Antibody phenotypes and probabilistic seroprevalence estimates during the emergence of SARS-CoV-2 in Sweden

Xaquín Castro Dopico[1#], Leo Hanke[1°], Daniel J. Sheward[1°], Murray Christian[1], Sandra Muschiol[2], Nastasiya F. Grinberg[3], Monika Adori[1], Laura Perez Vidakovics[1], ChangIl Kim[1], Sharesta Khoenkhoen[1], Pradeepa Pushparaj[1], Ainhoa Moliner Morro[1], Marco Mandolesi[1], Mattias Forsell[4], Jonathan Coquet[1], Martin Corcoran[1], Joanna Rorbach[5,6], Soo Aleman[7], Gordana Bogdanovic[2], Gerald McInerney[1], Tobias Allander[1,2], Chris Wallace[3,8], Ben Murrell[1*], Jan Albert[1,2*], Gunilla B. Karlsson Hedestam[1#]

Affiliations:
[1]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm 171 77, Sweden
[2]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm 171 76, Sweden
[3]Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, CB2 0AW
[4]Department of Clinical Microbiology, Umeå Universitet, Umeå 901 85, Sweden
[5]Department of Molecular Biochemistry & Biophysics, Karolinska Institutet, Stockholm 171 77, Sweden
[6]Max Planck Institute-Biology of Ageing, Karolinska Institutet Laboratory, Stockholm 171 77, Sweden
[7]Department of Infectious Diseases, Karolinska Universitetssjukhuset, Huddinge 141 52, Sweden
[8]MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge CB2 0SR, United Kingdom

°*Equal contribution
#Correspondence

Abstract

Serological studies are critical for understanding pathogen-specific immune responses and informing public health measures[1,2]. Here, we evaluate tandem IgM, IgG and IgA responses in a cohort of individuals PCR+ for SARS-CoV-2 RNA (*n*=105) representing different categories of disease severity, including mild and asymptomatic infections. All PCR+ individuals surveyed were IgG-positive against the virus spike (S) glycoprotein. Elevated Ab levels were associated with hospitalization, with IgA titers, increased circulating IL-6 and strong neutralizing responses indicative of intensive care status. Additional studies of healthy blood donors (*n*=1,000) and pregnant women (*n*=900), sampled weekly during the initial outbreak in Stockholm, Sweden (weeks 14-25, 2020), demonstrated that anti-viral IgG titers differed over 1,000-fold between seroconverters, highlighting the need for careful evaluation of assay cut-offs for individual measurements and accurate estimates of seroprevalence (SP). To provide a solution to this, we developed probabilistic machine learning approaches to assign likelihood of past infection without setting an assay cut-off, allowing for more quantitative individual and population-level Ab measures. Using these tools, that considered responses against both S and RBD, we report SARS-CoV-2 S-specific IgG in 6.8% of blood donors and pregnant women two months after the peak of spring COVID-19 deaths, with the SP curve and country death rate following similar trajectories.

Introduction

As SARS-CoV-2 only recently crossed the species barrier[3], the characterization of the humoral response to nascent outbreaks is central to optimizing approaches to tackle COVID-19 and further our understanding of human immunology[4–6].

Numerous SARS-CoV-2 studies have reported seroprevalence (SP) and disease-associated antibody (Ab) phenotypes[7–9], isolated virus-neutralizing monoclonal antibodies[10–13] and used convalescent individuals to define metrics for plasma therapy[14]. However, consensus on several key issues remains outstanding. For instance, the majority of serology data are derived from commercial kits utilizing spike derivatives or the nucleocapsid to detect pathogen-specific antibodies[7,15,16]. Several of these assays suffer from epitope loss[17], cross-reactivity[7,18] and suboptimal sensitivity[19–22]. Here, we develop highly sensitive and specific ELISA protocols based on native-like prefusion-stabilized spike (S) trimers[23] and the receptor-binding domain (RBD) produced in house, to accurately evaluate anti-viral Ab levels in COVID-19 patients and key community groups. We carefully analysed individuals with low anti-viral IgG responses and examined anti-viral Ab levels alongside *in vitro* virus neutralisation capacity and a descriptive set of clinical features to identify disease-associate Ab phenotypes.

Aside from the target antigen, a major consideration for Ab testing concerns setting the cut-off for positivity[24–26], which significantly affects seroprevalence (SP) estimates and individual clinical management. Currently employed approaches to the problem are severely limited by their high dependency on the representative nature of negative controls and the dichotomization of a continuous variable[27,28]; i.e. a sample is either *positive* or *negative*, but in reality, a wide spectrum of responses exists within the population. Therefore, to obtain more accurate SP estimates in blood donors and pregnant women, we strictly controlled our assay by the repeat sampling of historical controls (i.e. SARS-CoV-2 negative, n=295), and used data from SARS-CoV-2 PCR+ individuals to train probabilistic algorithms to handle ELISA measurements. We sampled blood donors (BDs) and pregnant women (PW) weekly throughout the first outbreak (late March-late June 2020). Blood donors are an important clinical resource, including for COVID-19 plasma therapy, while pregnant women require close clinical monitoring with respect to fetal-maternal health and are known to employ unique, yet poorly characterized immunological mechanisms that impact infectious pathology[29–31].

As Sweden did not impose a strict lockdown in response to the pandemic and has reported relatively high per-capita morbidity and mortality (ECDC, EU), understanding SARS-CoV-2 seroprevalence in these cohorts helps plan clinical need and understand the development of immunity in the population.

