

A statistical method for evaluating vaccine-induced immune correlates of protection against infection and disease progression: application to the ChAdOx1-S nCoV-19 phase 3 trial

Lucy R. Williams^a, Merryn Voysey^{b,c}, Andrew J. Pollard^{b,c}, Nicholas C. Grassly^{a,*}

^a Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, London, United Kingdom

^b Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

^c NIHR Oxford Biomedical Research Centre, Oxford, United Kingdom

ARTICLE INFO

Keywords:

Immune correlate of protection

COVID-19

Vaccine

Clinical trial

Statistical method

Bayesian statistics

ABSTRACT

Background: Correlates of protection (CoPs), defined as immune markers statistically correlated with vaccine efficacy (VE), can be used to accelerate vaccine development. Different components of the immune response may be important for protection against infection and against progression from asymptomatic infection to symptomatic or severe disease. However, CoPs are typically evaluated for these outcomes separately, which can lead to some CoPs not being identified. We propose a novel statistical framework for the integrated evaluation of CoPs for infections with multiple potential outcomes.

Methods: We developed a model of the natural history of an infection that can identify CoPs at each stage of infection and disease progression and implemented this model in a Bayesian estimation framework. We validated the model on simulated data then applied it to individual-level clinical and serum neutralising and binding antibody data from COV002 (NCT04400838), a phase II/III trial of the ChAdOx1 nCoV-19 (AZD1222) vaccine. We explored logistic and non-parametric (cubic spline) relationships between VE and the candidate CoPs.

Results: Both parametric and non-parametric forms of the model accurately estimated the relationships between the immune CoP and VE against infection (VE_{in}) and against progression to symptoms given infection (VE_{pr}) in 1000 simulated trial datasets. In the COV002 correlates subset (2227 participants, 5315 samples), SARS-CoV-2 spike-specific IgG was positively associated with both VE_{in} and VE_{pr} (average proportion of VE mediated by spike-specific IgG, 27 % (95 % CI 2–88 %) for VE_{in} and 41 % (95 % CI 0–96 %) for VE_{pr}). Pseudoneutralisation antibody titres and receptor binding domain (RBD) specific serum IgG showed similar correlations.

Conclusion: Integrated analysis of multiple disease outcomes and candidate CoPs enables the identification of CoPs that operate at different stages of disease progression, which are missed when evaluating outcomes separately.

1. Introduction

A central goal of vaccine development is the identification of immune markers that can be used to predict vaccine efficacy (VE) [1]. These correlates of protection (CoPs) can help to accelerate vaccine development by allowing early selection of candidates and, in some cases, vaccine licensure without the need for costly and large phase III efficacy trials [2]. Plotkin and Gilbert (2012) defined CoPs as immune markers that statistically correlate with protection after vaccination, and differentiate between CoPs that play a causal or non-causal role in

protection [3]. A mechanistic CoP (mCoP) is itself an effector of protection; it lies on the causal pathway between vaccination and the risk of the clinical outcome, while a non-mechanistic CoP (nCoP) does not itself induce protection, but it can be used to predict protection through its correlation with the (possibly unmeasured) immune effector.

Since the first formalisation of CoP classification by Prentice in 1989 [4], a drive to accelerate vaccine development has prompted increased interest in the use of CoPs and their validation methods [5]. Gilbert et al. [6] discussed the theoretical basis, assumptions, and applications of the state-of-the-art approaches in a recent review. Of note, they outline two

* Corresponding author at: Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, 90 Wood Lane, London, United Kingdom W12 0BZ.

E-mail address: n.grassly@imperial.ac.uk (N.C. Grassly).

<https://doi.org/10.1016/j.vaccine.2025.127856>

Received 30 July 2025; Received in revised form 8 October 2025; Accepted 9 October 2025

Available online 15 October 2025

0264-410X/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

distinct statistical frameworks that were widely applied in SARS-CoV-2 vaccine trials to elucidate the role of binding and neutralising antibodies in protection against COVID-19: “controlled VE” (cVE) and “mediation of VE” [7–11].

The cVE framework is based on the controlled effects framework of causal inference [12,13]. Introduced by Gilbert et al. (2022) [14], this approach provides an estimate for VE at different values of an immune marker if all trial participants were assigned to the vaccine arm at those values. It relies on a large dataset and on the assumptions that i) all confounders of the effect of the immune marker on the clinical outcome are controlled for; ii) all participants are capable of exhibiting each value of the immune marker.

Mediation of VE is a causal inference approach advocated by Cowling et al. (2019) [15], that is based on the natural effects framework of causal inference [16]. It aims to estimate causal rather than associational relationships between immune markers and clinical outcomes, by estimating the proportion of VE that is mediated by the CoP. This approach requires the same assumptions as the controlled effects approach, and additionally that iii) vaccination does not alter the relationship between the immune marker and the clinical outcome; iv) there is sufficient overlap in immune marker values across trial arms to be able to estimate VE not mediated by the immune marker (i.e. we can estimate the relative risk of the outcome of interest for placebo and vaccinated individuals at the same level of the immune marker).

Infection typically results in a range of clinical outcomes, reflecting disease natural history (e.g., varying severity) or different disease presentations (e.g., symptom groupings). Clinical trials collect data on many of these outcomes to not only define the primary outcome but also to gain a thorough understanding of a vaccine candidate’s impact and to enhance statistical power [17]. When a vaccine protects against disease progression in addition to infection, VE will vary depending on the clinical outcome of interest. For example, COVID-19 vaccines provided greater protection against severe clinical outcomes such as hospitalisation and death than milder outcomes like asymptomatic infection [18].

Identifying CoPs in the context of infections with multiple potential outcomes can be challenging, since different components of the immune system may vary in their importance for protection against infection or disease progression [19]. Additionally, intermediate outcomes may represent either a failure of a vaccine to protect low-risk individuals from infection, or successful protection against progression to more severe outcomes in high-risk individuals [20,21]. Typically, CoPs for each clinical outcome are evaluated separately; however, without explicitly modelling disease progression and its covariates, identification of immune CoPs against intermediate outcomes such as asymptomatic infection may not be possible. This was the case for the ChAdOx1-S nCoV-19 vaccine, where neutralising antibodies were identified as a CoP for symptomatic, but not asymptomatic infection [7]. The identification CoPs at each stage in the natural history of an infection would give a comprehensive understanding of vaccine protection, and could therefore be valuable for informing regulatory, vaccine development and policy decisions.

We propose a novel approach for the evaluation of CoPs for infections with multiple outcomes that explicitly models infection and disease progression [22]. We use a Bayesian approach to estimate cVE as a function of a candidate CoP, implemented as either a parametric or a non-parametric model. With additional assumptions, we take a natural effects (‘mediation of VE’) approach to estimate the proportion of VE that is mediated by each immune marker. We demonstrate both models on simulated datasets representing a range of potential scenarios, then apply them to data from a phase III trial of the ChAdOx1-S COVID-19 vaccine.

