OpenSAFELY Research Protocol:

COMPARES-vaccines: COMmon Protocol for the Analysis of Relative Effectiveness and Safety - of Covid-19 vaccine products

OpenSAFELY project number and summary: Project 174
https://jobs.opensafely.org/echo-evaluation-of-covid-19-vaccine-histories-using-opensafely/

GitHub repository: https://github.com/opensafely/COMPARES-vaccines

Authors

Martina Pesce[1], Edward PK Parker[2], Helen McDonald[3], Christopher Wood[1], Venexia Walker[4], Frederica Longfoot[1], Alain Amstutz[4], Andrea L Schaffer[1], William J Hulme[1]

1 Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
2 Vaccines and Immunisation Health protection Research Unit, London School of Hygiene and Tropical Medicine, UK
3 University of Bath, UK
4 Electronic Health Records Group, Population Health Sciences, Bristol Medical School, University of Bristol, UK

Version number	Version date	Summary of changes made
0.1	2025 - 08 - 18	First draft for receiving wider feedback beyond the core ECHO team.

Table of Contents

Title

COMPARES vaccines: COMmon Protocol for the Analysis of Relative Effectiveness and Safety of Covid-19 vaccine products

Abstract

Introduction

We present a study protocol outlining a target trial emulation that compares the safety and effectiveness of Covid-19 vaccine products during a single vaccination campaign in England. Covid-19 vaccination campaigns have been delivered each Spring and Autumn since 2021 and this is set to continue for the foreseeable future. The protocol is therefore designed to be adapted for each successive campaign, to accommodate changes to campaign dates, the vaccine products available, eligibility criteria, and other factors. The underlying analytical approach is the same from season to season.

Methods

The protocol describes a comparative safety and effectiveness analysis between two Covid-19 vaccine products available concurrently for a given vaccination campaign. Two approaches to control for confounding are used: one-to-one matching without replacement and inverse probability of treatment weighting. A variety of safety and Covid-19-related effectiveness endpoints are considered, informed by availability, reliability, and relevance. We will repeat analyses in a variety of population subgroups, and with various accompanying sensitivity analyses and baseline balance checks. The hypothetical randomised trial(s) that the study design attempts to emulate is also described.

We will use the OpenSAFELY research platform, which provides secure access to routinely-collected health records for millions of people in England, to conduct analyses.

All analyses are encoded explicitly in R scripts, which are fully executable against simulated dummy data before being run against real data in OpenSAFELY.

Discussion

The regularity and similarity of Covid-19 vaccination campaigns in England, including future campaigns, coupled with the availability of reliable routinely-collected health data on who is getting which vaccine and when, provides an opportunity to specify a *single* analysis protocol that can be reused across *multiple* campaigns.

Pre-registering this protocol enables us to obtain feedback on our analytic approach independently of results being known, enhancing robustness. It also enables eventual results to be rapidly published, as the methods have been pre-reviewed.

Plain english summary

[lay summary]

Background

The national Covid-19 vaccination programme in the UK began mass roll-out of Covid-19 vaccines on 8 December 2020 in a campaign designed to deliver a two-dose primary series, first in those at high risk of infection and death from Covid-19, and later in the entire adult population. This was followed by the first Covid-19 vaccine "booster" campaign in September 2021, and subsequent campaigns have continued in Spring and Autumn each year.

Each campaign to date has administered Covid-19 vaccines to millions of people. At least two vaccine products from different manufacturers have been used concurrently in the eligible population, to mitigate any unforeseen supply or safety issues. Products typically change from campaign to campaign, as newer variant-adapted versions become available and market conditions change. All vaccines have approval by the Medicines and Healthcare Products Regulatory Agency (MHRA) in the UK, following rigorous evaluation in randomised controlled trials to evaluate safety, immunogenicity, and efficacy against events such as test-confirmed SARS-CoV-2 infection and symptomatic disease.

However, observational studies provide vital information about vaccine effectiveness and safety once a vaccine is licensed. Follow-up in trials may be too short, or the recruited population too small, to evaluate longer-term or rarer outcomes, and vaccines may act differently in clinically diverse population groups, against newer viral variants, or in routine community vaccination settings. People's behaviours may also differ following a known vaccination, which may in turn influence the risk of experiencing outcomes and alter a vaccine's efficacy and safety profile compared with blinded randomised trials. Post-authorisation evaluations using observational data can fill these evidence gaps, and the concurrent availability of two or more vaccines in a single vaccine campaign, administered in millions of people, provides an invaluable opportunity to make direct comparisons between competing vaccine products for important effectiveness and safety endpoints.