Results

Antibody test development

We developed ELISA protocols to profile IgM, IgG and IgA specific for a stabilized spike (S) glycoprotein trimer[23], its RBD, and the nucleocapsid (N). The two S antigens were produced in mammalian cells in-house (Fig. S1A) and trimer conformation was confirmed by cryo-EM[32]. A representative subset of the study samples (Table 1) was used for assay development (Fig. S1B). No reactivity was recorded amongst the negative control samples during test development, except for two endemic coronavirus-positive individuals that displayed reproducible IgM reactivity to SARS-CoV-2 N and S, and two 2019 blood donors with low anti-S IgM reactivity (Fig. S1C). Further investigation is, therefore, required to establish whether cross-reactive memory or primordial B lymphocyte lineages contribute to SARS-CoV-2 responses *a priori* in a subset of individuals.

Our assay revealed over a 1,000-fold difference in anti-viral IgG titers between Ab-positive individuals when examining serially diluted sera. In SARS CoV-2 PCR+ individuals, anti-viral IgG titers were comparable for S (EC_{50}=3,064; 95% CI [1,197 - 3,626]) and N (EC_{50}=2,945; 95% CI [543 - 3,936]) and lower for RBD [EC_{50}=1,751; 95% CI 966 - 1,595]. A subset (*ca.* 10%) of the SARS CoV-2-confirmed individuals did not have detectable IgG responses against N (Fig. S1D), suggesting that tests using this antigen may miss an important fraction of Ab-positive individuals, as seen previously[7].

Elevated anti-viral Ab responses are associated with disease severity

We next screened all SARS-CoV-2 PCR+ individuals in the study (n=105) and detected potent IgG responses against S in all individuals and against RBD in 97% of the persons (Fig. 1A and S2A). In healthy blood donors and pregnant women, titers were generally lower (Fig. 1A and S2A) and varied greatly, with some persons displaying low but likely real titers similar to those observed at the higher end of the negative control range. The need to identify low titer individuals, which comprise a relatively large fraction of seropositive healthy individuals, highlighted the challenge of using a pre-determined assay cut-off.

The IgM and IgA responses against S and RBD were generally less potent and more variable between individuals than the IgG response (Fig. 1B and S2B). Therefore, we sought to investigate whether isotype responses segregated with clinical features. To do this, the SARS-CoV-2 PCR+ individuals were classified according to clinical disease severity: Category 1 – *non-hospitalized (mild and asymptomatic)*; Category 2 – *hospitalized*; Category 3 – *intensive care* (on mechanical ventilation) (Table 1).

In all PCR+ individuals, anti-S and anti-RBD responses were highly correlated (Fig. S2C). Furthermore, multivariate analyses revealed increased anti-viral IgM, IgG and IgA to be associated with disease severity (Fig 1C and S Table 1), in line with the lower titers observed in blood donors and pregnant women. This was most pronounced for pathogen-specific IgA, suggestive of advancing mucosal disease[33]. A more severe clinical picture was also strongly associated with elevated serum IL-6 (Fig. 1C), a cytokine that feeds Ab production[34–37]. IL-6 is dysregulated in polygenic metabolic diseases[38–40] and acute respiratory distress syndrome (ARDS)[41], risk factors for COVID-19-associated mortality[42,43]. Interestingly, we observed a lack of association between serum IL-6 and anti-viral IgM levels, supporting that levels of the

127 cytokine and IgA mark a protracted, more severe clinical course. Female IgA RBD responses
128 were lower in non-hospitalized and hospitalized groups, compared to males (Fig. S3A)[44].
129 In our studies, anti-viral IgG levels were maintained two months post-disease onset/positive
130 PCR test, while IgM and IgA decreased (Fig. S3B), in agreement with their circulating $t_{1/2}$.
131 Serum from SARS-CoV-2 PCR+ individuals was collected 6-61 days post-PCR, with the
132 median time from symptom onset to PCR test being 5 days. In longitudinal patient samples
133 where we observed seroconversion, IgM, IgG and IgA peaked with similar kinetics when all
134 three isotypes developed, although anti-viral IgA was not always generated in Category 1 and
135 2 individuals (Fig 1D and S3C). Anti-viral IgM could be maintained two months post-
136 PCR+/symptom onset at the individual level (Fig S3C). Overall, disease severity showed the
137 most consistent relationship with any measure and was the primary predictor of Ab levels (Fig.
138 S3D). After accounting for the effect of age, sex and days since PCR+ test, anti-viral IgA titers
139 were approximately 3-fold higher in intensive care vs. non-hospitalized individuals, while IL-
140 6 levels were *ca.* 10-fold increased (Fig. S3E).
141 We next characterized virus neutralizing Ab response, a key parameter for understanding the
142 potential for protective humoral immune responses and the selection of plasma therapy donors.
143 Using an *in vitro* pseudotype virus assay to measure block of viral entry into target cells, we
144 detected neutralizing antibodies in the serum of all SARS-CoV-2 PCR+ individuals, and in all
145 except two healthy Ab-positive donors (from n=56). Neutralizing responses were not seen in
146 samples before seroconversion. A range of neutralizing ID_{50} titers were apparent (Fig 1E), with
147 binding and neutralization titers being highly correlated (Fig. S3D). The strongest neutralizing
148 responses were observed in samples from patients in intensive care (Category 3) (geometric
149 mean ID_{50}=5,058; 95% CI [2,422 - 10,564]), in keeping with their elevated and presumably
150 more developed Ab response (Fig 1F). Sera from healthy blood donors and pregnant women
151 also displayed neutralizing responses, but consistent with the binding data, these were less
152 potent than those observed in individuals with severe disease (ID_{50}=600; 95% CI [357 – 1,010]
153 and ID_{50}=350; 95% CI [228 - 538], respectively, Fig. 1G). Across the two antigens and three
154 isotypes, anti-RBD IgG was most strongly correlated with neutralization.
155
156 <u>Seroprevalence estimates in blood donors and pregnant women</u>
157 As the Stockholm region is a busy urban area and Sweden did not impose strict lockdown in
158 response to the emergence of SARS-CoV-2, we sought to better understand the frequency and
159 nature of anti-SARS-CoV-2 Ab responses in healthy blood donors and pregnant women
160 sampled during weeks 14-25 (March 30 to June 22 2020) (Fig. 2A). We surveyed a large
161 number of historical controls, blood donors from the spring of 2019 (n=295), to inform
162 decisions about how to set the assay cut-off for positivity. We considered this critical since the
163 use of a small set of control samples can lead to an incorrectly set threshold and errors in the
164 measurement. To illustrate this, we randomly sampled six groups of 20 negative controls from
165 890 measurements in 295 historical donors and we calculated SP in blood donors and pregnant
166 women from weeks 17-19 based on a 6SD cut-off. Depending on which 20 negative control
167 samples were used from the pool, the SP ranged from 5.7 to 8.7%, a 35% difference (Fig. 2B).
168 *Weak responder* status is likely influenced by many factors, including genetic background,
169 health status, total serum Ab and protein levels, and assay variability. Critically, test samples
170 with low anti-viral titers may also fall into this range, highlighting the need to better understand

