Open Forum Infectious Diseases

MAJOR ARTICLE

Correlates of Protection Against Symptomatic COVID-19: The CORSER 5 Case–Control Study

Léopold Beeker,[1,2] Thomas Obadia,[1,3] Emma Bloch,[1] Laura Garcia,[1] Manon Le Fol,[1] Tiffany Charmet,[2] Laurence Arowas,[4] Rémy Artus,[5] Olivia Cheny,[6] Dorian Cheval,[5] Yanis Dahoumane,[6] Maurine Delhaye,[4] Delal Ergen,[5] Mariem Essaidani,[5] Christine Fanaud,[4] Sandrine Fernandes Pellerin,[6] Nathalie Jolly,[6] Hélène Laude,[4] Emmanuel Roux,[5] Marine Samson,[5] Linda Sangari,[5] Marie-Noëlle Ungeheuer,[4] Sophie Vacant,[5] Ayla Zayoud,[4] Françoise Donnadieu,[1] Stéphane Pelleau,[1] Simon Galmiche,[2] Arnaud Fontanet,[2,7] and Michael White[1]

[1]Infectious Disease Epidemiology and Analytics G5 Unit, Department of Global Health, Institut Pasteur, Université Paris-Cité, Paris, France, [2]Emerging Diseases Epidemiology Unit, Department of Global Health, Institut Pasteur, Université Paris-Cité, Paris, France, [3]Hub de Bioinformatique et Biostatistique, Institut Pasteur, Université Paris-Cité, Paris, France, [4]Plateforme d'Investigation Clinique (ICAReB-Clin), Institut Pasteur, Université Paris-Cité, Paris, France, [5]Biobanque de Collections de Ressources Biologiques d'Origine Humaine (ICAReB-Biobank), Center for Translational Science, Institut Pasteur, Université Paris-Cité, Paris, France, [6]Pôle de Coordination de la Recherche Clinique (PC-RC), Institut Pasteur, Université Paris-Cité, Paris, France, and [7]PACRI, Risques Infectieux et Emergents, Conservatoire National des Arts et Métiers, Paris, France

Background. Establishing correlates of protection often requires large cohorts. A rapid and adaptable case–control study design can be used to identify antibody correlates of protection against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in serum and saliva.

Methods. We designed a case–control study to compare antibody levels between cases of SARS-CoV-2 infection within 5 days of symptom onset and uninfected controls. Controls were matched on age, number of coronavirus disease 2019 vaccine doses, time since last dose, and past episodes of infection. We quantified anti-SARS-CoV-2 and seasonal coronavirus immunoglobulin (Ig) G in serum and saliva at inclusion, 1 month, and 6 months.

Results. We included 90 cases and 62 controls between February and September 2022. A boost and decay pattern of serum antibodies was observed in cases at 1 and 6 months, respectively, but not in controls. Anti-SARS-CoV-2 antibody levels were significantly higher in controls at inclusion both in serum (particularly antinucleocapsid IgG: 4.14 times higher compared with cases; 95% CI, 2.46–6.96) and saliva (particularly antispike for Delta variant IgG: 4.89 times higher compared with cases; 95% CI, 2.91–9.89). Saliva antibodies generally outperformed serum antibodies for case/control differentiation.

Conclusions. In this case–control study, we provided evidence of correlates of protection of anti-SARS-CoV-2 IgG in saliva and serum, with saliva antibodies often outperforming serum. The finding that antibodies in saliva are a better correlate of protection than antibodies in serum may inform vaccine development by highlighting the importance of robust induction of mucosal immune responses. This study design may be used during future epidemics for the prompt assessment of correlates of protection.

Keywords. antibody; case–control study; COVID-19; protection; saliva; serum.

Establishing correlates of protection is essential to the design of vaccine policies and to the study of the effectiveness of new vaccines that cannot be evaluated against placebo when an effective and indicated alternative exists. Correlates of protection against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection have been identified for anti-SARS-CoV-2 antibodies measured in blood samples, mainly in cohort studies [1, 2]. These often require prospective data with samples available at the beginning of the follow-up period. Such studies have been able to assess the evolution of antibody correlates of protection against the Omicron variant [3]. Case–control studies can also help define correlates of protection [4]. They offer cheaper and quicker alternatives to cohort studies but require either the existence of preexisting samples or inclusion immediately after exposure to SARS-CoV-2 for cases before any potential boost in immunity by the infection. In the very early days of infection (<6 days), the boosting of the immune response remains limited, potentially allowing the analysis of preexisting immunity in the few days following symptom onset [5, 6].

This study is part of the CORSER studies led by Institut Pasteur: in its capacity as an overarching study, CORSER (CORonavirus SERo-epidemiology) plays a pivotal role in coordinating and harmonizing various sero-epidemiological investigations conducted at Institut Pasteur. By unifying these diverse studies under a single framework, CORSER facilitates

Received 24 June 2024; editorial decision 31 December 2024; accepted 06 January 2025; published online 8 January 2025

Correspondence: Michael White, PhD, Infectious Disease Epidemiology & Analytics G5 Unit, Institut Pasteur, 25-28 Rue du Docteur Roux, 75015 Paris, France (michael.white@pasteur.fr); or Arnaud Fontanet, MD, PhD, Epidemiology of Emerging Infectious Diseases Unit, Institut Pasteur, 25-28 Rue du Docteur Roux, 75015 Paris, France (arnaud.fontanet@pasteur.fr).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofaf006

efficient collaboration, standardization of methodologies, and maximization of resources.

Serological analyses based on a combination of antibodies to different antigens of interest can yield better identification of people who are infected compared with a single antibody [7]. To our knowledge, this approach has not been applied to the identification of correlates of protection for respiratory viruses. Furthermore, the analysis of saliva IgG can provide an easily accessible alternative to serum sampling. Antibodies to SARS-CoV-2 have been demonstrated to persist in saliva following infection [8], but their role as a correlate of protection against infection has not been extensively studied. The role of mucosa-associated lymphatic tissue has been shown to be central to immunity against SARS-CoV-2 infection [9]. Thus, mucosal immunoglobulin (Ig) G could prove reliable correlates of protection.

In the present case–control study, we aimed to identify if serum and saliva IgG antibodies against a series of SARS-CoV-2 and non-SARS-CoV-2 antigens were correlated with protection against symptomatic SARS-CoV-2 infection.

