empirical evaluation and target trial emulation*
- Learnings
- Application
- Reproducibility
- Q&A

Background

- COVID-19 pandemic
- Massive vaccination campaign
- Very good trial results: 80-95% protective effect.
  - Which is its effectiveness in Real World Data?
  - How can we do it?
- How to design the study:
  - How do we select unvaccinated individuals? Problem with index date!
  - How to account for confounders: age, comorbidities associated with COVID-19 increased risk, …?
  - Cohort study
  - Which is the best method? PS matching, PS weighting, …

Objective

Evaluate effectiveness of 1st vaccine dose vs unvaccinated population:

1. To empirically evaluate the comparative performance of three different methods to minimize confounding in the study of COVID-19 vaccine effectiveness: **overlap weighting (OW), inverse probability of treatment weights (IPTW)** and propensity score (PS) with exact geographical and index date **matching**.
2. To conduct a **target trial emulation** study for the phase III randomized controlled trials which assessed effectiveness for the two different vaccine brands that were available first in England: BNT162b2 and ChAdOx1.

Population

CPRD AURUM Data base: primary care data of the UK with ~40M individuals. Linked COVID-19 testing and vaccination status.

Inclusion criteria:

- In observation on 4th January 2021
- >= 75 years old
- Prior observation > 180 days
- Location is present
- Two cohorts regarding if they were vaccinated 4th Jan - 28th Jan

Methods

- We want to address confounding between both cohorts:
  - Vaccinated are older

Characteristic	Unvaccinated (*n*=332 315)	Vaccinated					
		Any type (*n*=583 813)	SMD[d]	ChAdOx1[e] (*n*=235 538)	SMD[d]	BNT162b2[e] (*n*=348 275)	SMD[d]
Age (years), median [IQR]	78 [76–83]	81 [78–86]	0.285	80 [77–85]	0.208	82 [78–86]	0.341

  - More comorbidities:

Asthma	38 193 (11.5%)	70 385 (12.1%)	0.017	28 329 (12.0%)	0.017	42 056 (12.1%)	0.018
Autoimmune disease	13 920 (4.2%)	26 281 (4.5%)	0.015	10 893 (4.6%)	0.021	15 388 (4.4%)	0.011
COPD	28 627 (8.6%)	52 176 (8.9%)	0.011	21 619 (9.2%)	0.020	30 557 (8.8%)	0.006

-> Propensity scores (Probability to receive vaccine)

Propensity scores

Key confounders + data driven approach:

Data driven:

Covariates to be included in the large-scale PS:

- Conditions: 1–30 days, 31–180 days and 181 days to any time prior index date.
- Drugs: 1–30 days, 31–180 days prior index date.

Inclusion:

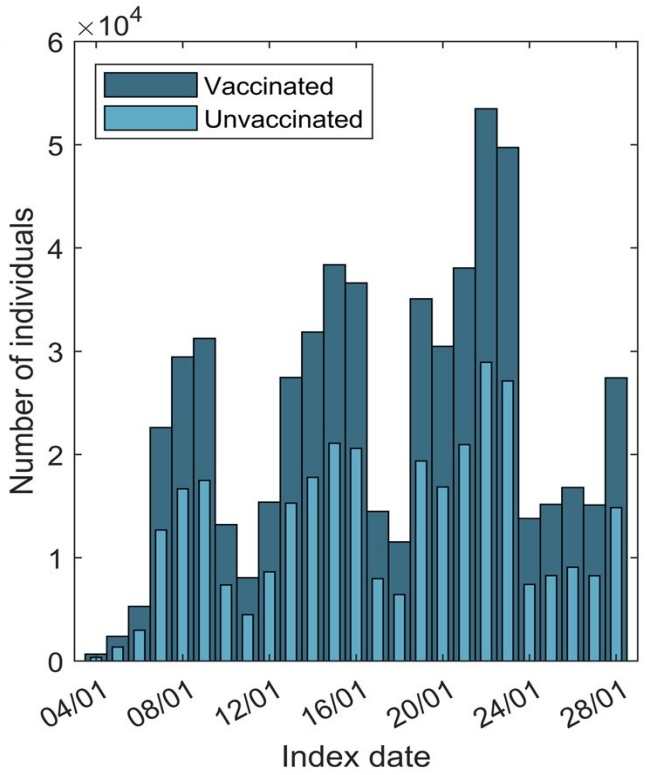
- At least a frequency >0.5% in the study population.
- Lasso regression for variable selection.
- Clinical review to exclude instrumental variables.

Key confounders:

location, age, prior observation, number of outpatient visits and number of previous COVID-19 PCR tests.

Study design - Weighting

- Index date of unvaccinated: random date following vaccinated distribution.
- Three sets of weights*:
  - No geographic identifier
  - Using regional identifier
  - Using GP (care site) identifier
- Two weighting methods:
  - Inverse Probability Treatment
  - Overlap weighting

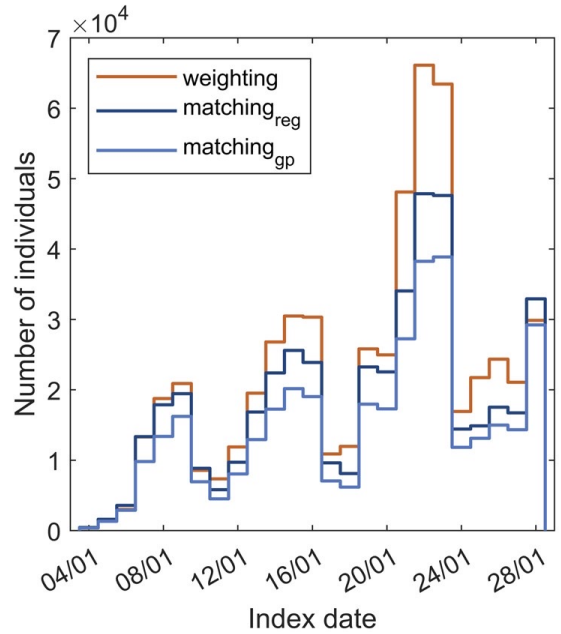


*concerns: roll-out and infection rates

Study design - Matching

- Matching :
  - Exact: age group (5-year bands), sex + geographic identifier
  - 0.2 caliper matching using nearest neighbors