171 the assay boundary. Taking a one-dimensional 6SD cut-off for anti-S IgG responses based on
172 all 890 values from 295 negative donors, 8% of healthy donors sampled in Stockholm tested
173 positive two months after the peak of deaths in the country (Fig. 2C and S4A and B).
174 To exploit individual antibody titers and improve our estimates, using an independent approach
175 we modelled the probability that a sample is IgG positive using machine learning approaches
176 that considered anti-S and -RBD responses together. After comparing different algorithms for
177 this purpose, we generated an ensemble SVM-LDA learner (*ENS*) to maximize sensitivity,
178 specificity and consistency across different cross-validation strategies and trained it using our
179 data from PCR+ individuals (please see: Materials and Methods). The *ENS* learner identified
180 6.5 and 6.6% IgG positivity in blood donors and pregnant women, respectively, two months
181 after the peak of outbreak deaths (S Table 2). Importantly, *ENS* identified several individuals
182 to have a 40-50% chance of being antibody positive, who could be targeted for further testing
183 (Fig. 2D). The accuracy of these tools will improve with larger training sets and will facilitate
184 future assays in providing multiplex antigen measurements in a single test, as well as
185 information on the strength of response.
186 Finally, to model population changes in SP over time, we developed and validated a cut-off-
187 independent Bayesian framework able to share information between weeks[45] (Fig. S4C and
188 Materials and Methods). Using this model, we found the steep increase in positivity at the start
189 of the pandemic to slow between weeks 17-25, suggesting that humoral immunity to the virus
190 developed slowly in these healthy cohorts despite considerable virus spread in the community.
191 Strikingly, seroprevalence inferred using our Bayesian approach exhibited the same trajectory
192 as Stockholm County deaths when lagged by two and half weeks (Fig. S4D).

Discussion

Serology remains the *gold standard* for estimating previous exposure to pathogens and benefits from a large historical literature[46]. It is a strong predictor of an anamnestic response and quicker recovery upon re-infection. Although the concept of herd immunity is based upon the study of antibodies, worryingly, there is no standardization for the many SARS-CoV-2 antibody tests currently available. Globally, hospital staff and health authorities are struggling with test choices, negatively impacting individual outcomes and efforts to contain SARS-CoV-2.

Benefitting from a robust antibody test developed alongside a diagnostic clinical laboratory responsible for monitoring sero-reactivity during the pandemic, we profiled SARS-CoV-2 Ab responses in three cohorts of clinical interest. COVID-19 patients receiving intensive care showed the highest anti-viral titers, developing augmented serum IgA and IL-6 with worsening disease, likely a consequence of age, sex and other risk factors, as well as prolonged immune stimulation and a higher infectious dose. Therefore, isotype-level measures may assist COVID-19 clinical management and determine, for example, whether all critically ill patients develop class-switched mucosal responses to SARS-CoV-2, potentially informing lung therapeutic delivery[47,48]. Our neutralization data showed that nearly all SARS-CoV-2 PCR+ individuals and healthy donors who seroconvert, develop Ab titers capable of preventing S-mediated cell entry *in vitro*, demonstrating the critical role of Abs to fight SARS-CoV-2.

Outside of the severe disease setting, it is important to determine how many people have seroconverted. Blood donors and pregnant women are both good sentinels for population health, although they are not enriched for high-risk groups, such as healthcare workers and public transportation employees, where SP may be higher. Blood donors are generally working age, active and mobile members of society with a good understanding of health, and the majority of pregnant women in Sweden would have been advised to take precautions against infectious diseases. Interestingly, in our study, both groups showed a similar SP during the time period analyzed.