METHODS

Participants Enrollment and Case–Control Matching

Study subjects were recruited among the participants of a case–control study (ComCor study) that was conducted online in mainland France and included cases with recent SARS-CoV-2 infection matched with controls enrolled by a market research opinion company and comparing exposures between cases and controls to identify settings at risk of transmission [10]. Between February and September 2022, we identified cases participating in this case–control study who resided in the greater Paris area. We included cases for the present study within 5 days of symptom onset to avoid the boosting effect of the ongoing infection [5, 6]. Among controls of the ComCor study (ie, without ongoing SARS-CoV-2 infection), we invited those who met the matching criteria with included cases. Both cases and controls were sampled at inclusion (serum and saliva antibodies, reverse transcriptase polymerase chain reaction [RT-PCR] on nasal swab to check case or control status), followed by 2 visits 1 and 6 months later. Saliva was sampled using a Lollisponge assay (Labelians)—this collection method was demonstrated to lead to lower variability in measured antibody levels compared with saliva collected directly into tubes. At each visit, we collected further information on comorbidities and exposures associated with risk of infection (public transport, presence of children in the household, in-person working, etc.) [11, 12]. The Omicron variant and its subvariants were the most frequent in France during the study period, mainly BA.2, until BA.5 became predominant in June 2022, according to weekly sequencing of a random selection of samples nationwide.

A total of 152 participants were included, consisting of 90 cases and 62 controls. Controls were recruited to reflect both cases' demographics and risk factors for coronavirus disease 2019 (COVID) exposure at the time of symptom onset. Matching was conducted based on 5 categorical variables: age group (younger/older than 60 years), number of documented COVID infections, vaccination history, time since last vaccination (less than a month, between 1 and 3 months, between 3 and 6 months, more than 6 months, no vaccination), and month of inclusion.

Processing of Biological Samples

Samples were received and stored by the Integrated Collections for Adaptive Research in Biomedicine biobank (Institut Pasteur, Paris, France). SARS-CoV-2 RNA in nasopharyngeal swabs were detected by RT-PCR by the National Center of Infectious Respiratory Disease (Institut Pasteur, Paris, France). Antibody levels in saliva and sera were measured on a Luminex Magpix platform, with a bead-based assay that captures fluorescence intensities directly proportional to the quantity of antibodies bound to antigens coupled on individually color-coded beads. A total of 24 different antigens were used for the multiplex assay: 16 SARS-CoV-2 antigens (wild-type [WT] and variants, for nucleocapsid protein [NP], spike [S], receptor-binding domain [RBD], and membrane envelope [ME]) and 8 antigens from seasonal coronaviruses (OC43, NL63, HKU1, and 229E for NP and S proteins). All antigens were sourced from the Native Antigen Company (Oxford, UK). This assay has previously been used to study the kinetics of the antibody response following infection with SARS-CoV-2 [7, 13]. Serum samples were run at a dilution of 1/200. Serum samples that had a saturation of fluorescence intensity were rerun at a higher dilution of 1/1600. Saliva samples were run at a dilution of 1/2. Saliva samples from month 6 were not available at the time of analysis. We have previously reported a strong correlation between neutralization titers and antibody titers, especially those targeting the spike and RBD [14].

Statistical Analyses

Median fluorescence intensities (MFIs) were converted to relative antibody units (RAUs) for serum and saliva samples to normalize the measured antibody response. This was done using a 5-parameter logistic curve fitted to the standard curve generated with a pool of 27 positive sera prepared from individuals with PCR-confirmed SARS-CoV-2, on the same plate to account for interassay variation, as validated in a prior study [13].

We matched cases with controls using optimal matching without caliper [15] on the following criteria: age (younger than 60 years or 60 years and older), number of documented past SARS-CoV-2 infections, number of past vaccine doses, time since last vaccination (less than a month, between 1 and 3 months, between 3 and 6 months, more than 6 months, no vaccination), and month of inclusion.

At each time point, antibody levels were compared between cases and controls using Wilcoxon signed-rank tests. The longitudinal trajectories were studied with generalized linear mixed models with a random effect accounting for within-individual variations. The outcome of these models was participant status (case or control), and explanatory covariates consisted of available antibody measurements. When comparing levels between cases and controls at each time point, fold-changes were calculated as the ratio of RAUs in matched pairs ($FC_{t,a} = \frac{RAU^{control}_{t,a}}{RAU^{case}_{t,a}}$ is the fold-change in a pair [control, case] at time point t for antibody a). Confidence intervals for mean fold-changes were calculated under the assumption that log-RAUs were normally distributed, and therefore their paired difference (*ie,* the log-fold-change) would follow a Student distribution with degree of freedom derived from the number of available pairs of cases and controls. The Student-derived point estimates and confidence intervals were subsequently back-transformed to the linear RAU scale.

Random forests were used for unsupervised learning and further classification of individual statuses using the Boruta algorithm for feature selection [16]. Briefly, the Boruta algorithm performs feature selection based on importance weighting of observed variables compared with repeated random permutations that simulate their distribution under the null hypothesis of independence with the outcome. The limit for number of permutations was set at 10 000 for the Boruta algorithm. The performance of the resulting classifiers was assessed [7] with receiver operating characteristic (ROC) curves denoting the trade-off between false-positive (1–specificity) and true-positive (sensitivity) rates. Areas under the curve (AUC) were also calculated. To assess the robustness of the findings to the choice of analysis method, the data were analyzed using logistic regression models (Supplementary Methods).

All analyses were performed using R (version 4.2.3) [17] and packages lme4 (version 1.1–32) [18], randomForest (version 4.7–1.1), pROC (version 1.18.2) [19], and Boruta (version 8.0.0) [16].

Ethics

The study was led by the Institut Pasteur (ClinicalTrials.gov identifier NCT04325646) and received ethical approval from the Comité de Protection des Personnes Île-de-France III on February 19, 2020. The study was conducted in accordance with the MR-001 reference methodology defined by the data protection authority (Commission Nationale de l'Informatique et des Libertés). All participants provided written informed consent. Participants received €30 of compensation for each visit to the Institut Pasteur for travel and time spent.