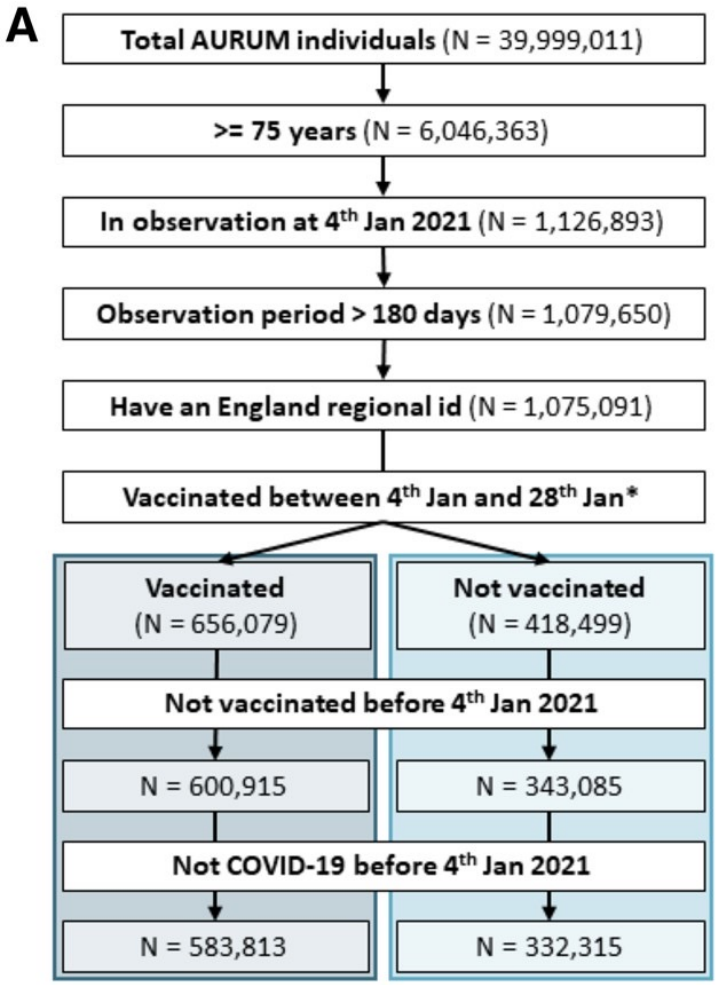
- Two sets of matching:
  - Regional geographic identifier
  - GP geographic identifier