2. Methods

2.1. Conceptual framework

Our analysis framework is based on the controlled effects approach for CoP evaluation [14]. Statistical modelling is used to contrast the risk of infection and progression in the vaccine arm to that in the placebo arm as a function of an immune marker. A cVE that increases with values of the immune marker indicates a good CoP, while a flat cVE curve indicates no relationship. In the causal inference literature, the role of the immune marker in vaccine-induced protection is known as its ‘indirect effect’, as the vaccine indirectly operates through this mediator [23]. All other unmeasured immune functions that contribute to protection are known as the ‘direct effects’. This terminology is potentially confusing when used in the vaccinology field due to existing definitions of direct VE (individual protection) and indirect VE (herd protection). In this paper, we therefore use the terms ‘marker-mediated’ for indirect VE and ‘non-marker-mediated’ for direct VE as defined in the causal inference literature. This terminology has been used previously in the natural effects framework of causal inference [6]. We refer to CoPs generally as either mechanistic or non-mechanistic.

Our framework differs to those previously proposed in the explicit modelling of the natural history of infection and disease, and the evaluation of candidate CoPs at multiple stages in the progression of an infection [6]. We illustrate this framework with the simple case of an infection that can present symptomatically or asymptotically (Fig. 1). We model protection against infection and against progression to symptoms as independent functions of vaccination, indirectly through correlation with a candidate CoP and directly through other unmeasured immune functions. This allows for the differing contributions of immune effectors at each stage. In a population with previous pathogen or vaccine exposure, immune measurements in the placebo arm may be non-zero and correlate with risk of infection or disease. In the application presented here, we assume that the immune marker is zero (or below the lower limit of quantification, LLOQ) in the placebo arm. This assumption can be relaxed; there just needs to be overlap in the immune marker values between the vaccine and placebo arm to allow estimation of protection provided by the vaccine through immune effectors other than the measured CoP (the non-marker-mediated VE; assumption iv above) [14]. If we also assume that there is no interaction between the marker-mediated and non-marker mediated VE (assumption iii above), we can take a natural effects approach to estimate the proportion of overall VE that is mediated by the CoP.

2.2. Model implementation

We adapted a model previously developed for the analysis of VE against multiple outcomes to allow for the analysis of CoPs [22]. The model is a Bayesian hierarchical generalized regression model, illustrated here for a clinical trial of a vaccine against an infection that can present symptomatically or asymptotically (but which can be extended to multiple disease stages; Supplementary Methods). We assume that the number of infections experienced by individual i in a given time period (e.g., length of follow-up) follows a Poisson distribution,

$$Z_i \sim \text{Poisson}(\Lambda_i) \quad (1)$$

with mean $\Lambda_i = E(Z_i)$ depending on their vaccination status, duration of follow-up (t_i), hazard of infection in the population during the follow-up period (in this example, assumed to be a constant α) and other potential covariates (x_i , e.g. age), such that in the placebo arm,

$$\log(\Lambda_i) = \alpha + \sum_{j=1}^m \beta_j x_{ij} + \log(t_i) \quad (2)$$

where β_j is the regression coefficients for variable j and m is the number

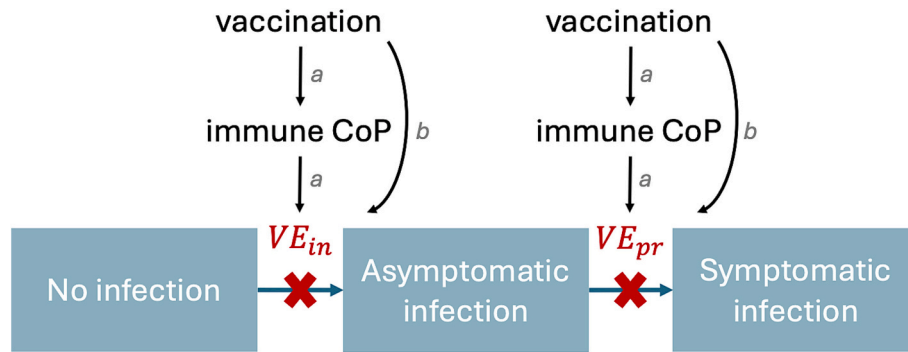


Fig. 1. Flow diagram of conceptual framework of the immune correlates of protection (CoP) model. Immune marker-mediated protection is shown with arrows labelled “a” and non-marker-mediated protection is shown in arrows labelled “b”. VE_{in} = vaccine efficacy against infection, VE_{pr} = vaccine efficacy against progression to symptoms.

of covariates. In a vaccinated individual, the cumulative hazard of infection is reduced by the vaccine induced protection against infection

$$\log(\Lambda_i) = \alpha + \sum_{j=1}^m \beta_j x_i + \log(1 - VE_{in}(y_i)) + \log(t_i) \quad (3)$$

where $VE_{in}(y_i)$ represents the VE against infection for that individual's level of the immune marker y_i .

The probability of symptoms in an infected individual i (θ_i) is estimated in the same Bayesian regression using a logistic link function and also can depend on vaccination status and other covariates (x_i). In the placebo arm, the probability of symptoms in those infected is estimated using logistic regression,

$$\text{logit}(\theta_i) = \varepsilon + \sum_{j=1}^m \kappa_j x_i \quad (4)$$

where ε is the log odds of symptom development, and κ_j is the regression coefficient for covariate j . In the vaccine arm, the probability of symptoms is reduced by the VE against progression, $VE_{pr}(y_i)$. VE_{in} and VE_{pr} are estimated as a function of the candidate CoP either parametrically or non-parametrically. In both models we assume VE_{in} and VE_{pr} range between 0 % and 100 %.

Average VE against symptomatic infection (VE_{sym}) and VE against asymptomatic infection (VE_{asym}) can be calculated from $VE_{in}(y_i)$ and $VE_{pr}(y_i)$ averaged across the vaccinated population [20] (Supplementary Methods).

We estimated model parameters for simulated and actual datasets by fitting a Poisson regression model using a Hamiltonian Markov Chain Monte Carlo (MCMC) algorithm in the Rstan package [24] in the R statistical environment [25]. Other regression models such as logistic or Cox proportional hazards regression could also be used but we do not explore that further here. We use MCMC to sample each parameter's posterior distribution and estimate the posterior distribution for VE across a range of CoP values. We take the mean and 2.5th and 97.5th percentiles of the distributions to give a point estimate and 95 % credible intervals (CrI). We ran all analyses with 4 chains, a warmup period of 1000 iterations and a total of 2000 iterations. Convergence was diagnosed based on the $\hat{R}$ convergence diagnostic (all $\hat{R}$ values were below 1.05 for all parameters) and trace plots that indicated good mixing. Code is available online [26].