RESULTS

Participants/Follow-up

Between March 24 and September 12, 2022, a total of 152 participants were enrolled, consisting of 90 cases and 62 controls. Propensity score optimal matching yielded 62 case–control pairs, leaving 28 cases unmatched due to absence of control individuals deemed comparable enough. Among the enrolled participants, 89 cases and 61 controls were followed up at 1 month and, respectively, 84 and 59 at 6 months. Attrition was therefore very low and did not differ across groups (Fisher exact test, $P = .74$ at 6 months).

The study population was 60% female, and the median age (interquartile range [IQR]) was 48 (37–57) years. Most participants (90.8%) had a full vaccination scheme at enrollment (3 doses at the time), with the last dose having been administered more than 3 months before enrollment for most (96.1%). Table 1 presents the demographics and epidemiological characteristics of the study. Notably, cases and controls did not differ in the variables used for propensity score matching.

SARS-CoV-2 Antibody Levels Over Time in Serum

IgG antibodies to 16 SARS-CoV-2 antigens and 8 seasonal coronavirus antigens were measured in the full set of serum samples from infected and noninfected participants (n = 450 serum samples).

Antibody levels to SARS-CoV-2 proteins were raised at 1 month among cases while they remained flat in controls, consistent with the immunological boost expected from the infection. Notably, there was substantial interindividual variation, with antibody levels varying by orders of magnitude between participants (Figure 1). For anti-S and anti-RBD antibodies, it was not possible to assess the waning of antibody levels due to saturation of our assay even after reruns at a lower dilution factor (see Figure 1 for S [WT and Omicron] and RBD [WT]). A significant decay was observed for NP (WT) at 6 months in cases, though levels remained significantly higher than at inclusion (Wilcoxon signed-rank test, M1 vs M0: $P < .0001$; M6 vs M0: $P < .001$). Contrary to what was observed for antibodies against SARS-CoV-2, levels remained flat throughout the study in both cases and controls for antibodies against the 4 human seasonal coronaviruses (NL63, 229E, OC43, and HKU1).

SARS-CoV-2 Antibody Levels Over Time in Saliva

The analyses presented above were extended by incorporating saliva samples. Data from a total of 302 saliva samples from M0 and M1 (saliva samples from month 6 were not available at the time of analysis) were added to the previous data from 450 serum samples at M0, M1, and M6.

The general trends observed with serum samples were preserved, notably the raising of SARS-CoV-2-related antibody titers in cases at M1 and overall lower titers in cases than

Table 1. Case–Control Summary and Demographics

		Case (n = 90)	Controls (n = 62)	*P* Value
Sex, No. (%)	Female	58 (64.4)	34 (54.8)	.307
	Male	32 (35.6)	28 (45.2)	
Age, mean (SD), y	…	48.18 (18.13)	48.05 (13.98)	.962
Age category, No. (%)	<60 y	73 (81.1)	46 (74.2)	.414
	≥60 y	17 (18.9)	16 (25.8)	
No. of past vaccine doses (%)	0	2 (2.2)	2 (3.2)	.677
	1	1 (1.1)	0 (0.0)	
	2	5 (5.6)	3 (4.8)	
	3	82 (91.1)	56 (90.3)	
	4	0 (0.0)	1 (1.6)	
Time since last vaccine dose, No. (%)	No vaccination	2 (2.2)	2 (3.2)	.098
	<1 mo ago	0 (0.0)	1 (1.6)	
	1–3 mo ago	1 (1.1)	0 (0.0)	
	3–6 mo ago	61 (67.8)	30 (48.4)	
	>6 mo ago	26 (28.9)	29 (46.8)	
Past SARS-CoV-2 infection, No. (%)	No, none	73 (81.1)	52 (83.9)	.201
	Yes, once	9 (10.0)	6 (9.7)	
	Yes, twice	1 (1.1)	3 (4.8)	
	Does not know	7 (7.8)	1 (1.6)	
Days elapsed from symptom onset to sample, No. (%)	1	3 (3.3)	NA	NA
	2	7 (7.8)	NA	
	3	31 (34.4)	NA	
	4	40 (44.4)	NA	
	5	9 (10.0)	NA	

Nonparametric tests are used for comparison between cases and controls (Kruskal test for non-normal continuous variables and Fisher exact test for categorical variables).

Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

controls. The magnitudes of RAUs were overall lower than for serum and less prone to saturation, with clearer peak values. In addition, cases showed a similar significant spiking trend at M1 for all 4 seasonal coronaviruses ($P < .0001$ for all), while the corresponding signal remained flat in controls (Figure 2).

Fold-Change Antibody Levels Analysis in Serum and Saliva at Inclusion

In serum only, controls had increased levels of anti-SARS-CoV-2 antibodies compared with cases at inclusion, as indicated by fold-changes being systematically above 1 (Figure 3). That difference was found to be significant in serum samples for NP WT (FC, 4.14; 95% CI, 2.46–6.96), spike Alpha (FC, 3.23; 95% CI, 1.73–6.05), spike Omicron (FC, 2.98; 95% CI, 1.61–5.51), spike Delta (FC, 2.44; 95% CI, 1.21–4.90), S2 subdomain WT (FC, 2.11; 95% CI, 1.23–3.63), and RBD Beta (FC, 1.81; 95% CI, 1.05–3.73). In addition, antibodies against human coronavirus OC43 spike and NP were also significantly higher (respectively, FC, 1.74; 95% CI, 1.21–2.69; and FC, 1.60; 95% CI, 1.01–2.47), as were antibodies against 229E NP (FC, 1.56; 95% CI, 1.06–2.07). The 9 largest fold-changes for serum samples corresponded to SARS-CoV-2 WT and variant antigens. Notably, 4 of the 5 largest fold-changes in serum sample titers between controls and cases were found in anti-SARS-CoV-2 spike antibodies to variants of concern/interest (Alpha, Omicron, Delta, and Zeta, in decreasing order, as well as the borderline significant WT).

All antibodies measured from saliva had fold-changes significantly above 1. This gain in resolution may be attributed to optimized measurement scaling, especially in cases, no longer subjected to an upper limit of quantification. Fold-changes derived from salivary samples were found to be systematically above their serum counterpart (Figure 3), with most values ranging from 3 to 5 (same order of magnitude as for serum). SARS-CoV-2 fold-changes were higher than those for seasonal coronaviruses, with the exception of NP (229E) and spike (OC43). Figure 3 shows that antibodies, when tested in saliva samples, present a higher baseline difference than serum samples for most of the antibodies tested in the assay. Anti-SARS-CoV-2 antibodies have the largest fold-changes (saliva and serum combined), but antibodies to seasonal coronavirus antigens have been shown to have significant fold-changes for saliva samples, in opposition to serum. The 15 largest fold-changes correspond to SARS-CoV-2 WT and variant antibodies.