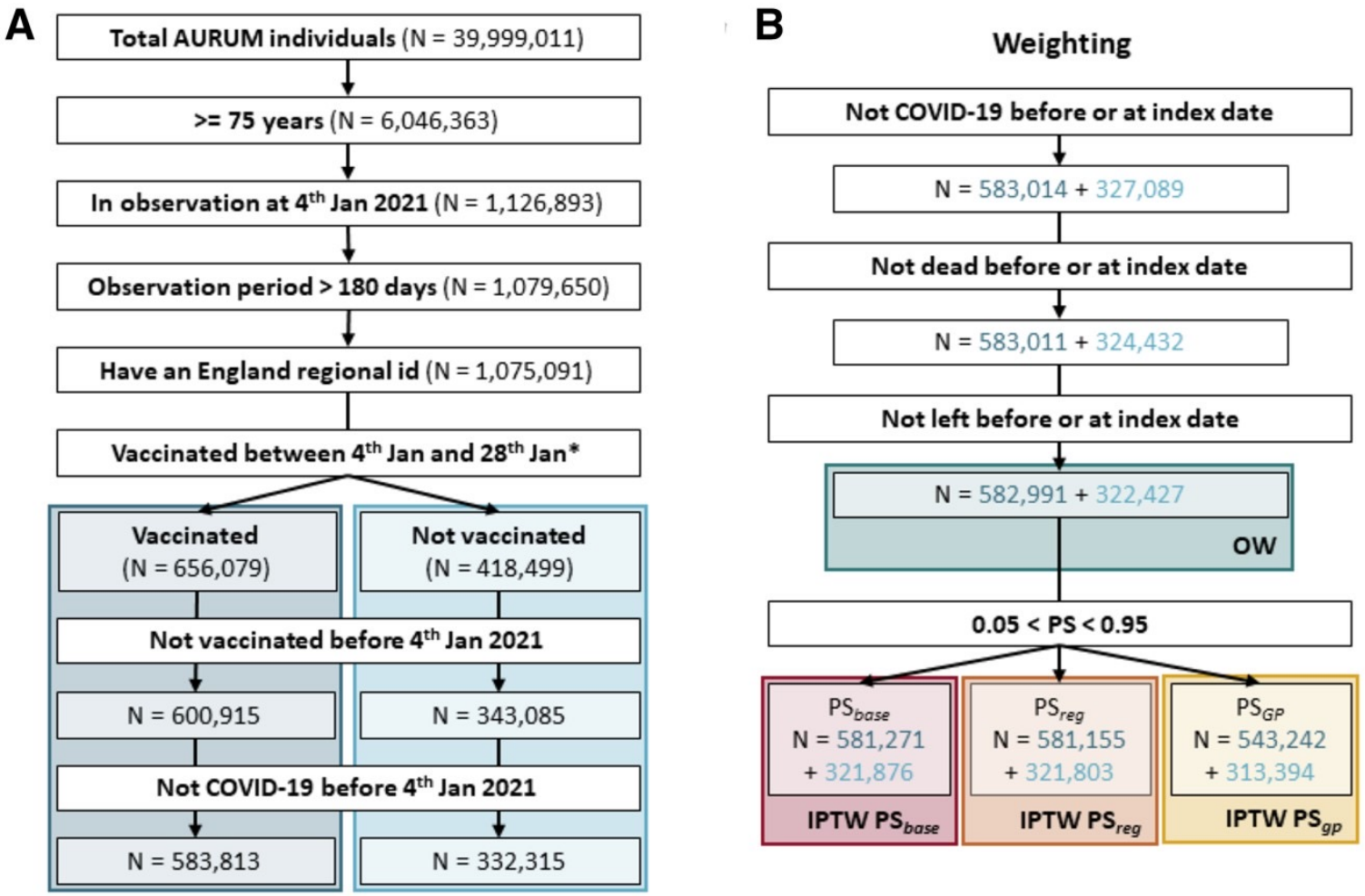


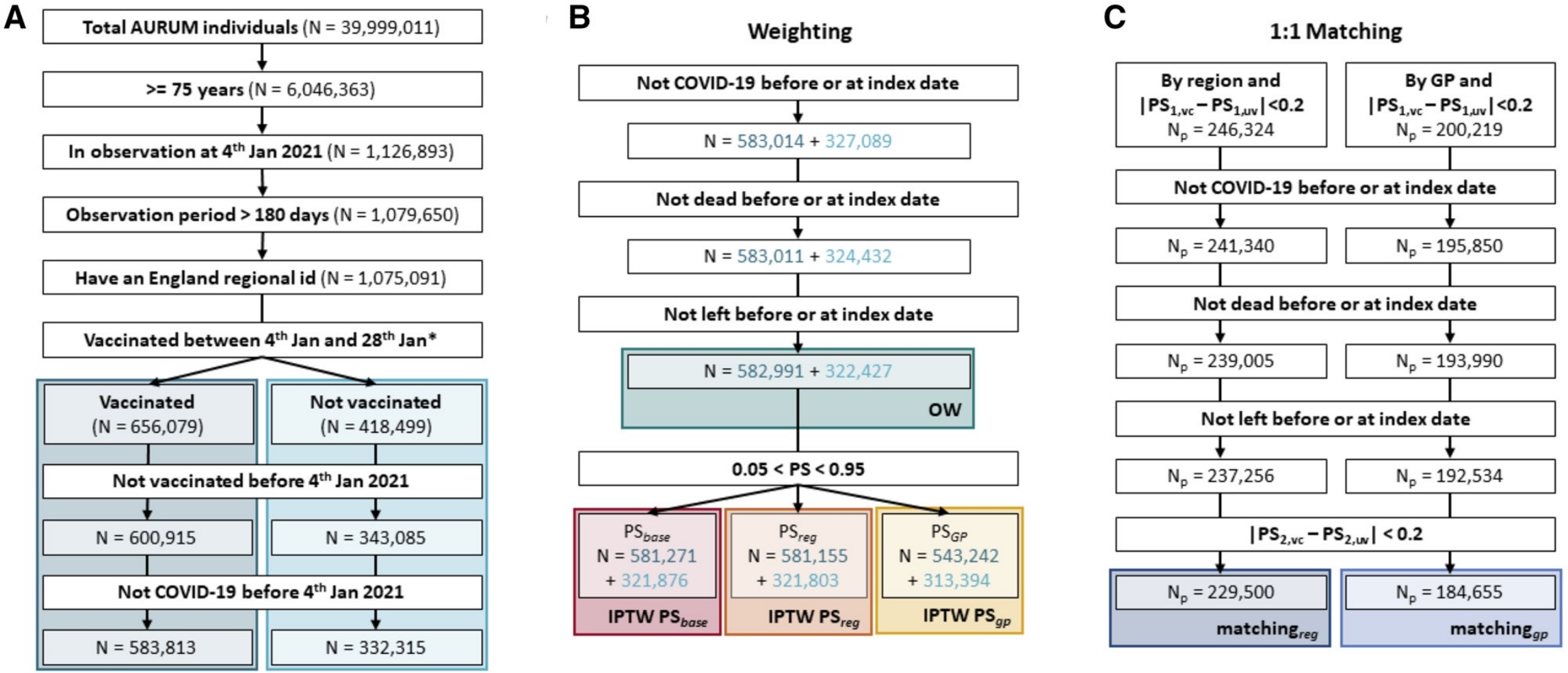
Outcome

- PCR testing
- PCR positive test
- PCR positive test or clinical COVID-19 diagnoses

Outcome model: Cox proportional hazard regression


A
Total AURUM individuals (N = 39,999,011)
>= 75 years (N = 6,046,363)
In observation at 4th Jan 2021 (N = 1,126,893)
Observation period > 180 days (N = 1,079,650)
Have an England regional id (N = 1,075,091)
Vaccinated between 4th Jan and 28th Jan*
Vaccinated
(N = 656,079)
Not vaccinated
(N = 418,499)
Not vaccinated before 4th Jan 2021
N = 600,915
N = 343,085
Not COVID-19 before 4th Jan 2021
N = 583,813
N = 332,315


A
Total AURUM individuals (N = 39,999,011)
>= 75 years (N = 6,046,363)
In observation at 4th Jan 2021 (N = 1,126,893)
Observation period > 180 days (N = 1,079,650)
Have an England regional id (N = 1,075,091)
Vaccinated between 4th Jan and 28th Jan*
Vaccinated
(N = 656,079)
Not vaccinated
(N = 418,499)
Not vaccinated before 4th Jan 2021
N = 600,915
N = 343,085
Not COVID-19 before 4th Jan 2021
N = 583,813
N = 332,315
B
Weighting
Not COVID-19 before or at index date
N = 583,014 + 327,089
Not dead before or at index date
N = 583,011 + 324,432
Not left before or at index date
N = 582,991 + 322,427
OW
0.05 < PS < 0.95
PSbase
N = 581,271
+ 321,876
IPTW PSbase
PSreg
N = 581,155
+ 321,803
IPTW PSreg
PSGP
N = 543,242
+ 313,394
IPTW PSgp


A
Total AURUM individuals (N = 39,999,011)
>= 75 years (N = 6,046,363)
In observation at 4th Jan 2021 (N = 1,126,893)
Observation period > 180 days (N = 1,079,650)
Have an England regional id (N = 1,075,091)
Vaccinated between 4th Jan and 28th Jan*
Vaccinated (N = 656,079)
Not vaccinated (N = 418,499)
Not vaccinated before 4th Jan 2021
N = 600,915
N = 343,085
Not COVID-19 before 4th Jan 2021
N = 583,813
N = 332,315
B
Weighting
Not COVID-19 before or at index date
N = 583,014 + 327,089
Not dead before or at index date
N = 583,011 + 324,432
Not left before or at index date
N = 582,991 + 322,427
OW
0.05 < PS < 0.95
PS_base N = 581,271 + 321,876
IPTW PS_base
PS_reg N = 581,155 + 321,803
IPTW PS_reg
PS_GP N = 543,242 + 313,394
IPTW PS_gp
C
1:1 Matching
By region and |PS1,vc − PS1,uv| <0.2 Np = 246,324
By GP and |PS1,vc − PS1,uv| <0.2 Np = 200,219
Not COVID-19 before or at index date
Np = 241,340
Np = 195,850
Not dead before or at index date
Np = 239,005
Np = 193,990
Not left before or at index date
Np = 237,256
Np = 192,534
|PS2,vc − PS2,uv| < 0.2
Np = 229,500
matching_reg
Np = 184,655
matching_gp

Follow-up

Individuals were followed till:

- End of data availability (~December 2021)
- Outcome of interest
- Vaccination (unvaccinated)
- Leave database
- Death
- Vaccination of the pair (vaccinated) [only matching]