2.3. Parametric VE estimation

We define total VE against infection (or disease progression) at a given level of the immune marker, y , as the sum of the marker-mediated ($M(y)$) and non-marker-mediated (N) VE:

$$VE(y) = M(y) + N \quad (5)$$

The marker-mediated VE is modelled as a logistic function of the CoP, scaled by the maximum total VE, σ .

$$M(y) = \frac{\sigma}{1 + e^{-k(y-y_{50})}} \quad (6)$$

where y_{50} is the measure of CoP at which 50 % of the maximum marker-mediated protection is provided and k is the slope of the relationship between VE and the candidate CoP]. The non-marker-mediated VE is given by $N = \sigma c$, where c is a constant to be estimated. In this application, we assume that c is the same for all vaccinated individuals, and that $M(y)$ depends only on the candidate CoP y . However, the model can be extended to allow these parameters to depend on other covariates (e. g. age).

We used non-informative, uniform priors on all parameters. Regression parameters were allowed to vary between -Inf and Inf, σ was bounded between 0 % and 100 %, y_{50} was restricted between the upper and lower bounds of the measured CoP distribution, and k restricted between 0 and 50 to aid model convergence, meaning that we assumed either a null or a positive relationship between the candidate CoP and cVE.

2.4. Non-parametric VE estimation

We also implemented the model of total VE against infection and against disease as (independent) non-parametric cubic spline regressions on the immune measurement with a logit link function. We assumed all individuals showed the same relationship between the candidate CoP and total VE. Knots were placed evenly along the distribution of the CoP in the vaccine arm, with one knot at the maximum and minimum values (Supplementary Methods). We used normal priors on the cubic spline average slope (mean = -5, standard deviation = 2) and basis weights (mean = 0, standard deviation = 1) which equates to a prior assumption for VE of 0 % across CoP values after transformation with the inverse logit. Consistent with the parametric model, we assumed total VE can range between 0 % and 100 %.

2.5. Application to simulated data

To demonstrate the applicability of the models to a range of disease and vaccine characteristics, we applied them to four simulated datasets. All datasets had a sample size of 10,000 individuals, with 1:1 allocation to vaccine or placebo. Participants were followed up for one year and censored after infection. For all participants, the force of infection, $\Lambda_i = 30\%$ per year before accounting for VE_{in} , and probability of symptoms, $\theta_i = 0.5$. We simulated a normally distributed $\log_{10}$ antibody distribution in the placebo arm with a mean of -8 and standard deviation of 1. A

fifth of participants in the vaccine arm were assumed to be non-responders with the same antibody distribution as the placebo arm. The remainder were given a $\log_{10}$ antibody distribution with a mean of 0 and standard deviation of 1.5 (Supplementary Fig. 1).

In scenario A ('baseline') we modelled a characteristic CoP that protects against both infection and progression to symptoms. We assumed the protection provided from unmeasured immune functions for VE_{pr} was greater than that for VE_{in} (non-marker-mediated $VE_{in} = 10\%$, non-marker-mediated $VE_{pr} = 30\%$). σ for VE_{in} was 70 %, while for VE_{pr} , $\sigma = 100\%$ protection was possible. The rate of increase of VE with the immune CoP was assumed to be equal for infection and progression ($k = 2$), but a lower antibody titre was necessary for some CoP-induced protection from symptoms than from infection ($y_{50} = 0$ for VE_{pr} and 2 for VE_{in}). All other scenarios were given the same parameter estimates except the following modifications. In scenario B ('all VE marker mediated'), all protection was mediated by the CoP and $\sigma = 100\%$ for VE_{in} and VE_{pr} . For scenario C ('no VE against progression'), $VE_{pr} = 0$ and $\sigma = 100\%$ for VE_{in} , and for scenario D ('no VE against infection'), $VE_{in} = 0$ and $\sigma = 100\%$ for VE_{pr} . VE_{sym} and VE_{asym} were calculated from the simulated VE_{in} and VE_{pr} . All parameters are summarised in Supplementary Table 1. Sensitivity analyses validating the approach for a smaller sample size and for case-cohort sampling are provided in the Supplementary Methods.

Both models were run on 1000 simulated datasets for each scenario and the relationship between the candidate CoP and VE_{in} and VE_{pr} was estimated across the range of CoP values for each simulation. To evaluate the performance of each estimate, we recorded the mean and 2.5th and 97.5th percentiles of the posterior mean estimated VE values, the coverage, and the precision at i) the assumed LLOQ for the CoP, ii) the median CoP value in the vaccinated responders, and iii) the 95th percentile of the CoP value in the vaccinated responders. Coverage is defined as the proportion of simulations that include the true value within the 95 % credible interval. Precision was calculated as the root mean square error (RMSE) across all simulations for the posterior mean estimated VE compared with the true VE.

2.6. Application to COV002

2.6.1. Description of COV002

The trial data used in this analysis comes from the University of Oxford sponsored phase 2/3 clinical trial, COV002 (NCT04400838). COV002 was a single-blind, multicentre, randomised phase 2/3 trial assessing the safety and efficacy of the SARS-CoV-2 ChAdOx1 nCoV-19 (AZD1222) vaccine. A detailed description of the trial and the protocol has been previously published [27–30]. The study was approved in the UK by the Medicines and Healthcare products Regulatory Agency (reference 21584/0428/001-0001) and the South-Central Berkshire Research Ethics Committee (reference 20/SC/0179). In this analysis we have clinical, socio-demographic, and immunogenicity data from May 28th 2020 until the February 25th, 2021 data cut.

The primary outcome was VE against PCR-confirmed symptomatic SARS-CoV-2 infection, detected through testing in response to reporting of symptoms consistent with COVID-19 (fever $\geq 37.8^\circ\text{C}$, cough, shortness of breath, anosmia or ageusia). VE against asymptomatic infection was estimated as a secondary outcome, and asymptomatic infections were detected with weekly nose and throat swabs conducted by trial participants at home.

Immunogenicity data were collected from a proportion of participants at 28 days post the second vaccine dose (post-boost +28 days, PB28). Here, we analyse i) Anti-SARS-CoV-2 spike IgG, ii) Anti-SARS-CoV-2 receptor binding domain (RBD) IgG and iii) pseudo SARS-CoV-2 spike neutralisation. Anti-SARS-CoV-2 spike and RBD IgG were measured by a multiplex immunoassay, as the number of binding antibody units per millilitre (BAU/ml), while antibody neutralisation was measured with a lentivirus-based pseudovirus expressing the SARS-CoV-

2 spike protein, as the neutralisation dilution for 50 % virus inhibition (ID_{50}). Full details of assay validation are given in [7]. All serum samples from symptomatic cases and 15 % of non-cases from the efficacy cohort were tested, where the sample volume allowed [7]. As in previous COV002 analysis [7], we assumed the missingness mechanism was missing at random and accounted for this using inverse probability weights (Supplementary Methods). We also accounted for variation in exposure to infection with covariates in the log-linear regression model for the number of infections (Supplementary Methods). Immune marker values were $\log_{10}$ -transformed prior to analysis. Values that were censored at the LLOQ were imputed with the value LLOQ/2.