Identifying Correlates of Protection in Serum

A comprehensive analysis of the highly dimensional, multicorrelated antibody measurements at inclusion was conducted using random forests for classification in conjunction with the

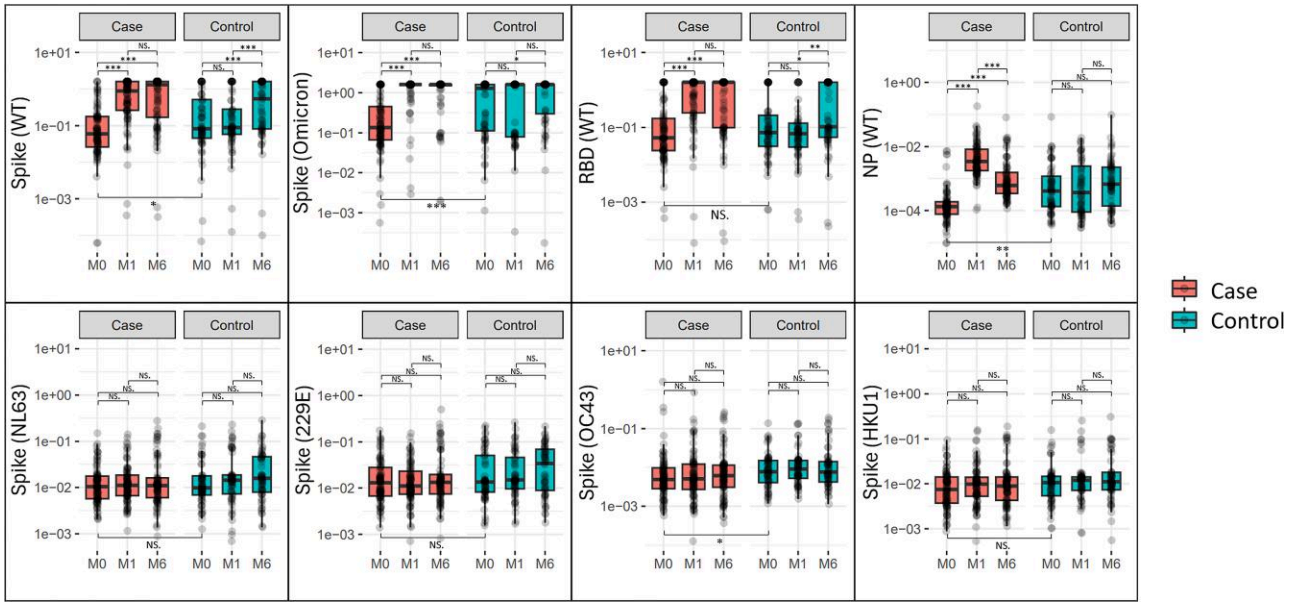

Figure 1. Boxplots of antibody levels measured from sera in cases and controls at scheduled visits. Antibody levels are expressed in RAU. Significance levels summarize the pairwise comparisons (M0 vs M1, M1 vs M6, M0 vs M6) in each study group or the comparisons between study groups at M0 using the Wilcoxon signed-rank test. The annotations on significance brackets correspond to *P* values: $^{NS}P > .05$; $^{***}P < .001$; $^{**}P < .01$; $^{*}P < .05$. Abbreviations: NS, nonsignificant; RAU, relative antibody units.

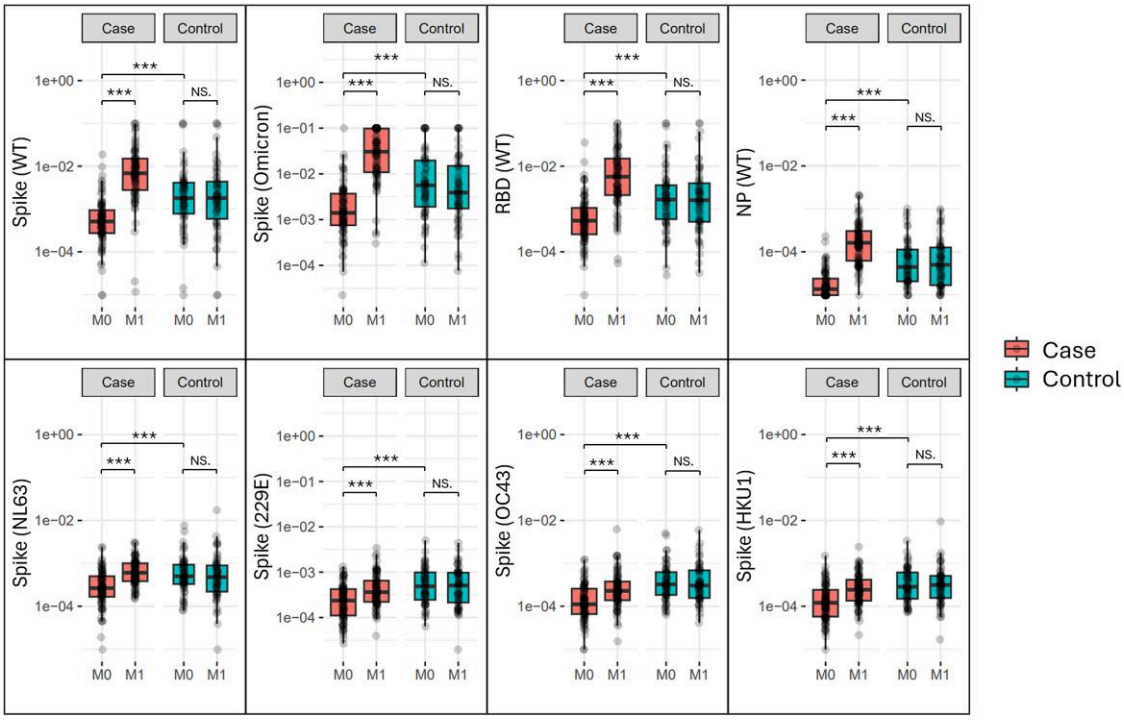

Figure 2. Boxplots of antibody levels measured from saliva in cases and controls at scheduled visits. Antibody levels are expressed in RAU. Significance levels summarize the pairwise comparisons (M0 vs M1) in each study group or the comparisons between study groups at M0 using the Wilcoxon signed-rank test. The annotations on significance brackets correspond to *P* values: $^{NS}P > .05$; $^{***}P < .001$; $^{**}P < .01$; $^{*}P < .05$. Abbreviations: NS, nonsignificant; RAU, relative antibody units.