Our group has previously compared effectiveness and safety of Covid-19 vaccines using the OpenSAFELY research platform, covering millions of patients registered at practices in England, and the ongoing twice-yearly booster campaigns allow us to continue this work.

The expected continuation of Covid-19 vaccination campaigns in Autumn and Spring each year provides an opportunity to specify a *single* protocol covering *multiple* studies, one study for each campaign.

This approach is bolstered by the anticipated similarity of the campaigns, and therefore the similarity of the methods that can be deployed to evaluate vaccines administered in each of them. Further, studies evaluating comparative safety and effectiveness between two vaccines are less vulnerable to confounding bias than comparing vaccination versus no vaccination due to active interventions in both groups. Similarly, for one-time interventions compared with

sustained or dynamic interventions where time-varying confounding is a greater problem. Thus, comparative vaccine evaluations are more amenable to a standard approach, with reduced need for bespoke, context-specific design choices to address confounding.

In such settings, confounding may be more tractable than with unexposed comparators, or with sustained or dynamic exposures. Thus they may be more amenable to analysis by standardised protocols given that fewer bespoke, context-specific design choices are required.

We present a *common analytic protocol* for an analysis comparing safety and effectiveness of two Covid-19 vaccines delivered in the same vaccination campaign in England, conducted in OpenSAFELY, that can be used across multiple vaccine campaigns.

Features of the Common Protocol

The following characteristics of each vaccine campaign will vary over time:

Study start date: the earliest date that somebody could meet the eligibility criteria, which will usually be the first day of the vaccination campaign of interest
Study end date: the last date that somebody could meet the eligibility criteria.
The vaccine products available: the two, or more, Covid-19 vaccine products being used, and to be compared. These will be referred to as product A and product B.
Eligibility: those who are eligible for inclusion in the study, which will incorporate both receipt of the vaccine products of interest, documented eligibility to receive those vaccines (e.g., due to age or clinical vulnerability).

The following characteristics of the analysis plan *may* vary for different campaigns depending on the changing context of each campaign, and to accommodate any changes in data availability:

- Effectiveness, safety, and negative control endpoints
- Inclusion criteria
- Subgroup analyses
- Confounding characteristics (inclusion and definition)

This protocol will be versioned, with date stamps. Amendments will be comprehensively documented and justified in a change log, and tied to a specific release of the github repository.

Objectives / Research questions

The overall objective is to compare the safety and effectiveness of Covid-19 vaccine products used concurrently during each vaccination campaign in England. Comparisons will be made within specific vaccine-eligible population groups, principally those eligible due to age or due to clinical vulnerability. The research question can be summarised as: amongst people in receipt of either product A or product B, does one product have a superior safety or effectiveness profile compared with the other?

Methods

The OpenSAFELY Platform

All data will be linked, stored and analysed securely using the OpenSAFELY platform, https://www.opensafely.org/, as part of the NHS England OpenSAFELY Covid-19 service. No

GP data from patients who have registered a Type-1 Opt out with their GP surgery will be included in this study. Data include pseudonymised data such as coded diagnoses, medications and physiological parameters. No free text data are included. All code will be shared openly for review and re-use under MIT open license. Detailed pseudonymised patient data is potentially re-identifiable and therefore will not be shared. A detailed description of the OpenSAFELY platform is available in Nab et al. (1)

Access to the underlying identifiable and potentially re-identifiable pseudonymised EHR data is tightly governed by various legislative and regulatory frameworks, and restricted by best practice. The data in the NHS England OpenSAFELY Covid-19 service is drawn from General Practice data across England where TPP are the data processors.

EMIS and TPP are the two software vendors providing electronic patient management systems to the vast majority of GPs in England. OpenSAFELY is deployed inside their respective data centres where patient records are stored, creating a Secure Data Environment for researchers to conduct analyses remotely, without seeing the sensitive data within. All analytic scripts are openly shared, and queries submitted to run against the database are publicly logged.

TPP developers initiate an automated process to create pseudonymised records in the core OpenSAFELY database, which are copies of key structured data tables in the identifiable records. These pseudonymised records are linked onto key external data resources that have also been pseudonymised via SHA-512 one-way hashing of NHS numbers using a shared salt. University of Oxford, Bennett Institute for Applied Data Science developers and PIs, who hold contracts with NHS England, have access to the OpenSAFELY pseudonymised data tables to develop the OpenSAFELY tools.

These tools in turn enable researchers with OpenSAFELY data access agreements to write and execute code for data management and data analysis without direct access to the underlying raw pseudonymised patient data, and to review the outputs of this code. All code for the full data management pipeline — from raw data to completed results for this analysis — and for the OpenSAFELY platform as a whole is available for review at github.com/OpenSAFELY.

Further details of the data sources planned for use in this protocol are described in the Data Sources section.