By tracking these cohorts over time, we modelled SP changes at the population level with two independent tools we make freely available. We found the steep climb in Ab positivity at the start of the pandemic, to increase at a slower rate during subsequent weeks, plateauing at *ca.* 7%, well below the levels required for herd immunity. We noted that when lagged by two and half weeks, the trajectory of the SP curve coincided with the decreasing caseload and number of fatalities in Sweden[49]. Indeed, ICU occupancy and deaths are likely a better proxy for viral spread than PCR+ diagnoses, which are highly dependent on the number of tests carried out.

Given the uniqueness of the approach in Sweden[50], the data we present here may inform the management of this and future pandemics especially given the high inter-individual variability in anti-viral responses observed against SARS CoV-2.

Materials and methods

Human samples and ethical declaration

Samples from PCR+ individuals and admitted COVID-19 patients (*n*=105) were collected by the attending clinicians and processed through the Departments of Medicine and Clinical Microbiology at the Karolinska University Hospital. Samples were handled and analyzed in accordance with approval by the Swedish Ethical Review Authority (registration no. 2020-02811). All personal identifiers were pseudo-anonymized, and clinical feature data were blinded to the researchers carrying out experiments until data generation was complete. PCR testing for SARS-CoV-2 RNA was by nasopharyngeal swab or upper respiratory tract sampling at Karolinska University Hospital. As viral RNA CT values were determined using different qPCR platforms between patients, we did not analyze these alongside other available features. PCR+ individuals were questioned about the date of symptom onset at their initial consultation and followed-up for serology during their care, up to 2 months post-diagnosis. In addition, longitudinal samples from 10 of these patients were collected to monitor seroconversion and isotype persistence.

Anonymized samples from blood donors (100/week) and pregnant women (100/week) were randomly selected and obtained from the department of Clinical Microbiology, Karolinska University Hospital. No metadata, such as age or sex information were available for these samples. Pregnant women were sampled as part of routine for infectious diseases screening during the first trimester of pregnancy. Blood donors (*n*=295) collected through the same channels a year previously were randomly selected for use as negative controls. Serum samples from individuals testing PCR+ for endemic coronaviruses, 229E, HKU1, NL63, OC43 (*n*=20, ECV+) in the prior 2-6 months, were used as additional negative controls. The use of these anonymized samples was approved by the Swedish Ethical Review Authority (registration no. 2020-01807). Stockholm County death and Swedish mortality data was sourced from the ECDC and the Swedish Public Health Agency, respectively. Study samples are defined in Table 1.

Serum sample processing

Blood samples were collected by the attending clinical team and serum isolated by the department of Clinical Microbiology. Samples were barcoded and stored at -20ºC until use. Serum samples were not heat-inactivated for ELISA protocols but were heat-inactivated at 56ºC for 60 min for neutralization experiments.

SARS-CoV-2 antigen generation

The plasmid for expression of the SARS-CoV-2 prefusion-stabilized spike ectodomain with a C-terminal T4 fibritin trimerization motif was obtained from[23]. The plasmid was used to transiently transfect FreeStyle 293F cells using FreeStyle MAX reagent (Thermo Fisher Scientific). The ectodomain was purified from filtered supernatant on Streptactin XT resin (IBA Lifesciences), followed by size-exclusion chromatography on a Superdex 200 in 5 mM Tris pH 8, 200 mM NaCl.

274 The RBD domain (RVQ – QFG) of SARS-CoV-2 was cloned upstream of a Sortase A
275 recognition site (LPETG) and a 6xHIS tag, and expressed in 293F cells as described above.
276 RBD-HIS was purified from filtered supernatant on His-Pur Ni-NTA resin (Thermo Fisher
277 Scientific), followed by size-exclusion chromatography on a Superdex 200. The nucleocapsid
278 was purchased from Sino Biological.
279
280 Anti-SARS-CoV2 ELISA
281 96-well ELISA plates (Nunc MaxiSorp) were coated with SARS-CoV-2 S, RBD or
282 nucleocapsid (100 µl of 1 ng/µl) in PBS overnight at 4ºC. Plates were washed six times with
283 300 ml PBS-Tween-20 (0.05%) and blocked using PBS-5% no-fat milk (Sigma). Human serum
284 samples were thawed at room temperature, diluted (1:100 unless otherwise indicated), vortexed
285 and incubated in blocking buffer for 1h before plating. Serum samples were incubated
286 overnight at 4ºC before washing, as before. Secondary HRP-conjugated anti-human antibodies
287 were diluted in blocking buffer and incubated with samples for 1 hour at room temperature.
288 Plates were washed a final time before development with TMB Stabilized Chromogen
289 (Invitrogen). The reaction was stopped using 1M sulphuric acid and OD values were measured
290 at 450 nm using an Asys Expert 96 ELISA reader (Biochrom Ltd.). Secondary antibodies (all
291 from Southern Biotech) and dilutions used: goat anti-human IgG (2014-05) at 1:10,000; goat
292 anti-human IgM (2020-05) at 1:1000; goat anti-human IgA (2050-05) at 1:6,000. All assays of
293 the same antigen and isotype were developed for their fixed time and samples were randomized
294 and run together on the same day when comparing binding between PCR+ individuals. All data
295 were log transformed for statistical analyses.
296
297 *In vitro* virus neutralisation assay
298 Pseudotyped viruses were generated by the co-transfection of HEK293T cells with plasmids
299 encoding the SARS-CoV-2 spike protein harboring an 18 amino acid truncation of the
300 cytoplasmic tail[23]; a plasmid encoding firefly luciferase; a lentiviral packaging plasmid
301 (Addgene 8455) using Lipofectamine 3000 (Invitrogen). Media was changed 12-16 hours post-
302 transfection and pseudotyped viruses harvested at 48- and 72-hours, filtered through a 0.45 µm
303 filter and stored at -80°C until use. Pseudotyped neutralisation assays were adapted from
304 protocols validated to characterize the neutralization of HIV, but with the use of HEK293T-
305 *ACE2* cells. Briefly, pseudotyped viruses sufficient to generate ~100,000 RLUs were incubated
306 with serial dilutions of heat-inactivated serum for 60 min at 37°C. Approximately 15,000
307 HEK293T-ACE2 cells were then added to each well and the plates incubated at 37°C for 48
308 hours. Luminescence was measured using Bright-Glo (Promega) according to the
309 manufacturer's instructions on a GM-2000 luminometer (Promega) with an integration time of
310 0.3s. The limit of detection was at a 1:45 serum dilution.
311
312 IL-6 cytometric bead array
313 Serum IL-6 levels were measured in a subset of PCR+ serum samples (n=64) using an
314 enhanced sensitivity cytometric bead array against human IL-6 from BD Biosciences (Cat #
315 561512). Protocols were carried out according to the manufacturer's recommendations and
316 data acquired using a BD Celesta flow cytometer.
317