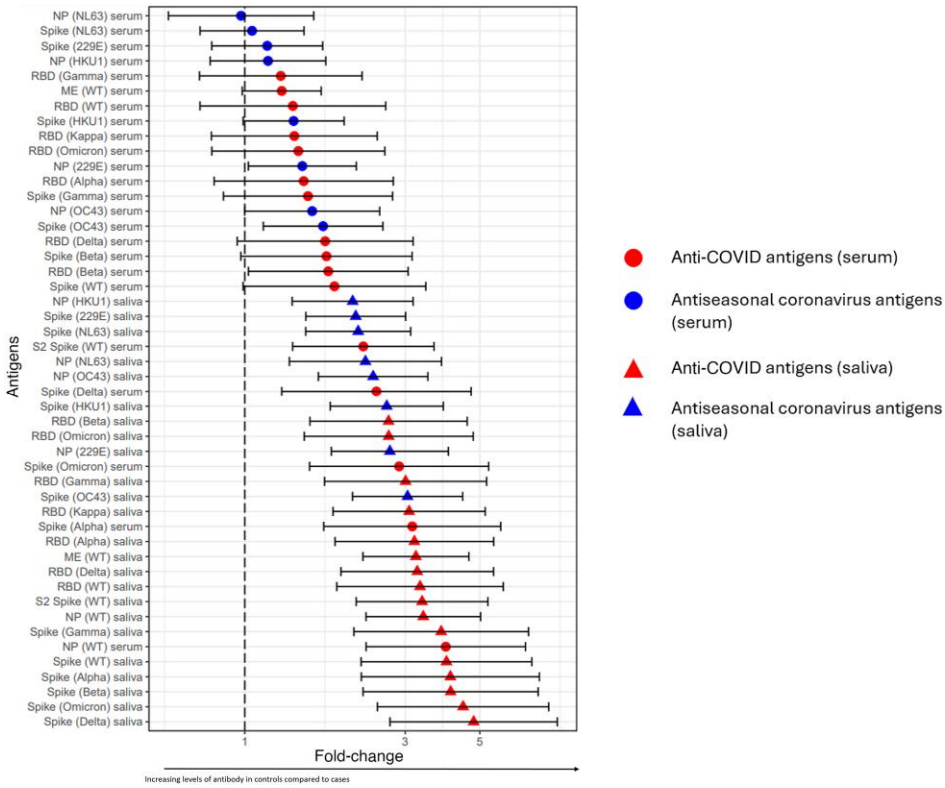

Figure 3. Fold-change of antibody levels for controls over cases at inclusion. All antibodies related to coronaviruses are presented (red: SARS-CoV-2; blue: seasonal coronaviruses; circle: serum; triangle: saliva). The vertical dashed line represents equality of antibody levels in controls and cases. Larger values (to the right) correspond to increased titers among controls. Confidence intervals are at 95%. Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Boruta feature selection algorithm. It identified 11 antibodies contributing significant information to a participant's group, that is, the ability to distinguish between cases and controls based solely on preboost M0 IgG levels. Among these 11 antibodies, 9 were SARS-CoV-2-specific, with the remaining 2 being NP for seasonal coronaviruses 229E and OC43. Notably, the top 4 features by order of Boruta importance were the same as their counterparts in the univariable fold-change analysis (Figure 4A).

An inflexion point in Z-score distributions can be noted starting at that fourth feature (spike [Alpha]), which denotes marginal improvement in predictive accuracy of the classifier as an individual progresses across the corresponding tree nodes. Figure 4B presents the ROC curves for antigens deemed of significant Boruta importance, either taken individually or combined into a trained random forest. The AUC for even the antigen with the best predictive value (NP [WT]: AUC, 0.714) remained lower than that of the random forest trained on the set of features selected by Boruta (AUC, 0.779).

Identifying Correlates of Protection in Serum and Saliva

The same analysis plan as above was implemented on the combined serum and saliva data. The Boruta algorithm identified 17 antibody responses significantly predictive of a participant's status (case or control). Among these, 5 were from seasonal coronaviruses: 229E (NP from saliva and serum), OC43 (spike and NP both from saliva), and HKU1 (spike from saliva). We note that NPs (229E and OC43) from saliva were consistent with our previous findings when including only serum samples. The 12 remaining were SARS-CoV-2-related, with the 3 of highest Boruta importance being ME (WT), NP (WT), and spike (Delta) from saliva.

Contrary to the analysis on serum antibodies alone, the random forest classifier based on pooled serum and saliva samples did not improve classification accuracy over a single-antigen predictor (AUC_{RF}, 0.811; and $AUC_{ME\ sal}$, 0.810). Figure 5B presents the ROC curves for antibodies deemed of significant Boruta importance. These findings are robust to the choice of statistical analysis method, with logistic regression models also demonstrating that anti-ME and anti-NP antibodies in saliva had the strongest association with protection (Supplementary Data).