Study Design

Study type An active-comparator new-user (ACNU) design, comparing individuals vaccinated with product A compared with product B.

We will use two strategies for confounder control:

- **Matching**. We will use 1:1 matching without replacement. This approach leads to a reduction in the effective sample size, and an alteration of the inferential population, as vaccine recipients who cannot be matched are excluded from the analysis. However, it

ensures tight control of matching variables and for continuous confounders does not require any specification of their underlying (uni- or multi-variate) functional form. Extensive balance checks will be used to evaluate baseline imbalance for factors not explicitly matched on.

- **Inverse Probability Weighting (IPW).** We will model the probability of receiving product A versus product B and construct pseudo-populations where the probability of receipt of either product is balanced between the two groups on average.

For completeness, we also provide the implied hypothetical randomised trial that this study design attempts to emulate, following Hernan et al 2025.

Index date / Time zero Date of vaccination.

Data sources

Primary care data - TPP & EMIS

SystmOne is a primary care clinical information system run by TPP, used by roughly one third of GP practices in England, with records for approximately 44% of the English population.(2)

This system captures information about a patient that is electronically recorded or accessed by GPs, including symptoms, investigations, test results, diagnoses, prescriptions, certain demographic and social characteristics, etc.

Patients who have opted out of allowing their data to be shared for research and planning (i.e., for purposes beyond direct care, known as Type 1 Opt Outs), are not included.

ONS registered deaths

Date and cause of death based on information recorded when deaths are certified and registered in England and Wales. The death record includes the underlying cause of death, i.e., the medical condition judged to be the *underlying* cause according to the rules of the *10th Revision of the International Classification of Diseases* (ICD-10) and up to 15 additional medical conditions considered to be *contributory* causes.

Hospital admissions

Admitted Patient Care (APC) is the national data set for hospital admissions. Admitted Patient Care Spells (APCS) is part of Hospital Episode Statistics (HES) and is provided to OpenSAFELY via NHS Digital's Secondary Use Service (SUS). It includes admission and discharge dates; routes of admission; reason for admission; diagnoses (ICD-10); treatments; discharge destination.

Emergency attendances

The Emergency Care Data Set (ECDS) is the national data set for urgent and emergency care. ECDS is part of HES and is provided to OpenSAFELY via NHS Digital's SUS. It includes attendance, admission and discharge dates; locations; presenting complaints; diagnoses (SNOMED-CT).

Study Population

Cohorts and Eligibility Criteria

People receiving product A or product B between the study start date and the study end date will be considered for inclusion. We will analyse three primary cohorts in each campaign:

- **Age:** people who are eligible for vaccination due to their age. The age-eligibility threshold may differ by campaign.
- **Clinically vulnerable:** people who are eligible for vaccination due to clinically vulnerability (“CV”). Clinical vulnerability encompasses various chronic or immunosuppressive conditions and medications as set out in the Green Book chapter 14a. The definition of clinical vulnerability differs by campaign; for example, in spring, only immunosuppressed meet this definition, whereas in autumn it includes many more conditions. We do not incorporate all clinically vulnerability criteria (e.g., care home staff) due to under-ascertainment in the health record.
- **Care home residency:** people who are eligible for vaccination due to living in a care home or nursing home.

Cohorts are not mutually exclusive and we do not restrict each cohort to those who meet one and only one criterion. Each cohort will be analysed independently. See Table 2 for more details on how these cohorts are defined.

Before selecting cohort-specific recipients, we will apply the following eligibility criteria:

- Registered at a GP practice at the time of vaccination for at least 12 weeks prior to vaccination;
- Complete information on age, sex, deprivation, and Integrated Care Boards (ICBs, a geographical grouping of NHS and Local Authorities);These characteristics will be assessed as at the date of vaccination;
- Aged between 16 and 99 years old, inclusive;
- Not a hospital in-patient at the time of vaccination.

For certain outcomes, we will apply outcome-specific exclusions based on susceptibility (for example, for menorrhagia we will only include women of relevant age). These are listed against each specific outcome where applied.

If availability of or eligibility for product A and product B differs, we will restrict study dates and eligibility so that receipt of each product is a possibility for each enrolled person. For example, if product A is available from the start of the campaign but product B is only available 4 weeks later, the study start date will be 4 weeks after the start of the campaign. Or for example if product A is unavailable (or only extremely rarely) to people aged under 50 years, we will restrict eligibility to people aged 50 years and over.

Study Measures

This section provides an overview of study measures used in the analyses. Supplementary Table 1 provides comprehensive details of the operational definitions, with links to codelists and logic, for each of these variables.