We defined the correlates population as participants who met the eligibility criteria for the primary outcome efficacy analysis: seronegative at baseline, PB28 visit between 14 and 42 days after the second dose, follow-up for at least 7 days after PB28, no prior evidence of SARS-CoV-2 infection. Only participants who received two doses of either vaccine or control were included, regardless of their dosage (standard or low dose). As study participants were allowed to unblind once SARS-CoV-2 vaccines became publicly available, we censored after i) SARS-CoV-2 infection, ii) unblinding, or iii) February 25th, 2021, whichever was earliest.

In the parametric CoP models, we tested the null hypotheses that each candidate CoP was not associated with VE_{in} or VE_{pr} (i.e. slope parameter, $k = 0$) using the Savage-Dickey density ratio (SDDR) (Supplementary Methods) [31].

3. Results

3.1. Simulation results

The simulated relationships between the candidate CoP and VE are shown in the left-hand column of Fig. 2. VE_{in} and VE_{pr} were assumed to follow a scaled logistic relationship with the CoP, which implied a similar relationship for VE_{sym} but a non-monotonic relationship for VE_{asym} . In scenarios A (baseline) and B (all VE marker mediated), VE_{asym} was lowest at intermediate levels of the candidate CoP (i.e. immunity at this level mainly protects against disease progression but not against infection, leading to a negative VE against asymptomatic infection).

Both the parametric and non-parametric models accurately estimated the relationship between the simulated CoP and controlled VE in all four scenarios (illustrated in Fig. 2 for a single simulation; full results in Supplementary Tables 2 and 3). Across the 1000 simulations, the coverage of the VE credible intervals for the parametric model at the LLOQ, median and upper 95th percentile of the CoP in the vaccine arm responders ranged 93 %–98 % for Scenario A (baseline), 83 %–96 % for Scenario B (all VE marker mediated), 92 %–98 % for Scenario C (no VE against progression), and 93 %–98 % for Scenario D (no VE against infection) for VE_{in} and VE_{pr} (Supplementary Table 2). The lowest coverage estimates were observed when the true VE_{in} or VE_{pr} was close to the limits of its possible range (i.e. 0 % or 100 %). The non-parametric model accurately estimated VE where there were sufficient data but was biased towards zero at the limits of the vaccine arm CoP distribution where there were few data points, reflecting the prior distribution chosen for the spline function (Supplementary Table 3). Estimated parameter values for the parametric model are provided in Supplementary Table 4.

Sensitivity analysis validating the model on data with unrepresentative sampling (preferential sampling of cases, particularly those in the vaccine arm) with inverse probability weighting showed good convergence to the true values of VE for both the parametric and non-parametric models, while a smaller trial size ($N = 1000$) gave greater uncertainty in VE estimation (Supplementary Figs. 4 and 5).

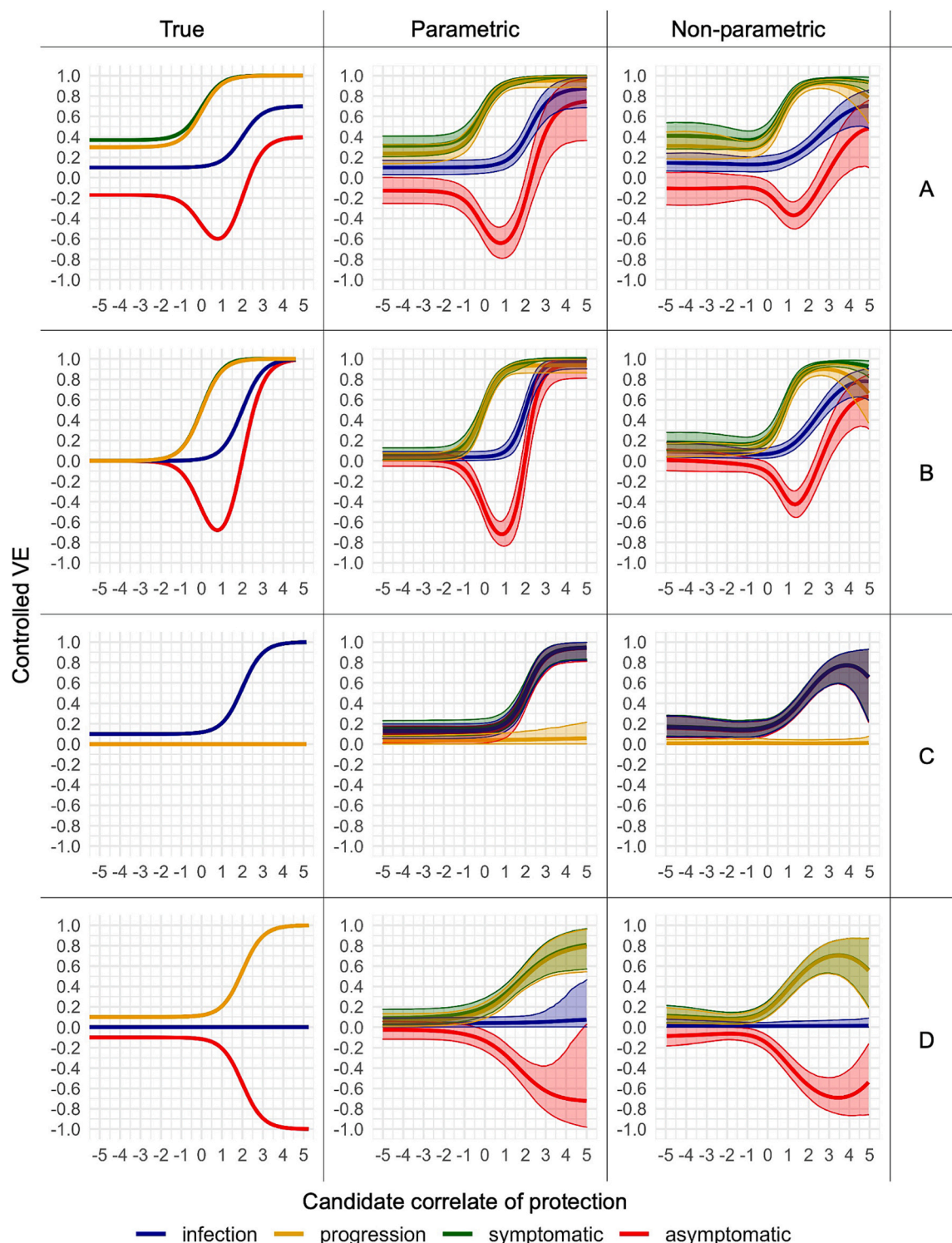


Fig. 2. Simulated (actual) and estimated relationships between an immune correlate of protection and controlled vaccine efficacy (VE). Rows show scenarios A - baseline, B - all VE marker mediated, C - no VE against progression and D - no VE against infection. Columns show the true simulated relationship, parametric, and non-parametric model estimates. Model-estimate figures show the mean and 95 % credible intervals of one randomly selected simulation. Model estimates for all 1000 simulations are summarised in Supplementary Tables 2 and 3.