DISCUSSION

By including cases within 5 days of symptom onset, we aimed to measure antibody levels before any infection-induced

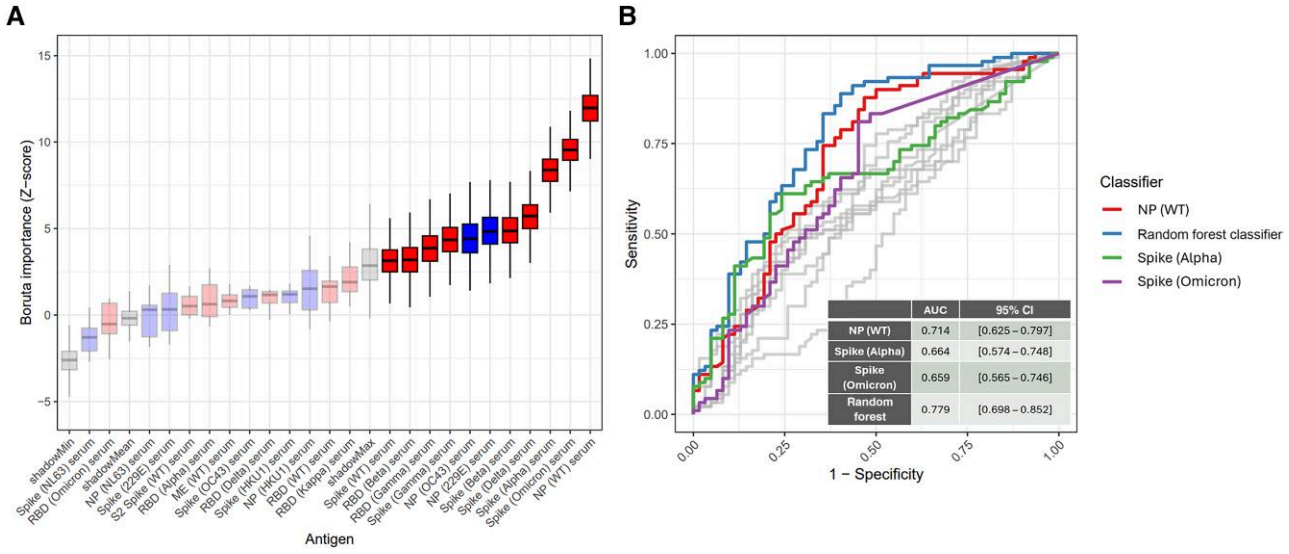

Figure 4. Feature selection and performance of classifiers trained with serum antibody responses. *A*, Boxplots of variable importance Z-scores calculated with the Boruta feature selection algorithm applied on the subset of serum antibodies against SARS-CoV-2 and seasonal coronaviruses. Antibodies are ordered from left to right by ascending median Z-score. Variable importance is defined as the Z-score of the mean decrease in accuracy of the classifier when that variable is shuffled to simulate the null (normalized permutation importance). Colors correspond either to antibodies against SARS-CoV-2 (red) or seasonal coronaviruses (blue), or to composite variables (gray) created by the Boruta algorithm to separate random noise from true classifier performance. Boxplots without transparency correspond to features deemed of importance by the feature selection algorithm. *B*, ROC curves for considered antibodies along with corresponding AUC. ROC curves are shown for threshold-based classification using single antibodies (translucent lines except for the 3 best: NP [WT] and spike [Omicron and Alpha]) and the final random forest (nontranslucent). Curves in gray correspond to the other 8 antibodies, with significant Z-scores as assessed using the Boruta algorithm. An AUC of 0.5 indicates a classifier that performs no better than random, while an AUC of 1 indicates a perfect classifier. Abbreviations: AUC, area under the curve; NP, nucleocapsid protein; ROC, receiver operating characteristic; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WT, wild-type.

immunity boost, based on data from previous studies [5, 6, 19]. A key assumption is that antibody levels measured within the first days of symptoms have not changed significantly from pre-infection antibody levels. Similar to Regev-Yochay et al., we were able to identify correlates of protection using a case–control approach [4]. Inclusion of potential cases right after exposure or symptom onset requires reactive inclusion and sampling. Compared with cohort studies, case–control studies require smaller sample sizes, are quicker to complete, and provide the additional advantage of studying immune response at time points very close to the infection. This design could prove useful for future epidemics, as it could be implemented rapidly to assess correlates of protection.

In this case–control study, we found that serum and saliva IgG levels against several SARS-CoV-2 antigens correlated with protection against SARS-CoV-2 infection with the Omicron variant. Antibody levels differed significantly between cases who were sampled within 5 days of symptom onset and uninfected controls who were matched for age, vaccine, and past infection. Differences in antibody levels between cases and controls were often greater in saliva than in serum. In the analysis investigating the association between antibodies in serum and saliva and protection against COVID-19, the top 3 selected biomarkers were anti-SARS-CoV-2 antibodies measured in saliva (Figure 5A). The stronger association between protection against COVID-19 and antibody levels in saliva, compared with antibodies levels in serum, may be attributable to the physical location of saliva close to the cells in the respiratory tract where viral invasion occurs. The observation that including data on antibodies measured in serum does not improve prediction performance (Figure 5B) is consistent with the hypothesis that serum antibodies have no direct role in protection against COVID-19. It is possible that other immune biomarkers such as IgA antibody responses or T-cell responses could have even stronger associations with protection against COVID-19.

At the time of the study, the predominant strains were the subvariants of Omicron, mainly BA.2 and BA.5. Participants had received vaccines designed on the wild-type strain. RBD is a subunit of the spike protein and displays higher variability among variants than the other subunits. This indicates a more variant-specific nature of the anti-RBD antibodies [20]. The full spike protein, being less variant-specific overall, could serve as a superior correlate of protection. A similar rationale applies to the NP and ME proteins, which play minimal roles in cell invasion by SARS-CoV-2 and, consequently, are less stressed by selection pressures. Moreover, ME (WT) and NP (WT) may act as indicators of prior infection. The higher mean anti-ME and anti-NP levels observed at baseline in controls suggest that a higher proportion of controls than cases had previously been infected with SARS-CoV-2 (or mounted a more