Exposure

Receipt of either product A or product B. The date of exposure, time-zero, is the date that the vaccine was received, as recorded in the health record.

If in a given campaign a third (or more) product is also available, and at least 10% of recipients in that campaign receive it, then this will be included in analyses, and pairwise comparisons will be made between all products.

If only one product is widely available, we will report baseline characteristics of those in receipt of the vaccine and post-vaccination incidence rates, and no comparative analyses will be conducted.

Outcomes

For each outcome, the first event occurring between the vaccination date and the end of follow up (inclusive) will be used.

Additionally, we will use outcome events occurring up to 24 weeks prior to the vaccination date to assess baseline balance.

The primary outcome horizon is 24 weeks after day zero. Outcomes rates on each day will be available in supplementary materials.

In general it is not possible to distinguish temporal order of vaccination and outcome events if they occur on the same day. The main exception is death, which necessarily follows vaccination. In this study, same-day events are assumed to occur after vaccination, and we do not exclude

people on the basis of same-day events. This is realised by adding a day to all outcome event dates. This is not a sensible strategy in general for *treat versus do not treat* comparisons, where the occurrence of an event might vastly reduce the likelihood of, or preclude, treatment. However, in head-to-head studies like this, this is only a problem if the occurrence of a given outcome prior to vaccination differentially impacts the likelihood of receiving product A versus product B, which is unlikely. Taking this approach also means we do not miss rapid-onset safety outcomes, such as anaphylaxis.

Table 1: Effectiveness, safety and negative control outcomes

Outcomes	**Additional exclusions**	**Data source**				**Notes**
		GP coded event	**A&E visit**	**Hospital admission**	**Death**	
Effectiveness events						
Covid-19 attendance to A&E			X			
Covid-19 admission to hospital				X		
Covid-19 admission to critical care				X		
Covid-19 death					X	
Safety events						
Guillain-Barre Syndrome		X		X		
Bell's Palsy		X	X	X		
Venous thrombotic event		X		X	X	
Arterial thrombotic event		X		X	X	
Thrombotic thrombocytopenia		X		X	X	
Myocarditis		X	X	X	X	
Pericarditis		X	X	X	X	
Menorrhagia	Woman aged 16 - 55, no hysterectomy	X	X	X		
Erythema Multiforme		X	X	X		

Anaphylaxis		X	X	X	X	
Negative controls events						
Otitis		X	X	X		Capturing differences in susceptibility to infection
Cellulitis		X	X	X		Capturing differences in susceptibility to infection

For all outcomes except death, we will report the occurrence of at least one of these events for each person during the 1-year period prior to vaccination, and compare between products as an evaluation of baseline balance.

Censoring and competing events

Death is a competing risk for all non-fatal outcomes in this study (that is, death precludes the occurrence of any subsequent non-fatal outcomes). As death is expected to occur rarely over the duration of follow-up, we will censor follow-up at death, rather than explicitly modelling death as a competing risk.

If a patient deregisters from the practice that they were registered with at baseline, then subsequent outcomes might be unobserved. Again, as de-registration is expected to occur rarely over the duration of follow-up, we will censor follow-up at deregistration.

Covariates

As specified elsewhere in this protocol, the following additional variables will be used for at least one of:

- Study eligibility
- Matching characteristics in the matched analysis
- Predictors of product received in the weighted analysis
- Checks of baseline balance between treatment groups
- Stratifying characteristics in subgroup analyses

Variables may be operationalised differently in different contexts. For example age will be categorical in subgroup analyses, but used continuously for a propensity-of-treatment model.

Table 2: Eligibility, subgroup, confounder, and other variables
[incomplete; to finish]

Variable	Possible values	Notes
Demographics		
Age	[integers]	Based on rounding birth dates down to the first of the month
Sex	Male; Female	
Ethnicity	White; Mixed; Asian or Asian British; Black or Black British; Chinese or Other Ethnic Groups; Not stated.	As recorded in the GP record. This may be explicitly self-reported assumed by the practice.
Area deprivation (Rank of English Index of Multiple Deprivation, IMD)	Rounded to nearest 100	
NHS region	East of England; London; Midlands; North East and Yorkshire; North West; South East; South West	
Care home residency	Yes; No	
Clinical history		
PRIMIS variables	TODO	
Clinical vulnerability		Our operationalisation of clinical vulnerability differs from Green Book and real-world vaccine eligibility assessment by GPs and other healthcare professions. This is a pragmatic decision based

		on reliability of health records to identify certain characteristics.
Documented prior SARS-CoV-2 infection		
Time since prior Covid-19 infection / disease	0-28 days, 29-84 days, 84+ days, never.	
Vaccination history		
Same-day vaccination with Influenza and / or RSV vaccines	Flu only; RSV only; Flu and RSV; None.	
Campaign of previous vaccine dose	Campaign n-1; Campaign n-2; Campaign n-3^{+}; No prior vaccine recorded.	
Product of previous vaccine dose	[campaign dependent]	
Heterologous prior vaccine	[campaign dependent]	

Missing data

In routinely-collected clinical data where the presence of clinical codes indicate the occurrence of an event or condition, “missing” values are only relevant for age, sex, area residence and area deprivation, IMD, ethnicity, and BMI, which we know have values but may not be recorded. Of these, people with missing age, sex and IMD are excluded (see Cohorts and Eligibility). Where ethnicity is missing, an “Unknown” category is used.