Metrics

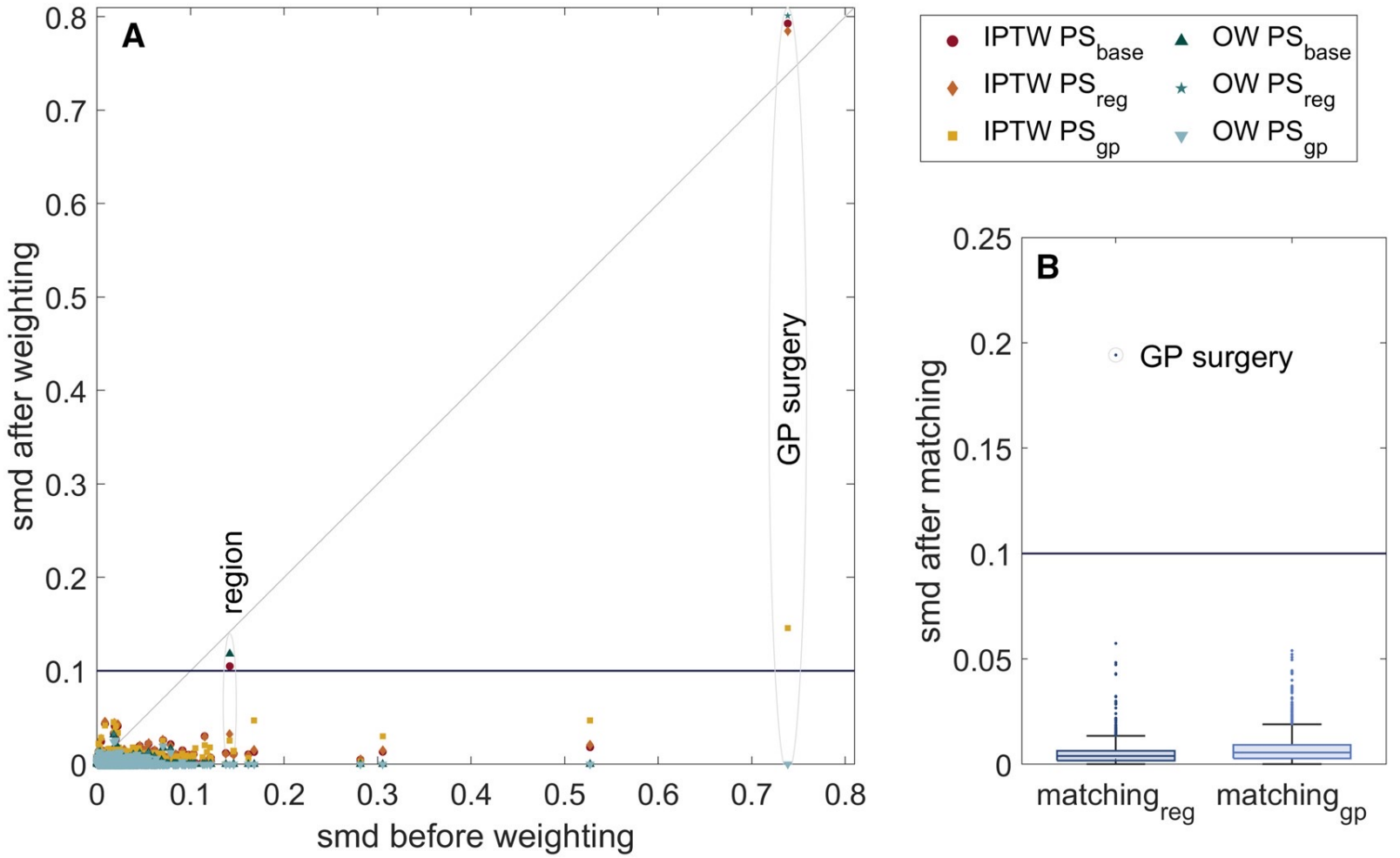
- Observed confounding -> Standardised mean differences
- Power -> Minimum detectable risk
- Unobserved confounding -> negative control outcomes

Metrics

Standardised mean differences

- Use 0.1 threshold to determine if a covariate was balanced or not.

Observed confounding



Metrics

Method[a]	Minimum detectable relative risk[b]
Any type of vaccinated comparison	
Unweighted	<0.93; >1.08
IPTW PS_{base}	<0.93; >1.08
IPTW PS_{reg}	<0.93; >1.08
IPTW PS_{gp}	<0.92; >1.08
OW PS_{base}	<0.93; >1.08
OW PS_{reg}	<0.93; >1.08
OW PS_{gp}	<0.93; >1.08
$Matching_{reg}$	<0.87; >1.15
$Matching_{gp}$	<0.86; >1.17

Metrics

- Unobserved confounding -> **Negative Control Outcomes**

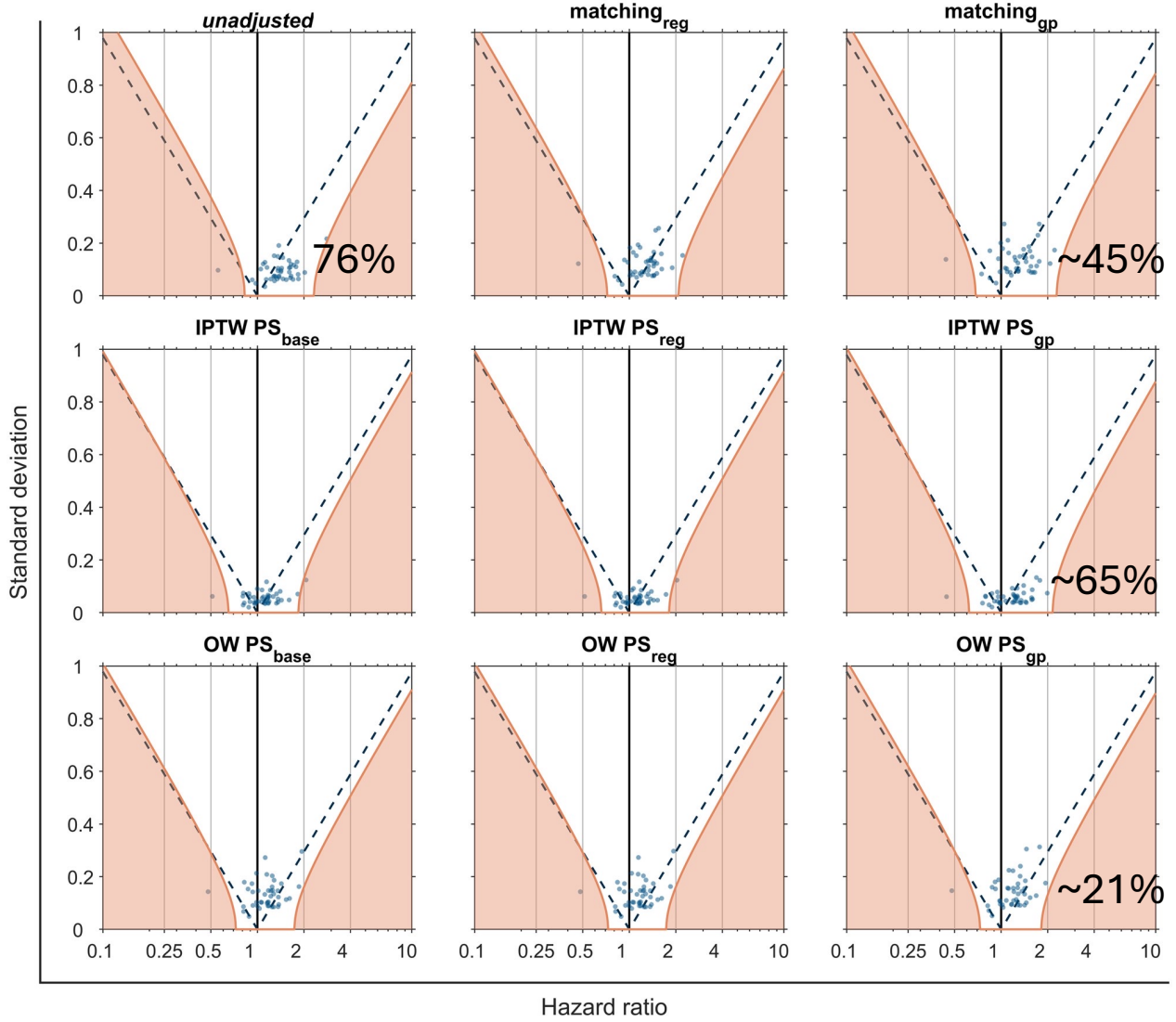
outcome that is not expected to be influenced by the exposure of interest but shares the same confounder structure.

43 negative control outcomes selected (based on clinical knowledge).