318 Statistical analysis of SARS-CoV-2 PCR+ data
319 All univariate comparisons were performed using non-parametric analyses (Kruskal-Wallis,
320 stratified Mann-Whitney, hypergeometric exact tests and Spearman rank correlation), as
321 indicated, while multivariate comparisons were performed using linear regression of log
322 transformed measures and Wald tests. For multivariate tests, all biochemical measures (IL-6,
323 PSV ID50 neut., IgG, IgA, IgM) were log transformed to improve the symmetry of the
324 distribution. As "days since first symptom" and "days since PCR+ test" are highly correlated,
325 we cannot include both in any single analysis. Instead, we show results for one, then the other
326 (Supplementary Table 1).
327
328 Probabilistic seroprevalence estimations
329 We employed two distinct probabilistic strategies for estimating seroprevalence without
330 thresholds, each developed independently. Our machine learning approach consisted of
331 evaluating different algorithms suited to ELISA data, which we compared through ten-fold
332 cross validation (CV): logistic regression, linear discriminant analysis (LDA), and support
333 vector machines (SVM) with a linear kernel. Logistic regression and linear discriminant
334 analysis both model log odds of a sample being case as a linear equation with a resulting linear
335 decision boundary. The difference between the two methods is in how the coefficients for the
336 linear models are estimated from the data. When applied to new data, the output of logistic
337 regression and LDA is the probability of each new sample being a case. Support vector
338 machines is an altogether different approach. We opted for a linear kernel, once again resulting
339 in a linear boundary. SVM constructs a boundary that maximally separates the classes (i.e. the
340 margin between the closest member of any class and the boundary is as wide as possible),
341 hence points lying far away from their respective class boundaries do not play an important
342 role in shaping it. SVM thus puts more weight on points closest to the class boundary, which
343 in our case is far from being clear. Linear SVM has one tuning parameter C, a cost, with larger
344 values resulting in narrower margins. We tuned C on a vector of values (0.001, 0.01, 0.5, 1, 2,
345 5, 10) via an internal 5-fold CV with 5 repeats (with the winning parameter used for the final
346 model for the main CV iteration). We also note that the natural output of SVM are class labels
347 rather than class probabilities, so the latter are obtained via the method of Platt[51]. We
348 considered three strategies for cross-validation: i) random: individuals were sampled into folds
349 at random, ii) stratified: individuals were sampled into folds at random, subject to ensuring the
350 balance of cases:controls remained fixed and iii) unbalanced: individuals were sampled into
351 folds such that each fold was deliberately skewed to under or over-represent cases compared
352 to the total sample. We sought a method that worked equally well across all cross-validation
353 schemes, as the true proportion of cases in the test data is unknown and so a good method
354 should not be overly sensitive to the proportion of cases in the training data. We found most
355 methods worked well, although logistic regression was sensitive to changes in the case
356 proportion in the training data. We chose to create an ensemble method combining that with
357 the highest specificity (LDA) and the highest sensitivity (SVM), defined as an unweighted
358 average of the probabilities generated under SVM and LDA. The ensemble learner had average
359 sensitivity > 99.1% and average specificity 99.8%. We then trained the ensemble learner on all
360 719 training samples and predicted the probability of anti-SARS-CoV-2 antibodies in blood
361 donors and pregnant volunteers sampled in 2020. We inferred the proportion of the sampled

population with positive antibody status each week using multiple imputation. We repeatedly (1,000 times) imputed antibody status for each individual randomly according to the ensemble prediction, and then analyzed each of the 1,000 datasets in parallel, combining inference using Rubin's rules, derived for the Wilson binomial proportion confidence interval[52].

Our Bayesian approach is explained in detail in Christian *et al*[45]. Briefly, we used a logistic regression over anti-RBD and -S training data to model the relationship between the ELISA measurements and the probability that a sample is antibody-positive. We adjusted for the training data class proportions and used these adjusted probabilities to inform the seroprevalence estimates for each time point. Given that the population seroprevalence cannot increase dramatically from one week to the next, we constructed a prior over seroprevalence trajectories using a transformed Gaussian Process, and combined this with the individual class-balance adjusted infection probabilities for each donor to infer the posterior distribution over seroprevalence trajectories.