3.2. COV002 results

8558 participants were eligible for the full analysis dataset and 2227 participants had a serology sample for at least one assay, making them eligible for the correlates subset. Among the eligible participants in the vaccine arm, 750 (17.6 %) had a neutralising antibody reading and 1239

(29.1 %) had a spike/RBD binding antibody reading. In the placebo arm, 431 (10.1 %) and 828 (19.4 %) had neutralising antibody and spike/RBD binding antibody readings, respectively. The baseline characteristics of the full analysis dataset and the correlates subset are provided in Supplementary Table 5. Antibody distributions at PB28 across the three candidate CoPs are shown in Supplementary Fig. 6.

All measured candidate CoPs showed an increase in controlled VE_{in} and VE_{pr} with increased concentration when using either the parametric or non-parametric model (Fig. 3 and Supplementary Fig. 7 respectively). In the parametric model, measured neutralising antibodies provided on average across the vaccine arm, 8 % (95 % CrI 0–37) protection against infection and 15 % (95 % CrI 0–39) protection against symptom development (marker-mediated VE; Table 1). Other (non-measured) vaccine-induced immunity provided on average a further 52 % (95 % CI 22–66) protection against infection and 23 % (95 % CrI 1–44) protection against symptom development (non-marker-mediated VE). This gave estimates of 13 % (95 % CrI 0–57) and 36 % (95 % CrI 0–94) for the proportion of protection against infection and progression mediated by

neutralising antibodies, respectively. Compared to neutralising antibodies, estimates for the proportion of protection mediated by measured spike and RBD IgG antibodies were slightly higher (Table 1). For spike IgG, 27 % (95 % CrI 2–88) and 41 % (95 % CrI 0–96) of the protection against infection and progression, respectively, was mediated by spike IgG, while for RBD IgG, these estimates were 24 % (95 % CrI 2–89) and 39 % (95 % CrI 0–96).

Between the LLOQ and the 95th percentile of the neutralising antibody distribution, the parametric model estimated an increase in VE_{in} from 58.4 % to 63.8 % and an increase in VE_{pr} from 30.2 % to 45.4 % (Table 2). At the spike IgG LLOQ, VE_{in} was estimated at 44.5 % and VE_{pr} at 28.0 %. Both were higher at the 95th percentile of the CoP, where VE_{in}

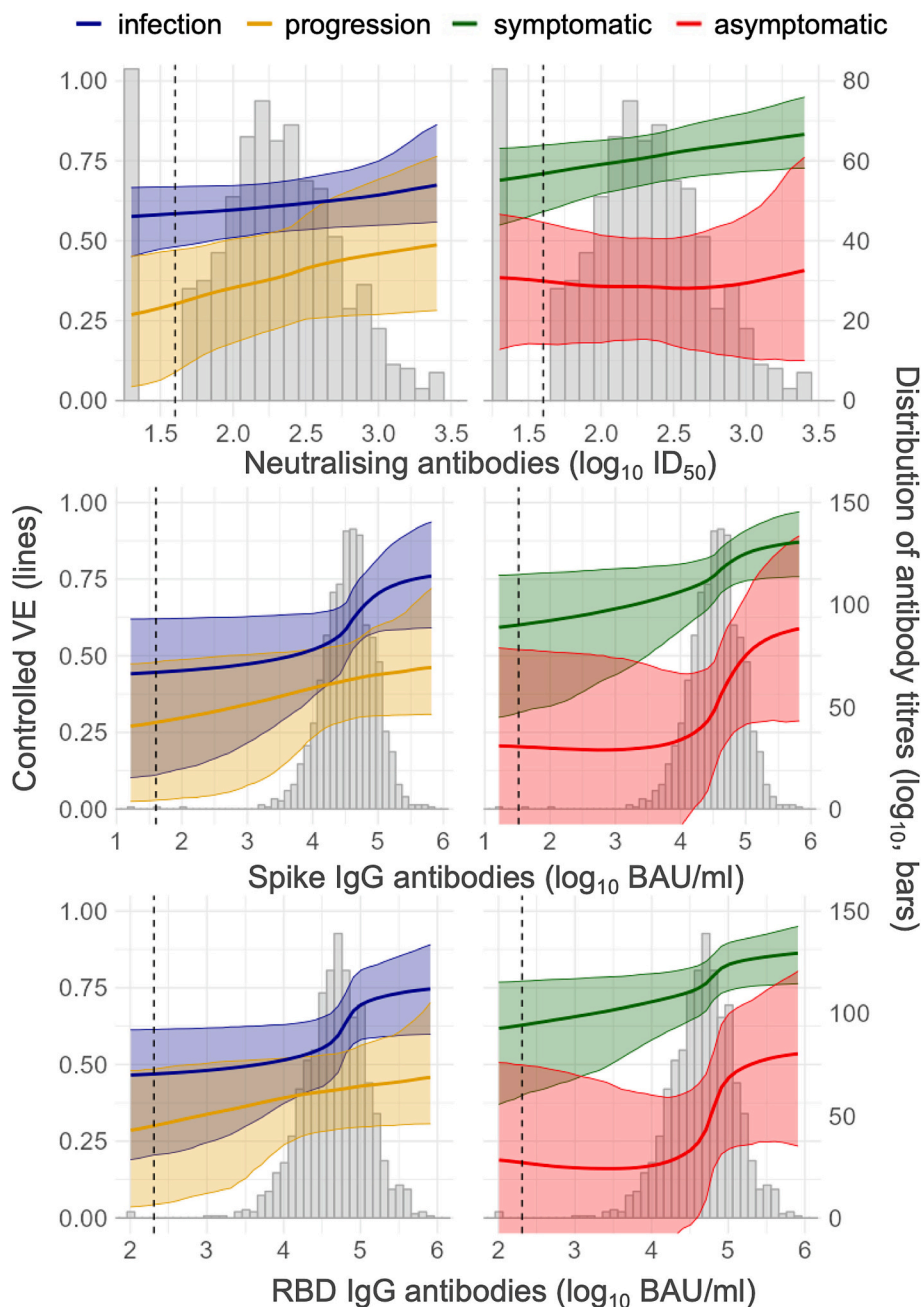


Fig. 3. Parametric model estimated controlled vaccine efficacy (VE_{in} and VE_{pr}) as a function of neutralising antibodies, spike IgG antibodies, and RBD IgG antibodies in the COV002 trial, alongside the implied VE_{asym} and VE_{sym} . Dashed line represents the lower limit of quantification (LLOQ) for each assay. Neutralising antibodies LLOQ = $1.60 \log_{10} ID_{50}$, spike IgG LLOQ = $1.52 \log_{10} BAU/ml$, RBD IgG LLOQ = $2.31 \log_{10} BAU/ml$. VE estimates are the mean (lines) and 95th percentiles (ribbons) of the distribution of VE estimates generated across antibody titre for each MCMC iteration. Bars represent the distribution of each candidate CoP in the vaccine arm. Values < LLOQ were given a value of LLOQ/2.