All remaining patient characteristics and outcomes are defined by the presence or absence of clinical codes or events in the health record. All analyses assume these events are recorded accurately and make no adjustments for any potential biases due to misclassification / underascertainment. Possible deviations from these assumptions are discussed as limitations only (for example under-ascertainment of safety events).

Statistical Analysis

Descriptive analyses

An inclusion flowchart will be created to report the flow of potentially eligible to actually eligible people, with the following stages:

1. Receipt of Covid-19 vaccination between start date and end date
2. Receipt of product A or product B only
3. No Covid-19 vaccination in the preceding X days
4. Aged 16 to 99 inclusive and no missing sex or IMD
5. [Additional cohort-specific inclusion criteria]

Baseline characteristics of recipients of each product will be tabulated, and standardised mean differences between groups will be calculated using love plots.

This will be done both before and after attempts to achieve balance via confounder adjustment (see below). Pre-adjustment differences will be shown to understand how recipients of each product differ naturally in the unadjusted eligible population. Post-adjustment differences will be described for each analysis approach (matching and IPW).

Comparative analyses

The main analysis proceeds in 3 steps:

1. **Estimate balancing weights:** We will use two complementary approach to estimate weights that will balance characteristics between recipients of different products:
    a. matching, which creates weights of 1 if a match is identified and 0 otherwise;
    b. propensity scores, which creates weights equal to the inverse of the estimated probability of receiving the product that was actually received.
2. **Estimate the cumulative incidence for each product:** The balancing weights from step 1 will be used in two complementary approaches to estimate the cumulative incidence of each outcome given the vaccine product received:
    a. Kaplan-Meier estimation, a non-parametric approach where only minimum assumptions about time-specific effects are required.
    b. g-computation using Pooled Logistic Regression, a parametric approach where the time-specific effects are modelled explicitly.
3. **Compare the cumulative incidences:** The product-specific cumulative incidences from step 2 will be contrasted using risk differences and cumulative incidence rate ratios.

There are therefore four comparisons for each outcome, one for each combination of balancing weights and cumulative incidence estimation. None of these approaches are a-priori more valid than the other, and all will be reported.

Full details for each approach is provided below.

Step 1: Estimate balancing weights

Matched analysis using 1-1 matching without replacement

Recipients of each product will be matched 1-1 without replacement on a set of characteristics known at the time of vaccination. Any person for whom a match is not found is discarded (weight=0), otherwise are included (weight=1).

The following characteristics will be matched on:

- date of vaccination (3-day caliper);
- campaign of previous Covid-19 vaccine dose (1, 2, or 3+ campaigns ago);
- sex (male or female);
- age (3-year caliper and within age groups 50-64, 65-75, 75-79, 80-84, 85+);
- Index of Multiple Deprivation (IMD, within a 5000 IMD-rank caliper);
- NHS region as a surrogate for geographical region;
- care home residency;
- any evidence of recent SARS-CoV-2 infection (including any of: positive SARS-CoV-2 test; infection as documented in primary care; Covid-19 hospital attendance or admission);
- current immunosuppression;
- morbidity count (grouped as 0, 1, or 2 or more conditions from the following list: diabetes; BMI over 40kg/m2; chronic heart disease; chronic kidney disease; chronic liver disease; chronic respiratory disease or severe asthma; chronic neurological disease; cancer within 3 years).

Baseline balance will be evaluated by comparing all people included in the matched analysis.

Justification for matched analysis there is no need to explicitly model the relationship between confounders and the outcome. Positivity violations in matched characteristics are automatically dealt with.

Matching will be performed using the MatchIt R package.

Weighted analysis using Inverse Probability of treatment Weights (IPW)

We will first create a model that estimates the propensity to receive either product A or product B based on variables observed at or before baseline. The weights are defined as the inverse of

these propensities, creating a pseudo-population where the propensity to receive each product is equal.