Data and code availability statement

Data generated as part of the study, along with custom code for statistical analyses, is openly available via our GitHub repositories: https://github.com/MurrellGroup/DiscriminativeSeroprevalence/ and https://github.com/chr1swallace/seroprevalence-paper.

Author contributions

GKH and XCD designed the study, analyzed the results and wrote the manuscript with input from co-authors. JA, TA, SM, GB and SA provided the study serum samples and clinical information. LH, LPV, AMM, DJS, KCI, BM and GM generated SARS-CoV-2 antigens and pseudotyped viruses. XCD and MF developed the ELISA protocols and XCD generated the data. DJS and BM developed and performed the neutralization assay. MCh and BM developed the Bayesian framework. CW and NFG assisted with patient data statistical analyses and executed machine learning approaches. MA, SK, PP, MM, JC, MCo and JR carried out wet lab experiments and assisted with data analysis.

Acknowledgments

We would like to thank the study participants and attending clinical teams. Secondly, we extend our thanks to Björn Reinius, Marc Panas, Julian Stark, Remy M. Muts and Darío Solis Sayago for their input and discussion. Funding for this work was provided by a Distinguished Professor grant from the Swedish Research Council (agreement 2017-00968) and NIH (agreement SUM1A44462-02). CW and NFG are funded by the Wellcome Trust (WT107881) and MRC (MC_UP_1302/5).

Conflict of interest

The study authors declare no competing interests related to the work.

References

1. Altman, D., Douek, D. & Boyton, R. What policy makers need to know about COVID-19 protective immunity. *Lancet* (2020) doi:https://doi.org/10.1016/S0140-6736(20)30985-5.
2. Mouquet, H. & Nussenzweig, M. C. Polyreactive antibodies in adaptive immune responses to viruses. *Cellular and Molecular Life Sciences* (2012) doi:10.1007/s00018-011-0872-6.
3. Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2001017.
4. Davis, M. M. A Prescription for Human Immunology. *Immunity* vol. 29 835–838 (2008).
5. Winter, A. K. & Hegde, S. T. The important role of serology for COVID-19 control. *The Lancet Infectious Diseases* (2020) doi:10.1016/S1473-3099(20)30322-4.
6. London, A. J. & Kimmelman, J. Against pandemic research exceptionalism. *Science* (2020) doi:10.1126/SCIENCE.ABC1731.
7. Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* (2020) doi:10.1038/s41591-020-0897-1.
8. Zhao, J. *et al.* Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa344.
9. Tan, W. *et al.* Viral Kinetics and Antibody Responses in Patients with COVID-19. *J Clin* **doi: 10.11**, (2020).
10. Robbiani, D. F. *et al.* Convergent Antibody Responses to SARS-CoV-2 Infection in Convalescent Individuals. *Nature* (2020) doi:10.1101/2020.05.13.092619.
11. Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science (80-.).* (2020) doi:10.1126/science.abc5902.
12. Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* (2020) doi:10.1038/s41586-020-2380-z.
13. Seydoux, E. *et al.* Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. *Immunity* (2020) doi:10.1016/j.immuni.2020.06.001.
14. Duan, K. *et al.* Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *Proc. Natl. Acad. Sci. U. S. A.* (2020) doi:10.1073/pnas.2004168117.
15. Jiang, S., Hillyer, C. & Du, L. Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. *Trends in Immunology* (2020) doi:10.1016/j.it.2020.03.007.
16. Van Elslande, J. *et al.* Diagnostic performance of seven rapid IgG/IgM antibody tests and the Euroimmun IgA/IgG ELISA in COVID-19 patients. *Clin. Microbiol. Infect.* (2020) doi:10.1016/j.cmi.2020.05.023.
17. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* (2020) doi:10.1016/j.cell.2020.02.058.
18. Cheng, M. *et al.* Cross-reactivity of antibody against SARS-coronavirus nucleocapsid protein with IL-11. *Biochem. Biophys. Res. Commun.* (2005) doi:10.1016/j.bbrc.2005.10.088.
19. Özçürümez, M. K. *et al.* SARS-CoV-2 antibody testing—questions to be asked. *J. Allergy Clin. Immunol.* (2020) doi:10.1016/j.jaci.2020.05.020.
20. GeurtsvanKessel, C. H. *et al.* Towards the next phase: evaluation of serological assays for diagnostics and exposure assessment. *Nat Comms* **doi: 10.10**, (2020).
21. Deeks, J. J. *et al.* Antibody tests for identification of current and past infection with