Table 1

Parametric model parameter estimates for vaccine efficacy against infection and disease progression as a function of neutralising antibodies, spike IgG antibodies, and RBD IgG antibodies in the COV002 trial. VE_{in} = vaccine efficacy against infection; VE_{pr} = vaccine efficacy against progression to symptoms.

VE Parameter	Mean VE, % (95 % CrI)	
	VE_{in}	VE_{pr}
Neutralising antibodies		
Non-marker-mediated	52 (22, 66)	23 (1, 44)
Marker-mediated	8 (0, 37)	15 (0, 39)
Proportion marker-mediated	13 (0, 57)	36 (0, 94)
Average ¹	61 (53, 68)	37 (23, 50)
Maximum	75 (58, 98)	54 (30, 93)
Spike IgG antibodies		
Non-marker-mediated	43 (5, 62)	25 (1, 46)
Marker-mediated	17 (1, 54)	17 (0, 43)
Proportion marker-mediated	27 (2, 88)	41 (0, 96)
Average ¹	60 (52, 67)	42 (29, 53)
Maximum	79 (60, 99)	49 (31, 84)
RBD IgG antibodies		
Non-marker-mediated	45 (11, 61)	25 (2, 47)
Marker-mediated	15 (2, 49)	16 (0, 42)
Proportion marker-mediated	24 (2, 89)	39 (0, 96)
Average ¹	60 (52, 67)	41 (29, 53)
Maximum	77 (61, 97)	49 (31, 84)

¹ Mean VE across trial population; CrI = credible interval.

Table 2

Parametric and non-parametric model estimates for vaccine efficacy against infection and against progression to symptoms at given antibody titres in COV002. LLOQ = lower limit of quantification, VE_{in} = vaccine efficacy against infection, VE_{pr} = vaccine efficacy against progression to symptoms. Neutralising antibodies LLOQ = 1.60 log₁₀ ID₅₀ and 95th percentile = 2.93 log₁₀ ID₅₀, spike IgG LLOQ = 1.52 log₁₀ BAU/ml and 95th percentile = 5.08 log₁₀ BAU/ml, RBD IgG LLOQ = 2.31 log₁₀ BAU/ml and 95th percentile = 5.54 log₁₀ BAU/ml.

	Parametric estimated VE (% 95 % CrI)		Non-parametric estimated VE (%, 95 % CrI)	
	VE_{in}	VE_{pr}	VE_{in}	VE_{pr}
Neutralising antibodies				
LLOQ	58.4 (48.1–66.9)	30.2 (8.8–47.1)	56.8 (42.5–68.2)	29.0 (12.6–48.5)
95th percentile	63.8 (54.6–73.7)	45.4 (26.8–68.0)	65.5 (48.1–78.3)	38.6 (15.2–65.0)
Spike IgG antibodies				
LLOQ	44.5 (10.9–62.0)	28.0 (2.7–47.7)	52.5 (14.2–87.7)	30.0 (5.1–71.0)
95th percentile	71.4 (58.3–83.7)	44.1 (30.4–60.7)	72.5 (60.6–82.1)	42.5 (23.7–61.9)
RBD IgG antibodies				
LLOQ	46.9 (20.5–61.4)	30.0 (4.3–48.5)	63.1 (25.5–90.1)	31.9 (5.8–72.6)
95th percentile	70.4 (58.5–81.6)	43.2 (29.7–56.9)	72.9 (60.2–83.3)	36.4 (13.8–61.3)

CrI = credible interval.

was estimated at 71.4 % and VE_{pr} at 44.1 %. For RBD IgG, VE_{in} and VE_{pr} increased from 46.9 % to 70.4 % and 30.0 % to 43.2 % between the LLOQ and 95th percentile, respectively. Estimates from the non-parametric model were consistent with the parametric model but tended towards $VE = 0$ % where data were sparse as a result of the choice of priors (Table 2; Supplementary Fig. 7).

Hypothesis testing with the SDDR on the parametric model slope parameters found ‘moderate’ evidence for a relationship between neutralising antibodies and VE_{pr} (SDDR = 3.77) and ‘anecdotal’ evidence with VE_{in} (SDDR = 2.15). For spike and RBD IgG, the evidence was ‘moderate’ for both VE_{in} and VE_{pr} (SDDR spike VE_{in} = 4.03, spike VE_{pr} = 3.61, RBD VE_{in} = 5.07, VE_{pr} = 3.40).

4. Discussion

In this paper we demonstrated how a novel method for analysing candidate CoPs for infections with multiple outcomes can characterise their roles at each stage in the natural history of an infection. We validated parametric and non-parametric forms of the model using simulated data for an infection with two potential outcomes (symptomatic or asymptomatic), then estimated the roles of three candidate CoPs for protection against infection and symptom development for the ChAdOx1 nCoV-19 vaccine.

Simulations showed that non-monotonic relationships between VE_{asym} and a candidate CoP can arise when a vaccine protects against both infection and symptom development. VE_{asym} can be negative when a greater proportion of vaccinated individuals are protected from developing symptoms rather than from infection, such that vaccinated individuals are more likely to experience an asymptomatic infection than placebo recipients. This can lead to complex relationships between the CoP and VE against intermediate outcomes in the disease progression pathway, such as asymptomatic infection, when the CoP is correlated with both VE_{in} and VE_{pr} . For example, as seen in scenarios A (baseline) and B (all VE marker mediated), if protection against progression to symptoms occurs at lower levels of the CoP than protection against infection, VE_{asym} will decrease as the CoP and VE_{pr} increase. Once the CoP reaches high enough levels to protect against infection, the rate of decline in VE_{asym} will decrease and may ultimately increase. In these cases where a CoP contributes to protection from infection and progression, traditional separate analysis of symptomatic and asymptomatic outcomes may identify the CoP for VE against symptomatic but not asymptomatic infection. This was the case in the original analysis of the COV002 antibody data, where no relationship between immune markers and VE_{asym} were found [7]. Here we show that spike and RBD IgG and neutralising antibodies provide a degree of protection against both infection and progression to symptoms – although other immune functions (e.g. T cells) likely also play a role (proportion of protection spike IgG-mediated for VE_{in} and VE_{pr} was 27 %, and 41 % respectively). By describing the process of disease progression, we are able to characterise CoPs that would otherwise not be identified.