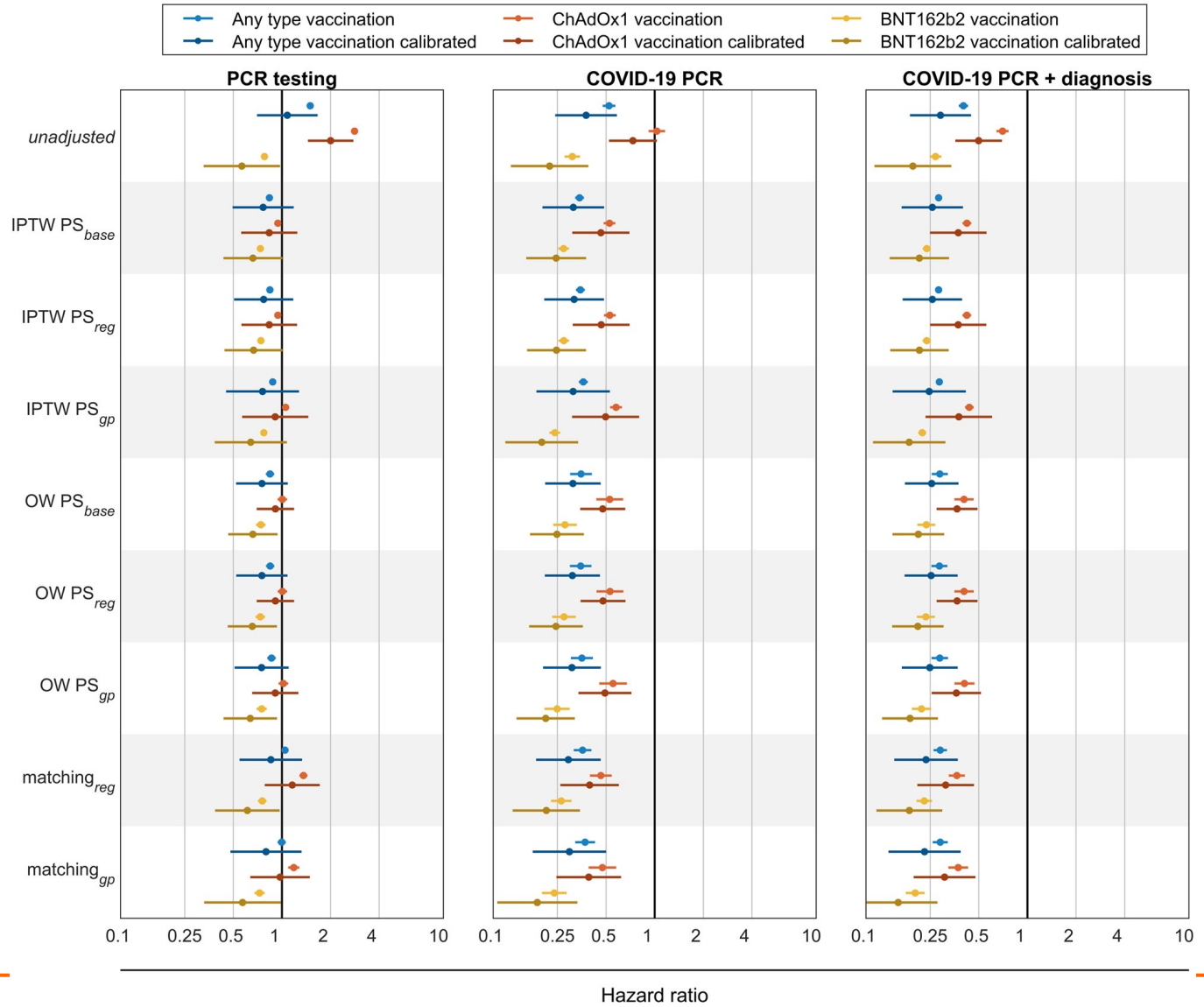
We would expect only 5% of NCO to be significant.

We used **empirical calibration** to account for unobserved confounding.

Negative control outcome



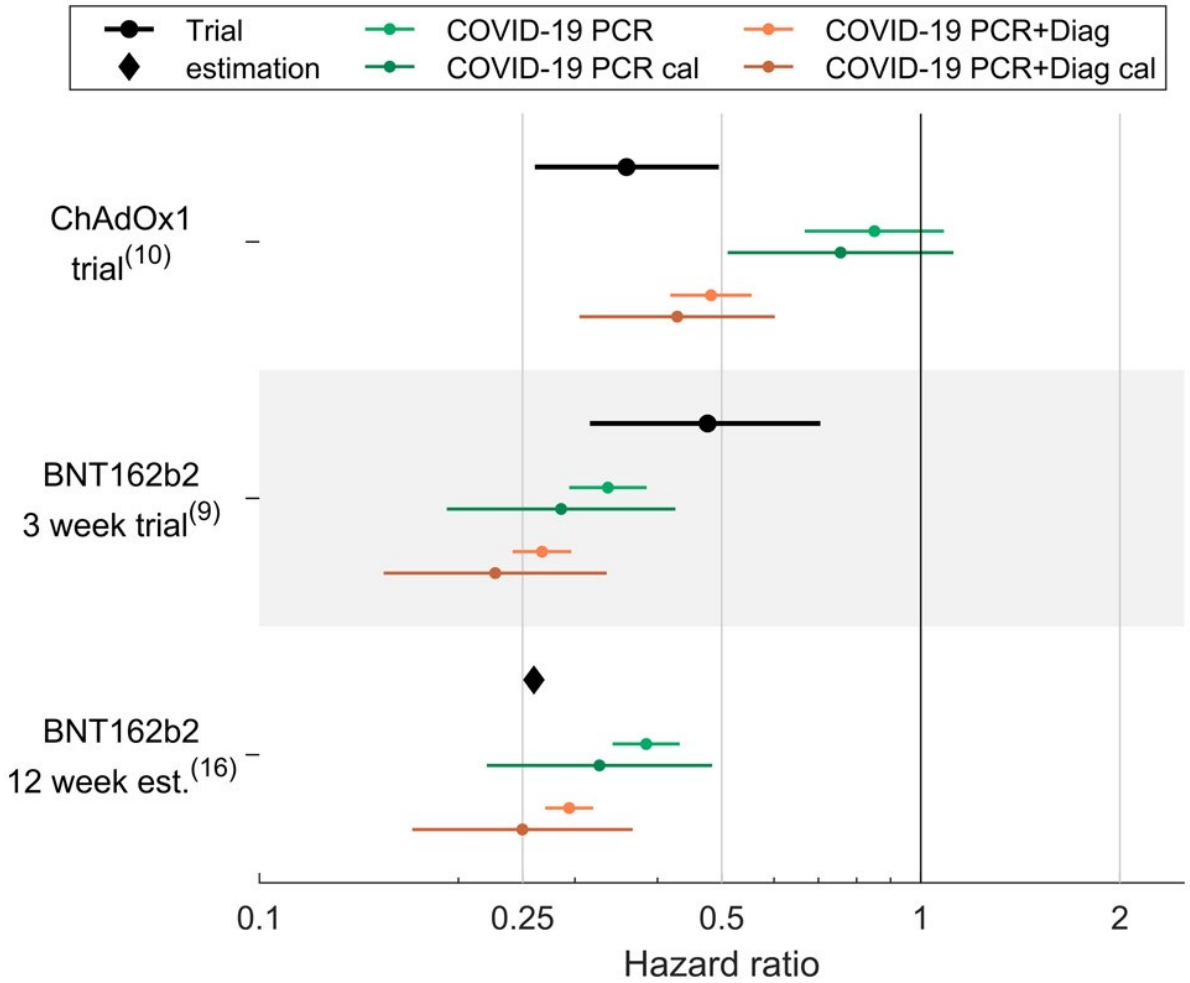
Outcomes


Any type vaccination
Any type vaccination calibrated
ChAdOx1 vaccination
ChAdOx1 vaccination calibrated
BNT162b2 vaccination
BNT162b2 vaccination calibrated
PCR testing
COVID-19 PCR
COVID-19 PCR + diagnosis
unadjusted
IPTW PS base
IPTW PS reg
IPTW PS gp
OW PS base
OW PS reg
OW PS gp
matching reg
matching gp
0.1
0.25
0.5
1
2
4
10
Hazard ratio