The following characteristics will be used in the propensity model:

- date of vaccination (3-knot spline);
- campaign of previous Covid-19 vaccine dose (1, 2, or 3+ campaigns ago);
- sex (male or female);
- age (3-knot spline);
- IMD rank (3-knot spline);
- STP as a surrogate for geographical region;
- care home residency;
- any evidence of prior SARS-CoV-2 infection (including any of: positive SARS-CoV-2 test; infection as documented in primary care; Covid-19 hospital attendance or admission);
- current immunosuppression;
- morbidity count (grouped as 0, 1, or 2 or more conditions from the following list: diabetes; BMI over 40kg/m2; chronic heart disease; chronic kidney disease; chronic liver disease; chronic respiratory disease or severe asthma; chronic neurological disease; cancer within 3 years).

Baseline balance will be evaluated by comparing the weighted population.

Justification for weighted analysis Given that severe outcomes are more reliably recorded in the health record, we focus on these outcomes in this analysis. Such outcomes are rarer, and so it is harder to reliably capture exposure-outcome relationships across all confounders due to fewer events. By modelling confounder-exposure relationships directly instead, we avoid this issue and gain statistical precision.

Weighting will be performed using the WeightIt R package.

Step 2: Estimate the product-specific cumulative incidence

Kaplan-Meier estimation

The cumulative incidence of the outcome is estimated using the Kaplan Meier (KM) estimator, for each (weighted) vaccine group. Confidence limits for the cumulative incidences will be derived using standard errors of the log of the cumulative incidence.

Pooled logistic regression

A pooled logistic regression model will be fitted that estimates, for each day of follow-up, the probability of experiencing the outcome of interest under receipt of each product. A parametric cumulative incidence curve can be derived using this model, estimating the cumulative incidence if (counterfactually) the entire population received product A or B.

Step 3: Compare the cumulative incidences

For each combination of step 1 and 2, we will estimate cumulative Risk Differences (RDs) and cumulative Incidence Rate Ratios (IRRs) comparing the products for each outcome. Confidence limits for the risk differences will be derived from the sum of squares of the standard errors of cumulative incidence, using Greenwood's formula. Confidence limits for the risk ratios and incidence rate ratios will be derived from the sum of squares of the standard errors of the log of the cumulative incidence.

We will also estimate IRRs within period-specific intervals defined by splits on weeks 1, 4, 8, 12, 16, and 20 and report in supplementary materials.

For effectiveness outcomes, we will report IRRs at 1 week to evaluate the potential for implausibly rapid divergence of the cumulative incidence, indicating baseline imbalance. Rapid divergence of cumulative incidence between products for these outcomes is implausible due to delays in immunological protection from vaccination.

Subgroup analyses

To explore any potential effect modification, we will repeat the main analyses independently within the following subgroups:

- Age (season dependent; 50-64, 65-74, 75-84, 85+)
- Previous product in season n-1 (A / B / none)
    - In Spring campaigns, for which vaccines are typically offered to a subset of those eligible in the preceding Autumn campaign, subgroups will include:
        - Product A in previous Autumn
        - Product B in previous Autumn
        - No vaccination in previous Autumn
    - In Autumn campaigns, in which previous vaccination will depend on eligibility (and uptake) across multiple preceding campaigns, subgroups will include:
        - Product A in previous Spring
        - Product B in previous Spring
        - Product A in previous Autumn, no vaccination in previous Spring
        - Product B in previous Autumn, no vaccination in previous Spring
        - No vaccination in previous Autumn or previous Spring
- Clinical vulnerability
- Immunosuppression
- Care-home residency
- Sex
- Ethnicity (White, Black, Asian, Mixed, Other, Unknown)
- IMD (quintiles)

Variables defining the subgroups will be dropped from the matching or weighting procedure for each subgroup analysis as appropriate.

Meta-analysis across different GP systems

Where both TPP and EMIS records are available, summary statistics will be generated independently within each system, and meta-analysed as follows:

- Vaccine coverage counts will be aggregated.
- Vaccine effectiveness and safety estimates will be meta-analysed using inverse variance weighting.

Hypothetical Target Trial

There are two hypothetical randomised trials implied by this study design, one each for the matched and weighted analyses.