SARS-CoV-2. *Cochrane Database Syst. Rev.* (2020) doi:10.1002/14651858.cd013652.
22. Traugott, M. *et al.* Performance of Severe Acute Respiratory Syndrome Coronavirus 2 Antibody Assays in Different Stages of Infection: Comparison of Commercial Enzyme-Linked Immunosorbent Assays and Rapid Tests. *J. Infect. Dis.* (2020) doi:10.1093/infdis/jiaa305.
23. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.).* (2020) doi:10.1126/science.aax0902.
24. Saraswati, K., Phanichkrivalkosil, M., Day, N. P. J. & Blacksell, S. D. The validity of diagnostic cut-offs for commercial and in-house scrub typhus IgM and IgG ELISAs: A review of the evidence. *PLoS Negl. Trop. Dis.* (2019) doi:10.1371/journal.pntd.0007158.
25. Xu, H., Lohr, J. & Greiner, M. The selection of ELISA cut-off points for testing antibody to Newcastle disease by two-graph receiver operating characteristic (TG-ROC) analysis. *J. Immunol. Methods* (1997) doi:10.1016/S0022-1759(97)00128-2.
26. Greiner, M., Franke, C. R., Böhning, D. & Schlattmann, P. Construction of an intrinsic cut-off value for the sero-epidemiological study of Trypanosoma evansi infections in a canine population in Brazil: a new approach towards an unbiased estimation of prevalence. *Acta Trop.* (1994) doi:10.1016/0001-706X(94)90044-2.
27. MacCallum, R. C., Zhang, S., Preacher, K. J. & Rucker, D. D. On the practice of dichotomization of quantitative variables. *Psychol. Methods* (2002) doi:10.1037//1082-989x.7.1.19.
28. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *British Medical Journal* (2006) doi:10.1136/bmj.332.7549.1080.
29. Aghaeepour, N. *et al.* An immune clock of human pregnancy. *Sci. Immunol.* (2017) doi:10.1126/sciimmunol.aan2946.
30. Mor, G. & Cardenas, I. The Immune System in Pregnancy: A Unique Complexity. *American Journal of Reproductive Immunology* (2010) doi:10.1111/j.1600-0897.2010.00836.x.
31. Kourtis, A. P., Read, J. S. & Jamieson, D. J. Pregnancy and infection. *New England Journal of Medicine* (2014) doi:10.1056/NEJMra1213566.
32. Hanke, L. *et al.* An alpaca nanobody neutralizes SARS-CoV-2 by blocking receptor interaction. *bioRxiv* (2020) doi:10.1101/2020.06.02.130161.
33. Cervia, C. *et al.* Systemic and mucosal antibody secretion specific to SARS-CoV-2 during mild versus severe COVID-19. *bioRxiv* (2020) doi:10.1101/2020.05.21.108308.
34. Eto, D. *et al.* IL-21 and IL-6 are critical for different aspects of B cell immunity and redundantly induce optimal follicular helper CD4 T cell (Tfh) differentiation. *PLoS One* **6**, e17739 (2011).
35. Dienz, O. *et al.* The induction of antibody production by IL-6 is indirectly mediated by IL-21 produced by CD4 + T cells. *J. Exp. Med.* (2009) doi:10.1084/jem.20081571.
36. Maeda, K., Mehta, H., Drevets, D. A. & Coggeshall, K. M. IL-6 increases B-cell IgG production in a feed-forward proinflammatory mechanism to skew hematopoiesis and elevate myeloid production. *Blood* (2010) doi:10.1182/blood-2009-07-230631.
37. Beagley, K. W. *et al.* Interleukins and IgA synthesis. Human and murine interleukin 6 induce high rate IgA secretion in IgA-committed B cells. *J. Exp. Med.* (1989) doi:10.1084/jem.169.6.2133.
38. Mauer, J. *et al.* Signaling by IL-6 promotes alternative activation of macrophages to limit endotoxemia and obesity-associated resistance to insulin. *Nat. Immunol.* (2014) doi:10.1038/ni.2865.
39. Kristiansen, O. P. & Mandrup-Poulsen, T. Interleukin-6 and diabetes: The good, the bad, or the indifferent? *Diabetes* (2005) doi:10.2337/diabetes.54.suppl_2.S114.

40. Ferreira, R. C. *et al.* Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and Influences Risk of Diverse Inflammatory Diseases. *PLoS Genet.* **9**, (2013).
41. Park, W. Y. *et al.* Cytokine balance in the lungs of patients with acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* (2001) doi:10.1164/ajrccm.164.10.2104013.
42. Magro, G. SARS-CoV-2 and COVID-19: Is interleukin-6 (IL-6) the 'culprit lesion' of ARDS onset? What is there besides Tocilizumab? SGP130Fc. *Cytokine: X* (2020) doi:10.1016/j.cytox.2020.100029.
43. Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0602-4.
44. Jin, J. M. *et al.* Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. *Front. Public Heal.* (2020) doi:10.3389/fpubh.2020.00152.
45. Christian, M. & Murrell, B. Discriminative Bayesian Serology: Counting Without Cutoffs. *BioRxiv* (2020).
46. Behring & Kitasato. Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. *Dtsch. Medizinische Wochenschrift* (1890) doi:10.1055/s-0029-1207589.
47. Larios Mora, A. *et al.* Delivery of ALX-0171 by inhalation greatly reduces respiratory syncytial virus disease in newborn lambs. *MAbs* (2018) doi:10.1080/19420862.2018.1470727.
48. Sheridan, C. Ablynx's nanobody fragments go places antibodies cannot. *Nat. Biotechnol.* (2017) doi:10.1038/nbt1217-1115.
49. Intensivvårdsregistret, S. Antal som intensivvårdas med Covid-19 per dag. *Svenska Intensivvårdsregistrets utdataportal* https://portal.icuregswe.org/siri/report/corona.co (2020).
50. Paterlini, M. 'Closing borders is ridiculous': the epidemiologist behind Sweden's controversial coronavirus strategy. *Nature* (2020) doi:10.1038/d41586-020-01098-x.
51. Platt, J. & others. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. large margin Classif.* (1999).
52. Lott, A. & Reiter, J. P. Wilson Confidence Intervals for Binomial Proportions With Multiple Imputation for Missing Data. *Am. Stat.* (2020) doi:10.1080/00031305.2018.1473796.