Our model showed that a greater proportion of protection against infection was mediated by spike and RBD binding antibodies measured using multiplex immunoassays than measured in a pseudoneutralisation antibody assay. Much of the COVID-19 literature has focussed on the role of neutralising antibodies as a CoP [32–34]. This is reasonable as they are expected to provide a mechanistic role in protection against invasive disease by preventing pathogen entry to the cell, and non-human primate studies have demonstrated their role as a mechanistic CoP [34,35]. The greater correlation with binding antibodies may simply reflect the sensitivity and dynamic range of these assays relative to the neutralisation assay. However, the role of binding antibodies in alerting the immune system to initiate other immune effectors should not be overlooked. Multiple studies have demonstrated statistical correlation between protection from COVID-19 and binding antibodies [11,36,37] and they tend to be easier and cheaper to measure. Validated, repeatable assays for spike and RBD binding antibodies could therefore be useful as mechanistic and non-mechanistic CoPs. While other immune functions such as mucosal antibodies, non-neutralising antibody functions and cellular immunity are expected to be important for protection from infection and severe outcomes, there were insufficient data to evaluate these in COV002 at the time of our analysis.

The five infectious diseases that cause the most mortality globally all exhibit a range of severity of disease outcomes [38]. For vaccines in development that may target different stages in the progression of these infections, our analysis framework could be useful for understanding at which stage, and in what way, the vaccine operates. For example, our understanding of the importance of different immune effectors at each stage in the progression of tuberculosis (TB) is incomplete, and the

currently licensed live-attenuated Bacille Calmette-Guérin (BCG) vaccine protects young children against the most severe forms of tuberculosis but offers little protection against pulmonary tuberculosis for adolescents and adults [39]. The identification of CoPs against these different outcomes would facilitate the rational design, optimisation, and evaluation of TB vaccine candidates currently in development, and our integrated analytical approach could be used to simultaneously and efficiently identify CoPs for each stage in the progression from infection to severe disease. Additionally, given the potential population-level impact of vaccines that protect against infection and transmission, interest in mucosal vaccines that prevent infection with the organisms causing diseases such as TB, COVID-19 and RSV disease has heightened. Again, our approach would be useful for disentangling the roles of different immune components at each stage of the infection.

While both the parametric and non-parametric models of the relationship between VE and candidate CoPs perform well when there are sufficient data, the non-parametric model performs less well when data are sparse. In the non-parametric model, we used priors that favoured a VE of zero, such that any observed VE was informed by a true protective effect. The decline in estimated VE at the bounds of the antibody distribution seen in the non-parametric COV002 results should be interpreted in light of this prior probability. Both models also show greater bias when the true VE is close to zero or one. This is a result of using the inverse logit transformation in the estimation of VE, as VE on the logit scale is allowed to vary between negative and positive infinity, the MCMC rarely samples the edges of the possible ranges. Despite this, both models produce estimates close to the true values and accurately characterise the simulated relationships across the CoP distribution. Finally, a limitation of our analysis is that VE_{pr} estimation is informed solely by the infected population, which we assume is comparable across trial arms. However, study arm is not randomised among infected individuals and a sampling bias can be introduced if VE_{in} or the risk of infection vary across the same variables that influence VE_{pr} . If these variables are known, it is possible to minimise this bias by controlling for covariates in the regression models for each of these model parameters (e.g. including age as a covariate with the risk of infection and VE_{pr}).

In conclusion, independent analysis of primary and secondary outcomes in a clinical trial when estimating CoPs is liable to miss those CoPs that contribute to protection against intermediate stages in the natural history of an infection. Our integrated approach facilitates the characterization of each candidate CoP at the stage of disease progression at which they act. This could be particularly important for mucosal vaccines that target the prevention of infection and for vaccines against diseases with a range of severity.

Contributors

LRW and NCG conceptualised the study and planned the statistical analysis. MV and AJP contributed to trial implementation, data collection, statistical analysis and concept development. LRW performed the formal analysis and data visualisation and wrote the original draft. NCG supervised the study and edited the manuscript. All authors critically reviewed the manuscript and approved the decision to submit for publication.

Data sharing

Anonymised participant data is available upon request directed to the alternative corresponding author. Proposals will be reviewed and approved by the sponsor, investigator, and collaborators on the basis of scientific merit. After approval of a proposal, data can be shared through a secure online platform after signing a data access agreement. All data will be made available for a minimum of 5 years from the end of the trial.

CRedit authorship contribution statement

Lucy R. Williams: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Merryn Voysey:** Writing – review & editing, Supervision, Methodology, Data curation, Conceptualization. **Andrew J. Pollard:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Nicholas C. Grassly:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nicholas Grassly reports financial support was provided by UK Research and Innovation Medical Research Council. Andrew Pollard and Merryn Voysey report financial support was provided by National Institute for Health and Care Research. Lucy Williams reports financial support was provided by UK Research and Innovation Medical Research Council and Imperial College London. Andrew Pollard and Merryn Voysey report financial support was provided by UK Research and Innovation. Nicholas Grassly is a member of the UK Joint Committee on Vaccination and Immunisation. Andrew Pollard has received grant funding from, Gates Foundation, Wellcome Trust, CEPI, AstraZeneca, the European Commission and Serum Institute of India in the past 36 months. Andrew Pollard is a contributor to intellectual property licensed by Oxford University Innovation to AstraZeneca. Andrew Pollard has undertaken consultancy for Shionogi and the Oxford Ellison Institute. Andrew Pollard is Chair of the UK Joint Committee on Vaccination and Immunisation, was a Member of WHO's SAGE until 2022 and Chair of WHO's Salmonella TAG until end of 2024 and remains a current member.

Acknowledgements

The COV002 study was funded by UK Research and Innovation, National Institute for Health Research (NIHR), Coalition for Epidemic Preparedness Innovations, NIHR Oxford Biomedical Research Centre, Thames Valley and South Midland's NIHR Clinical Research Network, and AstraZeneca. This report is independent research funded by the UK NIHR and UK Research and Innovation.

This work was supported by payments made to the Imperial College London, MRC Centre for Global Infectious Disease Analysis (reference MR/X020258/1), funded by the UK Medical Research Council (MRC). This UK funded award is carried out in the frame of the Global Health EDCTP3 Joint Undertaking. This work was also supported by the MRC Doctoral Training Partnership (to LRW) and the Imperial College President's Scholarship (to LRW).