Component		**Target trial Specification**	**Emulation**
Causal Estimand	**Eligibility criteria**	Eligible to be vaccinated in the current campaign as per Green Book;	Same, as ascertained in routinely-collected NHS data
	Treatment strategies	Receipt of Product A versus receipt of Product B. There are no post-baseline requirements. For example the strategy does not require that a person must not receive another Covid-19 vaccine within 20 weeks.	Same, as ascertained in routinely-collected NHS data
	Assignment	**Matched analysis** Randomised assignment, no blinding, fully balanced within pairs who are matched at enrollment / within strata defined by the matched characteristics. If a match is not found for an enrolled person, then they are discarded (ineligible) from the study. **Weighted analysis** Randomised assignment, no blinding. Assignment is weighted in accordance with the IPW model.	The vaccine product received, as recorded in routinely-collected data
	Outcomes	Various; see Table 1.	Same, as recorded in

			routinely-collected NHS data
	Start and end of follow-up	Date of vaccination until 24 weeks after vaccination.	Same. Temporal resolution is one day. Therefore, the ordering of events occurring on the same cannot be determined, except for death which is necessarily last. In this analysis, vaccination is assumed to occur at the start of the day, with outcomes occurring at the end of the day. Time-to-event for a person experiencing an outcome on the same day as vaccination would therefore be 1 day.
	Causal contrasts	Risk ratios and risk differences of the cumulative incidence. For the **Matched analysis**, the estimand is the *average treatment effect in the overlap (ATO) (i.e., in those who were matched)* For the **Weighted analysis**, the estimand is the *average treatment effect* (ATE)	Same.
Identifying assumptions		Conditional exchangeability for post-baseline death or deregistration. Note that vaccination is a one-time intervention, administered at (hypothetical) enrollment, and there are no post-baseline events that cause non-adherence to the assigned treatment strategy. For instance, the treatment strategy does not require individuals to receive no further vaccination within 6 months. Therefore, protocol adherence is 100% and, assuming perfect	Exchangeability conditional on measured confounders. See analysis plan.

	ascertainment of vaccination status, the intention-to-treat and per-protocol analyses coincide in both the hypothetical and emulated trial. Follow up is lost at death or deregistration, whereupon participants are censored.	
Estimator	For more details, see the statistical analysis plan.	

Limitations

Ascertainment of vaccine eligibility and outcomes is imperfect.

Although confounding in more tractable when comparing two vaccines than when compare vaccination against non-vaccination, unmeasured confounding is still a possibility

Disclosure Control

To manage the risk of disclosure–that is, the reidentification of individuals within the data– only aggregated results will be viewed and requested for release from the secure environment, and statistical disclosure control (SDC) will be applied to all released results as follows:

- All patient counts will be rounded. Values between 1 and 7 will be expressed as “<8” in tables. Values of eight or higher will be rounded to the nearest 5. Zeroes will remain zero.
- Proportions will be calculated using rounded numerator and denominator values. If the numerator is between 1 and 7 and the denominator is >=8, a value of “<8/denominator” will be used. For example if the numerator is 5 and the denominator is 16, the value reported will be “<8/16” or “<0.5”. If both the numerator and denominator are between 1 and 7, the value will be redacted outright.
- For cumulative incidence rates, the cumulative counts of outcome, censoring, and competing events, and the count of at-risk individuals, will be rounded to the nearest 5, and the cumulative incidences will be re-calculated from these rounded values.

To ensure SDC has been properly applied, all results will be reviewed by two trained output checkers prior to release from the secure environment, as per OpenSAFELY release policies.

Software and Reproducibility

Datasets will be created from the underlying database using ehrQL. All subsequent analyses will be conducted using R. All analytic code as well as codelists are archived online on GitHub.

Patient and Public Involvement and Engagement

The ECHO project has a Public Advisory Group (PAG), comprising 8 members of the public who meet regularly to discuss the project, feedback on public-facing materials, and provide experiences on Covid-19 vaccine availability, accessibility and desirability.

OpenSAFELY has involved patients and the public in various ways: we developed a public website that provides a detailed description of the platform in language suitable for a lay audience (https://opensafely.org); we have participated in two citizen juries exploring public trust in OpenSAFELY; we have co-developed an explainer video (https://www.opensafely.org/about/**);** we have patient representation who are experts by experience on our OpenSAFELY Oversight Board; we have partnered with Understanding Patient Data to produce lay explainers on the importance of large datasets for research; we have presented at various online public engagement events to key communities (e.g., Healthcare Excellence Through Technology; Faculty of Clinical Informatics annual conference; NHS Assembly; HDRUK symposium); and more. To ensure the patient voice is represented, we are working closely to decide on language choices with appropriate medical research charities (e.g., Association of Medical Research Charities). We will share information and interpretation of our findings through press releases, social media channels, and plain language summaries.

Dissemination

This Common Protocol will be used as a template for each campaign-specific analysis, with a separate protocol created for each campaign.

Analytic results from each study will follow the STROBE-RECORD reporting guidelines.

Manuscripts will be submitted to medRxiv at the same time as submission to a peer-reviewed journal.

Results and outputs will be discussed with our PAG.

Administrative

Timelines and approvals

NHS England will oversee the final approval for all publication ready papers, reports or presentations, to check that the outputs align with the stated application purpose.