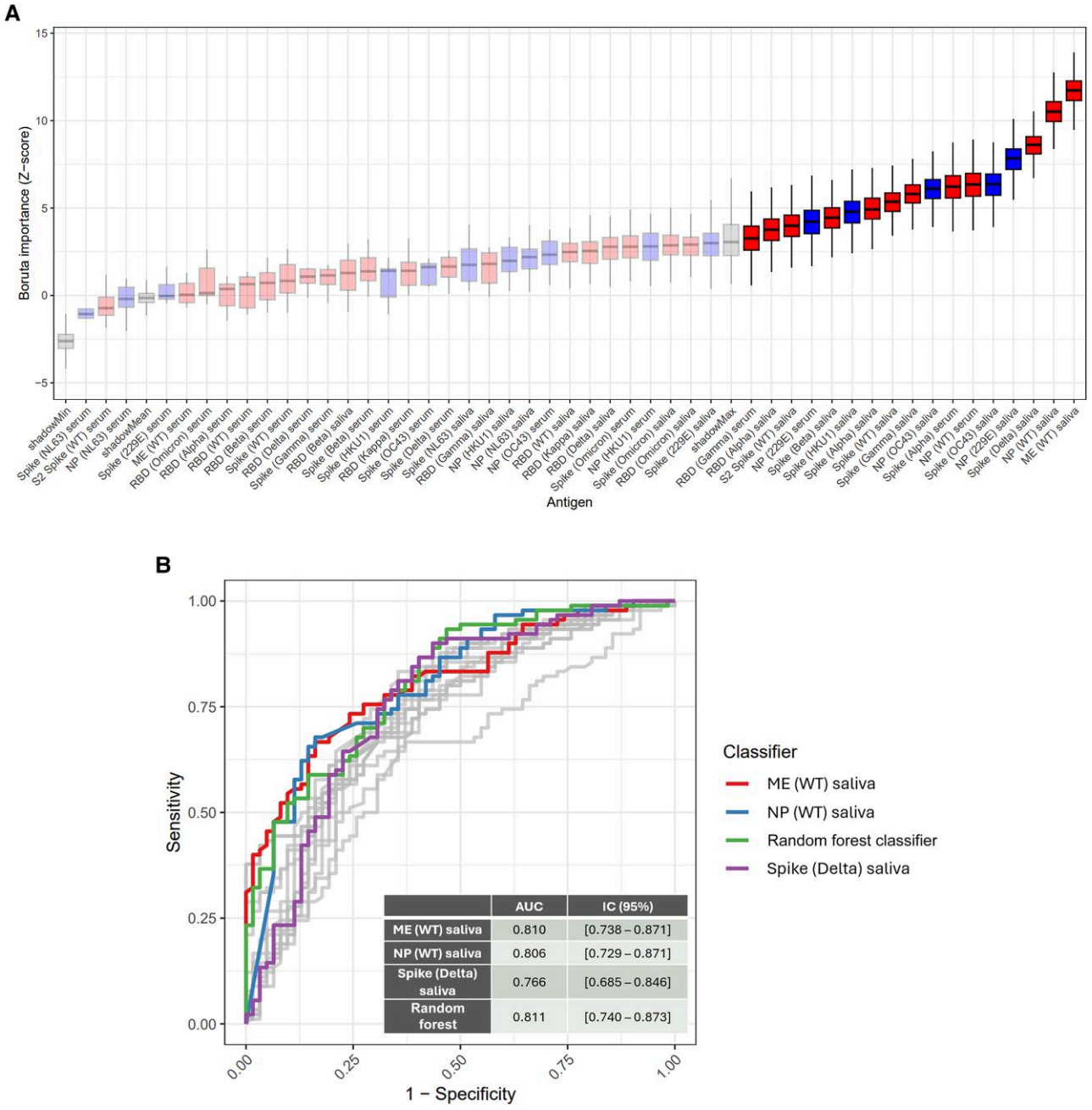

Figure 5. Feature selection and performance of classifiers informed with serum antibody responses. *A*, Boxplots of variable importance Z-scores calculated with the Boruta feature selection algorithm applied on the subset of serum and saliva antibodies against SARS-CoV-2 and seasonal coronaviruses. Antibodies are ordered from left to right by ascending median Z-score as previously shown in Figure 4. Colors correspond either to antibodies against SARS-CoV-2 (red) or seasonal coronaviruses (blue), or to composite variables (gray). Boxplots without transparency correspond to features deemed of importance by the feature selection algorithm. *B*, ROC curves for considered antibodies along with corresponding AUC. ROC curves are shown for threshold-based classification using single antibodies (translucent lines) and the final random forest (nontranslucent except for the 3 best: ME [WT], NP [WT], and spike [Delta]). Curves in gray correspond to the 14 antibodies with the lowest, yet significant Z-scores, as assessed using the Boruta algorithm. An AUC of 0.5 indicates a classifier that performs no better than random, while an AUC of 1 indicates a perfect classifier. Abbreviations: AUC, area under the curve; ME, membrane envelope; NP, nucleocapsid protein; ROC, receiver operating characteristic; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WT, wild-type.

immunogenic response to infection), despite our matching on the number of declared past episodes of infection, potentially resulting from undetected and possibly asymptomatic infections in controls. The role played by antibody responses to the ME antigen is poorly understood; however, we suspect that anti-ME antibodies are acting as correlates of the broader immune response rather than providing any direct functional protection. Most participants had received a complete vaccine

primary series and a booster, which confirms previous evidence on superior protection provided by hybrid immunity against symptomatic SARS-CoV-2 infection [21].

Using random forest–based classifiers yielded a moderate yet significant enhancement in classification performance compared with the top single-antigen classifier, notably when focusing on serum samples only but not when combining serum and saliva samples. This could be due to the high degree of correlations between antibodies observed in saliva samples. The use of a multiplex approach can prove advantageous in sero-epidemiological studies when exploring multiple biomarkers [7, 22]. However, the combination of these biomarkers using machine learning techniques may not always yield significant improvements, especially in highly correlated, highly dimensional data sets.

The stronger correlation of saliva antibody levels with protection compared with serum antibody levels may reflect the strong involvement of mucosal immunity in protection against infection [9]. It may also result from a higher occurrence of saturation of the assay in serum levels, a saturation that a higher dilution (1/1600) partially resolved as it helped recover one-third of the saturated data. This remaining assay saturation after sample dilution represents an important limitation of our study.

There were significant boosts in antibody levels to seasonal coronaviruses in cases compared with controls. This is potentially due to cross-reactivity, owing to the phylogenetic similarity in spike proteins, where conserved protein segments exist across various coronaviruses. Alternatively, this could suggest polyclonal activation, a phenomenon previously documented in other pathogens such as malaria [23]. While we found significant correlates of protection for seasonal coronaviruses, the literature suggests that this might reflect a general trend within the immune system rather than mechanistic correlates [24]. Investigation of the levels of neutralizing antibodies may offer additional insight.