We are grateful to the NIHR infrastructure provided through the NIHR Biomedical Research Centres and the NIHR Clinical Research Network at the UK study sites. We are grateful for the support of the Department of Health and Social care and the Office for Life Sciences for swabbing kits and access to PCR testing and the dedication of the Lighthouse laboratory teams. The views expressed in this publication are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. AstraZeneca reviewed the data from the study and the final manuscript prior to submission, but the authors retained editorial control. We thank the volunteers who participated in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.vaccine.2025.127856>.

Data availability

Data will be made available on request.

References

- Plotkin SA. Correlates of protection induced by vaccination. *Clin Vaccine Immunol* 2010;17:1055–65.
- Edwards KM. Development, acceptance, and use of immunologic correlates of protection in monitoring the effectiveness of combination vaccines. *Clin Infect Dis* 2001;33(Suppl. 4):S274–7.
- Plotkin SA, Gilbert PB. Nomenclature for immune correlates of protection after vaccination. *Clin Infect Dis* 2012;54:1615–7.
- Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–40.
- Plotkin SA. Recent updates on correlates of vaccine-induced protection. *Front Immunol* 2022;13:1081107.
- Gilbert PB, Fong Y, Hejazi NS, Kenny A, Huang Y, Carone M, et al. Four statistical frameworks for assessing an immune correlate of protection (surrogate endpoint) from a randomized, controlled, vaccine efficacy trial. *Vaccine* 2024;42:2181–90.
- Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat Med* 2021;27:2032–40.
- Fong Y, McDermott AB, Benkeser D, Roels S, Stieh DJ, Vandebosch A, et al. Immune correlates analysis of the ENSEMBLE single Ad26. COV2. S dose vaccine efficacy clinical trial. *Nat Microbiol* 2022;7:1996–2010.
- Benkeser D, Fong Y, Janes HE, Kelly EJ, Hirsch I, Sproule S, et al. Immune correlates analysis of a phase 3 trial of the AZD1222 (ChAdOx1 nCoV-19) vaccine. *npj Vaccines* 2023;8:36.
- Benkeser D, Montefiori DC, McDermott AB, Fong Y, Janes HE, Deng W, et al. Comparing antibody assays as correlates of protection against COVID-19 in the COVE mRNA-1273 vaccine efficacy trial. *Sci Transl Med* 2023;15:eade9078.
- Gilbert PB, Montefiori DC, McDermott AB, Fong Y, Benkeser D, Deng W, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* 2022;375:43–50.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–55.
- Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics* 2009;65:530–8.
- Gilbert PB, Fong Y, Kenny A, Carone M. A controlled effects approach to assessing immune correlates of protection. *Biostatistics* 2023;24:850–65.
- Cowling BJ, Lim WW, Perera RAPM, Fang VJ, Leung GM, Peiris JSM, et al. Influenza Hemagglutination-inhibition antibody titer as a mediator of vaccine-induced protection for influenza B. *Clin Infect Dis* 2018;68:1713–7.
- VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17–32.
- Prunas O, Willemsen JE, Bont L, Pitzer VE, Warren JL, Weinberger DM. Incorporating data from multiple endpoints in the analysis of clinical trials: example from RSV vaccines. *Epidemiology* 2024;35:103–12.
- Wu N, Joyal-Desmarais K, Ribeiro PAB, Vieira AM, Stojanovic J, Sanuade C, et al. Long-term effectiveness of COVID-19 vaccines against infections, hospitalisations, and mortality in adults: findings from a rapid living systematic evidence synthesis and meta-analysis up to December, 2022. *Lancet Respir Med* 2023;11:439–52.
- Pollard AJ, Bijker EM. A guide to vaccinology: from basic principles to new developments. *Nat Rev Immunol* 2021;21:83–100.
- Williams LR, Ferguson NM, Donnelly CA, Grassly NC. Measuring vaccine efficacy against infection and disease in clinical trials: sources and magnitude of Bias in coronavirus disease 2019 (COVID-19) vaccine efficacy estimates. *Clin Infect Dis* 2022;75:e764–73.
- Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. *Lancet Infect Dis* 2021;21:e26–35.
- Williams LR, Voysey M, Pollard AJ, Grassly NC. A novel approach for estimating vaccine efficacy for infections with multiple outcomes: application to a COVID-19 vaccine trial. *medRxiv* 2024. 2023.03.02.23286698.
- Lim WW, Leung NHL, Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Distinguishing causation from correlation in the use of correlates of protection to evaluate and develop influenza vaccines. *Am J Epidemiol* 2020;189:185–92.
- Stan Development Team. RStan: the R interface to Stan. 2020.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2021.
- Williams L. Correlates of protection (2024). [GitHub repository](#); 2024.
- Barrett JR, Belij-Rammerstorfer S, Dold C, Ewer KJ, Folegatti PM, Gilbride C, et al. Phase 1/2 trial of SARS-CoV-2 vaccine ChAdOx1 nCoV-19 with a booster dose induces multifunctional antibody responses. *Nat Med* 2021;27:279–88.
- Folegatti PM, Ewer KJ, Aley PK, Angus B, Becker S, Belij-Rammerstorfer S, et al. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* 2020;396:467–78.
- Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet* 2021;396:1979–93.
- Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 2021;397:99–111.
- Wagenmakers EJ, Lodewyckx T, Kuriyal H, Grasman R. Bayesian hypothesis testing for psychologists: a tutorial on the savage-dickey method. *Cogn Psychol* 2010;60:158–89.
- Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med* 2021;27:1205–11.
- Cromer D, Steain M, Reynaldi A, Schlub TE, Khan SR, Sasson SC, et al. Predicting vaccine effectiveness against severe COVID-19 over time and against variants: a meta-analysis. *Nat Commun* 2023;14:1633.
- Gilbert PB, Donis RO, Koup RA, Fong Y, Plotkin SA, Follmann D. A Covid-19 milestone attained - a correlate of protection for vaccines. *N Engl J Med* 2022;387:2203–6.
- McMahan K, Yu J, Mercado NB, Loos C, Tostanoski LH, Chandrashekar A, et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* 2021;590:630–4.
- Regev-Yochay G, Lustig Y, Joseph G, Gilboa M, Barda N, Gens I, et al. Correlates of protection against COVID-19 infection and intensity of symptomatic disease in vaccinated individuals exposed to SARS-CoV-2 in households in Israel (ICoFS): a prospective cohort study. *Lancet Microbe* 2023;4:e309–18.
- Hertz T, Levy S, Ostrovsky D, Oppenheimer H, Zismanov S, Kuzmina A, et al. Correlates of protection for booster doses of the SARS-CoV-2 vaccine BNT162b2. *Nat Commun* 2023;14:4575.
- Naghavi M, Ong KL, Aali A, Ababneh HS, Abate YH, Abbafati C, et al. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Lancet* 2024;403:2100–32.
- Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 1994;271:698–702.