Ethics

This study is classified as a Service Evaluation. Dr Jamie Lopez-Bernal, Consultant Epidemiologist at UKSHA, is the senior sponsor.

In addition to this sponsorship, all Covid-19 related studies led by the Bennett Institute have ethics approval up to April 2027 from the Health Research Authority (REC reference 20/LO/0651).

Funding

This study is part of the Evaluating Covid-19 Vaccine Histories using OpenSAFELY (ECHO) project, funded by NIHR (NIHR206900).

The OpenSAFELY platform is principally funded by grants from:

- NHS England [2023-2025];
- The Wellcome Trust (222097/Z/20/Z [2020-2024] and 311535/Z/24/Z [2025-2031]);
- Medical Research Council (MR/V015737/1 [2020-2021]).

Additional contributions to OpenSAFELY have been funded by grants from:

- Medical Research Council (MRC) via the National Core Study programme Longitudinal Health and Wellbeing strand (MC_PC_20030, MC_PC_20059 [2020-2022]) and the Data and Connectivity strand (MC_PC_20058 [2021-2022]);
- The National Institute for Health Research (NIHR) and the Medical Research Council (MRC) via the CONVALESCENCE programme (COV-LT-0009, MC_PC_20051 [2021-2024]);
- NHS England via the Primary Care Medicines Analytics Unit [2021-2024].

The views expressed are those of the authors and not necessarily those of the NIHR, NHS England, UK Health Security Agency (UKHSA), the Department of Health and Social Care, or other funders.

Funders have no role in the study design, collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Information governance

NHS England is the data controller of the NHS England OpenSAFELY Covid-19 Service; TPP is the data processor; all study authors using OpenSAFELY have the approval of NHS England.(3) This implementation of OpenSAFELY is hosted within the TPP environment which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant.(4)

Patient data has been pseudonymised for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymised datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the NHS England OpenSAFELY Covid-19 service is via a virtual private network (VPN) connection; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; only aggregate statistical outputs leave the platform environment following best practice for anonymisation of results such as statistical disclosure control for low cell counts.(5)

The service adheres to the obligations of the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018. The service previously operated under notices initially issued in February 2020 by the the Secretary of State under Regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 (COPI Regulations), which required organisations to process confidential patient information for Covid-19 purposes; this set aside the requirement for patient consent.(6) As of 1 July 2023, the Secretary of State has requested that NHS England continue to operate the Service under the Covid-19 Directions 2020.(7) In some cases of data sharing, the common law duty of confidence is met using, for example, patient consent or support from the Health Research Authority Confidentiality Advisory Group.(8)

Taken together, these provide the legal bases to link patient datasets using the service. GP practices, which provide access to the primary care data, are required to share relevant health information to support the public health response to the pandemic, and have been informed of how the service operates.

References

1. Nab L, Schaffer AL, Hulme W, DeVito NJ, Dillingham I, Wiedemann M, et al. OpenSAFELY: A platform for analysing electronic health records designed for reproducible research. Pharmacoepidemiology and Drug Safety. 2024;33(6):e5815.
2. Andrews C, Schultze A, Curtis H, Hulme W, Tazare J, Evans S, et al. OpenSAFELY: Representativeness of electronic health record platform OpenSAFELY-TPP data compared to the population of England. Wellcome Open Res. 2022;7:191.
3. NHS Digital. The NHS England OpenSAFELY COVID-19 service - privacy notice [Internet]. 2023 [cited 2023 Jul 5]. Available from: https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/the-nhs-england-opensafely-covid-19-service-privacy-notice
4. NHS Digital. NHS Digital. 2023 [cited 2023 Jul 5]. Data Security and Protection Toolkit. Available from: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit
5. NHS Digital. NHS Digital. 2022 [cited 2023 Jul 5]. ISB1523: Anonymisation Standard for Publishing Health and Social Care Data. Available from: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1523-anonymisation-standard-for-publishing-health-and-social-care-data
6. UK Department of Health and Social Care. GOV.UK. 2022 [cited 2023 Jul 5]. [Withdrawn] Coronavirus (COVID-19): notice under regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 – general. Available from: https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information/coronavirus-covid-19-notice-under-regulation-34-of-the-health-service-control-of-patient-information-regulations-2002-general--2
7. NHS Digital. NHS Digital. 2022 [cited 2023 Jul 5]. Secretary of State for Health and Social Care: COVID-19 Public Health Directions 2020. Available from: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notices/secretary-of-state-directions/covid-19-public-health-directions-2020
8. NHS Health Research Authority. Health Research Authority. [cited 2023 Jul 5]. Confidentiality Advisory Group. Available from: https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/