This study had several limitations. The difficulty in enrolling controls meeting the matching criteria led to discarding 28 cases for the subset of statistical analyses requiring paired matching. Our study assessed correlates of protection for a mixture of vaccine-induced and hybrid vaccine- and infection-induced immunity, and it was not possible to determine the relative contribution of these sources of immunity. Given the high overlap of antibody titers between cases and controls, we could not assess potential thresholds associated with protection. We matched cases with controls on vaccine and timing of inclusion (month) but could not account for further potential differences in exposure to infection, for instance due to differences in behaviors that may result in potential residual confounding if controls were less likely to encounter SARS-CoV-2. The absence of confirmed asymptomatic cases represents another limitation in determining whether the observed effects represent protection against symptomatic infection or against any infection. In our investigation of correlates of protection, we focused on antibody levels. Other studies are warranted to evaluate other aspects of the immune response, including other immunoglobulins such as IgA, or the cellular immune response. Finally, although we observed no association between antibody levels in the first sample and days following symptom onset, we cannot entirely rule out that the infection-induced immune boost had not begun in all participants. However, if this had had an impact, it would have led to underestimation of the contrasts we observed between cases and controls.

In this case–control study, we provided evidence of correlates of protection of anti-SARS-CoV-2 IgG levels in saliva and serum against SARS-CoV-2 infection, with saliva levels often outperforming serum levels. These findings may contribute to the design of vaccines tailored to inducing mucosal immune responses. This study design could be used during future epidemics for the prompt assessment of correlates of protection.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

Author contributions. L.B. analyzed the data and wrote the first draft of the manuscript. E.B., L.G., S.P., and F.D. analyzed the samples. L.A., R.A., O.C., D.C., Y.D., M.D., D.E., M.E., C.F., S.F.P., N.J., H.L., E.R., M.S., L.S., M.N.U., S.V., A.X., and T.C. collected samples, data, and coordinated biobanking. T.C. oversaw project coordination. T.O. and S.G. supervised the analysis plan. A.F. and M.W. designed the study and supervised the research.

Financial support. This work was supported by the French government's "Integrative Biology of Emerging Infectious Diseases" (Investissement d'Avenir grant ANR-10-LABX-62-IBEID) and INCEPTION (Investissement d'Avenir grant ANR-16-CONV-0005) programs, the European Research Council (MultiSeroSurv 852373 to M.W.), and the "URGENCE COVID-19" fundraising campaign of Institut Pasteur.

Data availability. All data produced in the present study are available on request to the authors.

Potential conflicts of interest. All authors report that they have no conflicts of interest.

References

1. Goldblatt D, Fiore-Gartland A, Johnson M, et al. Towards a population-based threshold of protection for COVID-19 vaccines. Vaccine **2022;** 40:306–15.
2. Gilbert B, Donis RO, Koup RA, et al. A COVID-19 milestone attained—a correlate of protection for vaccines. N Engl J Med **2022;** 387:2203–6.
3. Hertz T, Levy S, Ostrovsky D, et al. Correlates of protection for booster doses of the SARS-CoV-2 vaccine BNT162b2. Nature Commun **2023;** 14:4575.
4. Regev-Yochay G, Lustig Y, Joseph G, 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.
5. Borremans B, Gamble A, Prager K, et al. Quantifying antibody kinetics and RNA detection during early-phase SARS-CoV-2 infection by time since symptom onset. eLife **2020;** 9:e60122.
6. Killingley B, Mann AJ, Kalinova M, et al. Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults. Nat Med **2022;** 28:1031–41.

7. Rosado J, Pelleau S, Cockram C, et al. Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: an antibody-based diagnostic and machine learning study. Lancet Microbe **2021;** 2:e60–9.
8. Isho B, Abe KT, Zuo MJ, et al. Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients. Sci Immunol **2020;** 5: eabe5511.
9. Puhach O, Bellon M, Adea K, et al. SARS-CoV-2 convalescence and hybrid immunity elicits mucosal immune responses. eBioMedicine **2023;** 98:104893.
10. Galmiche S, Charmet T, Schaeffer L, et al. Exposures associated with SARS-CoV-2 infection in France: a nationwide online case-control study. Lancet Reg Health Eur **2021;** 7:100148.
11. Grant R, Charmet T, Schaeffer L, et al. Impact of SARS-CoV-2 Delta variant on incubation, transmission settings and vaccine effectiveness: results from a nationwide case-control study in France. Lancet Reg Health Eur **2022;** 13:100278.
12. Galmiche S, Charmet T, Rakover A, et al. Risk of SARS-CoV-2 infection among households with children in France, 2020–2022. JAMA Netw Open **2023;** 6: e2334084.
13. Pelleau S, Woudenberg T, Rosado J, et al. Kinetics of the severe acute respiratory syndrome coronavirus 2 antibody response and serological estimation of time since infection. J Inf Dis **2021;** 224:1489–99.
14. Woudenberg T, Pinaud L, Garcia L, et al. Estimated protection against COVID-19 based on predicted neutralisation titres from multiple antibody measurements in a longitudinal cohort, France, April 2020 to November 2021. Euro Surveill **2023;** 28:2200681.
15. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stats **2011;** 10:150–61.
16. Kursa MB, Rudnicki WR. Feature selection with the boruta package. J Stat Software **2010;** 36:1–13.
17. R Core Team. *R: A Language and Environment for Statistical Computing* [Computer Software]. R Foundation for Statistical Computing; **2023**.
18. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv 1406.5823 [Preprint]. June 23, 2014. Available at: https://doi.org/10.48550/arXiv.1406.5823. Accessed June 2024
19. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics **2011;** 12:77.
20. Aydillo T, Rombauts A, Stadlbauer D, et al. Immunological imprinting of the antibody response in COVID-19 patients. Nature Commun **2021;** 12:3781.
21. Bobrovitz N, Ware H, Ma X, et al. Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the Omicron variant and severe disease: a systematic review and meta-regression. Lancet Inf Dis **2023;** 23: 556–67.
22. Grimée M, Tacoli C, Sandfort M, et al. Using serological diagnostics to characterize remaining high-incidence pockets of malaria in forest-fringe Cambodia. Malar J **2024;** 23:49.
23. Yman V, White MT, Asghar M, et al. Antibody responses to merozoite antigens after natural *Plasmodium falciparum* infection: kinetics and longevity in absence of re-exposure. BMC Med **2019;** 17:22.
24. Anderson EM, Goodwin EC, Verma A, et al. Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. Cell **2021;** 184:1858–64.e10.