Learnings

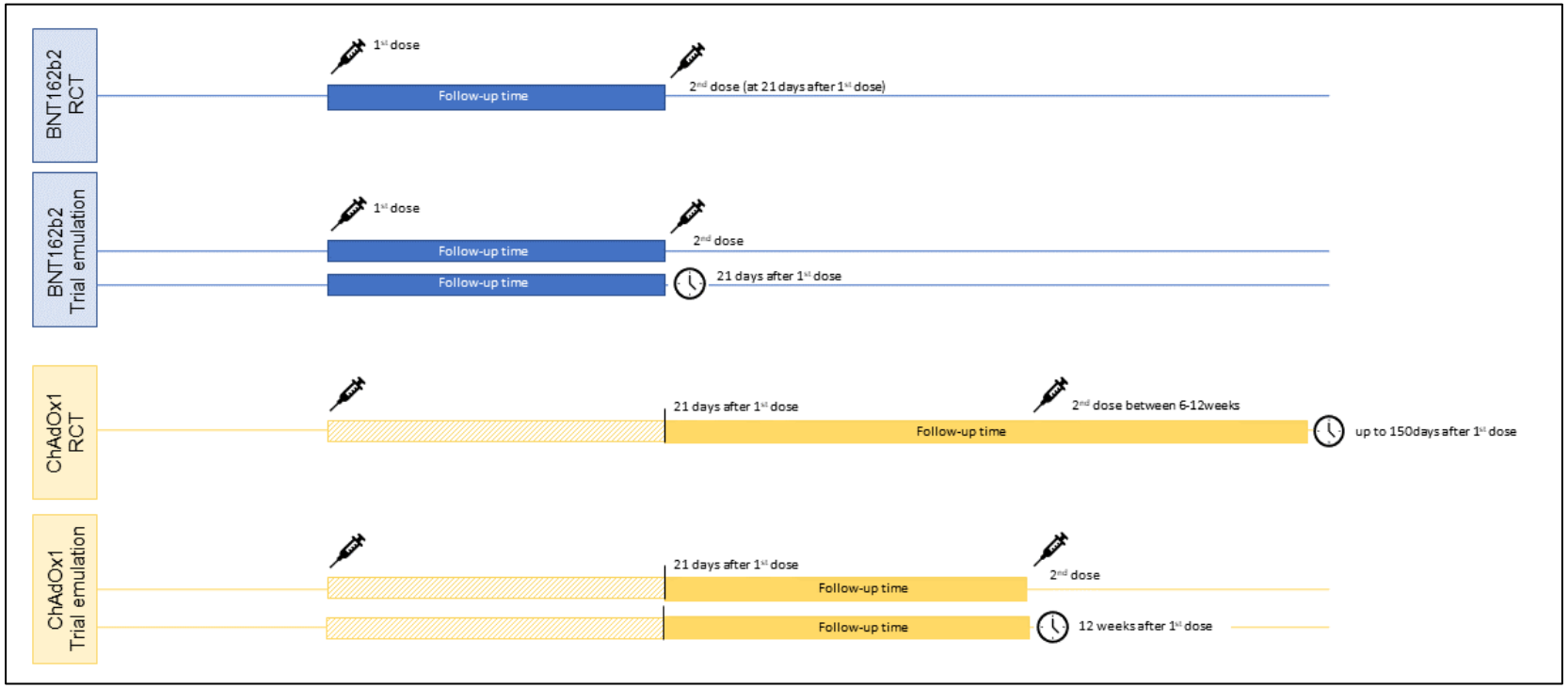
- Overlap weighting is the one that minimises better observed and unobserved confounding.
- Region is important for covid-19 related research.

Trial emulation



Trial emulation limitations

- Differences in study population (e.g. age)
- Follow-up time:



Learnings

- Real-world evidence successfully replicated the findings of phase 3 trials for COVID-19 vaccine effectiveness.
- Despite a lack of trial data, our findings suggest that first-dose BNT162b2 provides effective protection against SARS-COV-2 infection for up to 12 weeks, in line with UK's Joint Committee on Vaccination and Immunisation modelling and subsequent vaccination strategies.

Volume 53, Issue 1
February 2024

Article Contents

JOURNAL ARTICLE

Observational methods for COVID-19 vaccine effectiveness research: an empirical evaluation and target trial emulation

Martí Català, Edward Burn, Trishna Rathod-Mistry, Junqing Xie, Antonella Delmestri, Daniel Prieto-Alhambra ✉, Annika M Jödicke

International Journal of Epidemiology, Volume 53, Issue 1, February 2024, dyad138,
https://doi.org/10.1093/ije/dyad138
Published: 13 October 2023 **Article history** ▾

https://doi.org/10.1093/ije/dyad138

Application

- The effectiveness of COVID-19 vaccines to prevent long COVID symptoms: staggered cohort study of data from the UK, Spain, and Estonia
- The role of COVID-19 vaccines in preventing post-COVID-19 thromboembolic and cardiovascular complications

Long covid study

CPRD GOLD	United Kingdom	primary care EHR	3·1 million
CPRD AURUM	United Kingdom		13·3 million
SIDIAP	Catalonia (Spain)	primary care EHR with linked hospital data	5·9 million
TARTU ÜLIKOOL 1632 UNIVERSITAS TARTUENSIS	Estonia	National Health insurance claims	440’000

Databases had linked COVID-19 testing and vaccination status

All mapped to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to enable federated analytics.

Long covid study

Four staggered cohort studies based on UK Government vaccination priority groups:

<table>
<tr><td rowspan="2"></td><td colspan="5">Study period</td></tr>
<tr><td colspan="4">Enrolment periods</td><td>Follow-up</td></tr>
<tr><td></td><td>04/01–27/01</td><td>28/01– 28/02</td><td>01/03-13/04</td><td>14/04–31/07</td><td></td></tr>
<tr><td rowspan="2">STUDY COHORT
Age ≥ 75
(risk groups 2+3)</td><td>Vaccinated</td><td colspan="4"></td></tr>
<tr><td>Unvaccinated</td><td colspan="4"></td></tr>
<tr><td colspan="2" rowspan="2">STUDY COHORT
Age ≥ 65,
clinically extremely vulnerable/ at-risk patients
(risk groups 4-6)
+
Eligible unvaccinated adults in risk groups 2+3</td><td>Vaccinated</td><td colspan="3"></td></tr>
<tr><td>Unvaccinated</td><td colspan="3"></td></tr>
<tr><td colspan="3" rowspan="2">STUDY COHORT
Age ≥ 50 (risk groups 7-9)
+
Eligible unvaccinated adults in risk groups 2-6</td><td>Vaccinated</td><td colspan="2"></td></tr>
<tr><td>Unvaccinated</td><td colspan="2"></td></tr>
<tr><td colspan="4" rowspan="2">STUDY COHORT
Age ≥ 18 (risk group 10)
+
Eligible unvaccinated adults in risk groups 2-10</td><td>Vaccinated</td><td></td></tr>
<tr><td>Unvaccinated</td><td></td></tr>
</table>

Follow-up censored at end of observation and subsequent vaccine dose

Long covid study

Federated analyses

- Common analytical script was developed
- Adaptation to mimic country-specific vaccine rollout
- Analyses run locally by each data partner

Statistical analyses

- Large-Scale Propensity Scores incl. key variables
- Overlap weighting
- Diagnostics: assessment of covariate balance
- Fine-grey regression to estimate Hazard Ratios while accounting for death as competing risk
- Unmeasured confounding: Negative control outcomes and empirical calibration
- Random effect meta-analyses

Long covid study

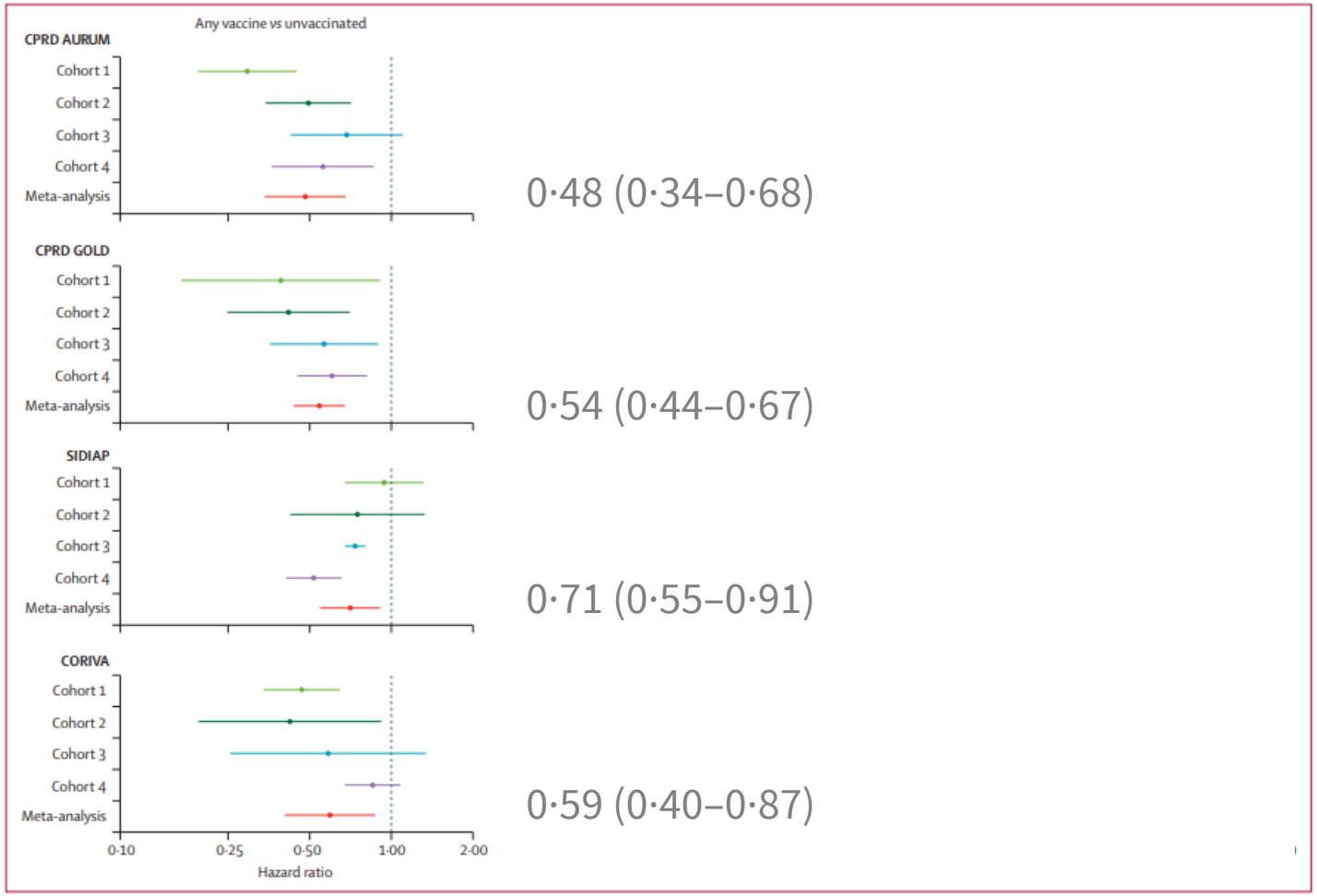


Figure 2: **Forest plots of vaccine effectiveness against long COVID**
Calibrated subdistribution hazard ratios from CPRD GOLD, CPRD AURUM, SIDIAP, and CORIVA for cohorts one to four and meta-analyses. Comparative effectiveness analyses for ChAdOx1 in SIDIAP and CORIVA were not fully conducted due to 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.

Long covid study

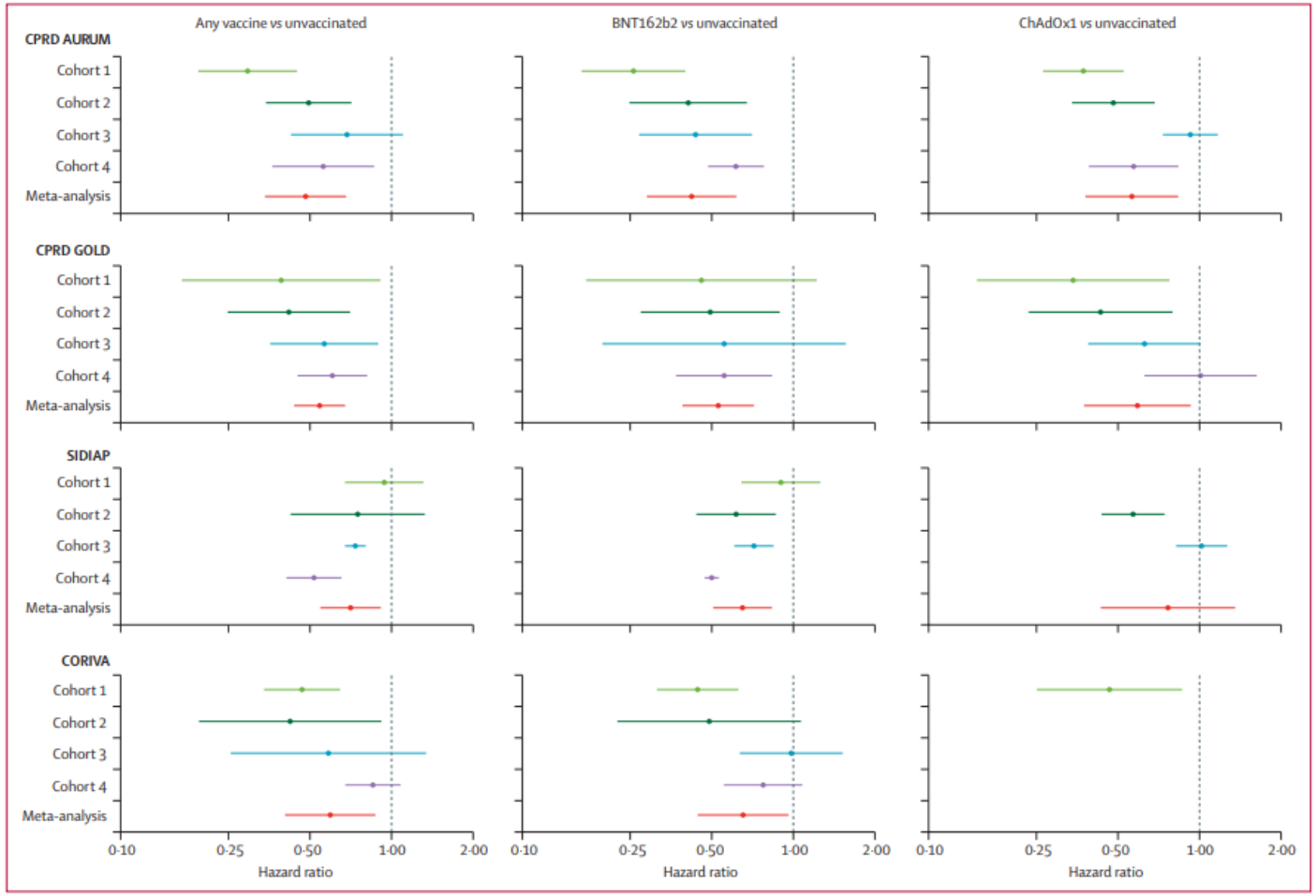


Figure 2: **Forest plots of vaccine effectiveness against long COVID**
Calibrated subdistribution hazard ratios from CPRD GOLD, CPRD AURUM, SIDIAP, and CORIVA for cohorts one to four and meta-analyses. Comparative effectiveness analyses for ChAdOx1 in SIDIAP and CORIVA were not fully conducted due to 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.

Long covid study

ARTICLES · Volume 12, Issue 3, P225-236, March 2024 · ***Open Access***

Download Full Issue

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

Martí Català, PhD [a] · Núria Mercadé-Besora, BA [c] · Raivo Kolde, PhD [d] · Nhung T H Trinh, PhD [f] · Elena Roel, PhD [c] · Edward Burn, PhD [a] · et al. Show more

Affiliations & Notes · Article Info · Linked Articles (1)

Shiny app: https://dpa-pde-oxford.shinyapps.io/LongcovidVaccineEffectiveness/

doi: 10.1016/S2213-2600(23)00414-9

Post-covid-19 events

Cardiac risk factors and prevention
Original research

The role of COVID-19 vaccines in preventing post-COVID-19 thromboembolic and cardiovascular complications

PDF

PDF + Supplementary Material

Núria Mercadé-Besora [1, 2, 3], Xintong Li [1], Raivo Kolde [4], Nhung TH Trinh [5], Maria T Sanchez-Santos [1], Wai Yi Man [1], Elena Roel [3], Carlen Reyes [3], Antonella Delmestri [1], Hedvig M E Nordeng [6, 7], Anneli Uusküla [8], Talita Duarte-Salles [3, 9], Clara Prats [2], Daniel Prieto-Alhambra [1, 9], Annika M Jödicke [1], Martí Català [1]

Shiny app: https://dpa-pde-oxford.shinyapps.io/PostCovidComplications/
doi: 10.1136/heartjnl-2023-323483

Reproducibility

- Effectiveness of COVID-19 vaccines to prevent long COVID: data from Norway

Same analyses replicated in Norwegian Linked Health Registries

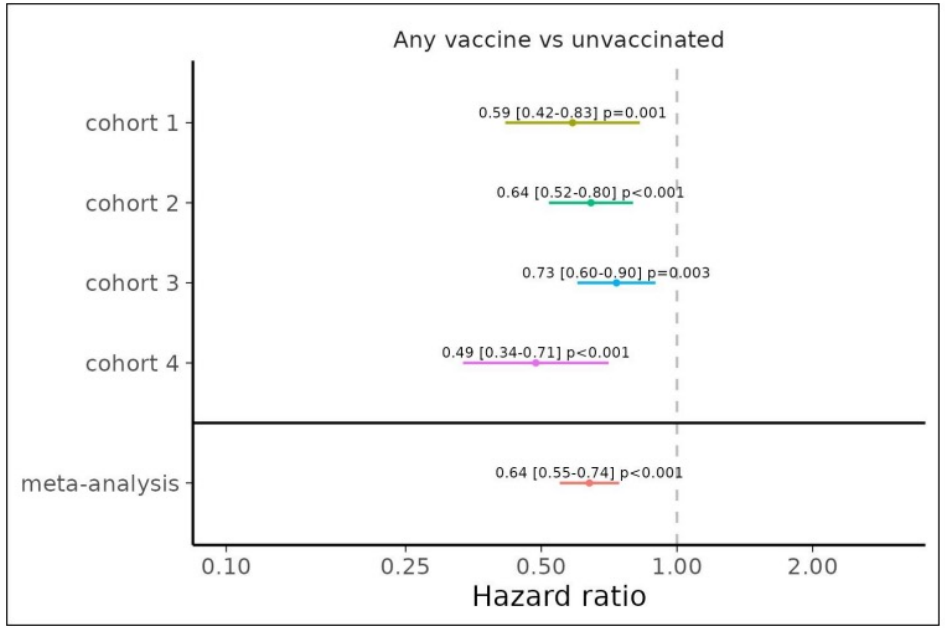
NLHR cover entire Norwegian population (5·4 million inhabitants)

Study population:

- 2 364 651 vaccinated people
- 1 532 935 unvaccinated people

Vaccination with any COVID-19 vaccine reduced the risk of developing long COVID symptoms:

Meta-analytic HR: 0·64 (0·55–0·74)



Reproducibility

CORRESPONDENCE · Volume 12, Issue 5, E33-E34, May 2024

Download Full Issue

Effectiveness of COVID-19 vaccines to prevent long COVID: data from Norway

Nhung TH Trinh [a] ✉ · Annika M Jödicke [b] · Martí Català [b] · Núria Mercadé-Besora [b] · Saeed Hayati [a] · Angela Lupattelli [a] · et al. Show more

Affiliations & Notes · Article Info · Linked Articles (1)

10.1016/S2213-2600(24)00082-1

Team effort

Thank you very much for your attention ☺

marti.catalasabate@ndorms.ox.ac.uk