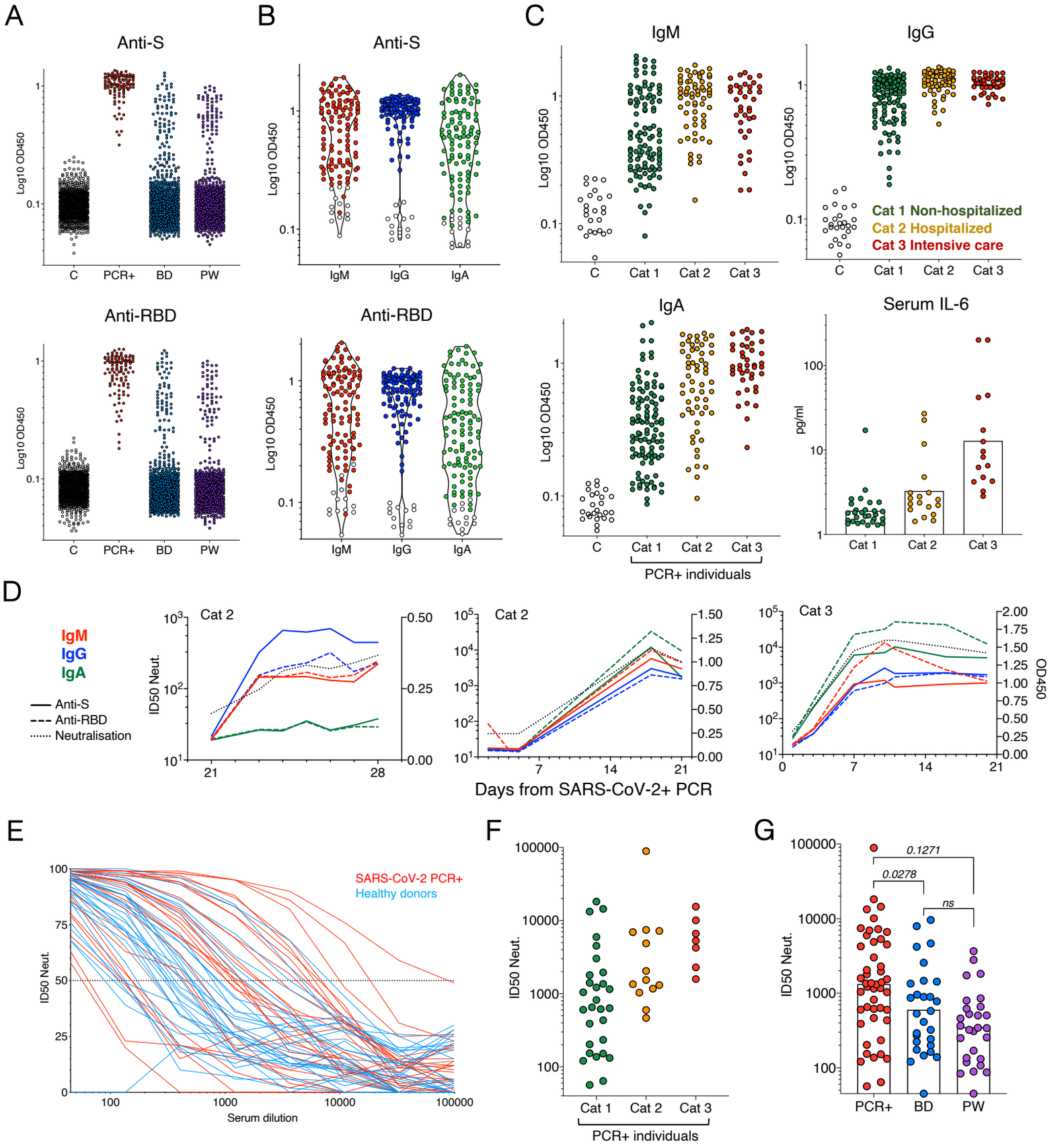

Figure 1: Anti-SARS-CoV-2 Ab phenotypes in PCR+ individuals, COVID-19 patients, blood donors and pregnant women. (**A**) Log transformed anti-S and -RBD IgG responses are shown in all SARS-CoV-2 PCR+ individuals (*n*=105), blood donors (BD, *n*=1,000) and pregnant women (PW, *n*=900). Controls (C) represent *n*=295 blood donors from 2019 analyzed repeatedly (to *n*=890). (**B**) IgM, IgG and IgA responses against S and RBD in PCR+ individuals (*n*=105), with a limited number of controls for each assay represented by open circles. Cat 1: non-hospitalized, Cat 2: hospitalized patients, Cat 3: ICU patients – from *n*=105 PCR+ (**C**) Anti-viral antibody levels in PCR+ individuals (*n*=105) and serum IL-6 levels responses according to clinical disease severity. S and RBD responses are graphed together, alongside a limited number of controls. (**D**) Three discordant longitudinal patient profiles of seroconversion and neutralisation capacity are shown in hospitalized individuals. (**E**) Pseudotyped virus neutralization. Serial dilutions of serum from representative PCR+ individuals (*n*=30, *red*) and healthy donors (*n*=30). (**F**) Neutralization ID_{50} titers according to disease severity category. (**G**) Comparison of neutralization ID_{50} titers between patients, blood donors and pregnant women. Bars represent geo.mean and *P* values are from a two-tailed Mann-Whitney test.

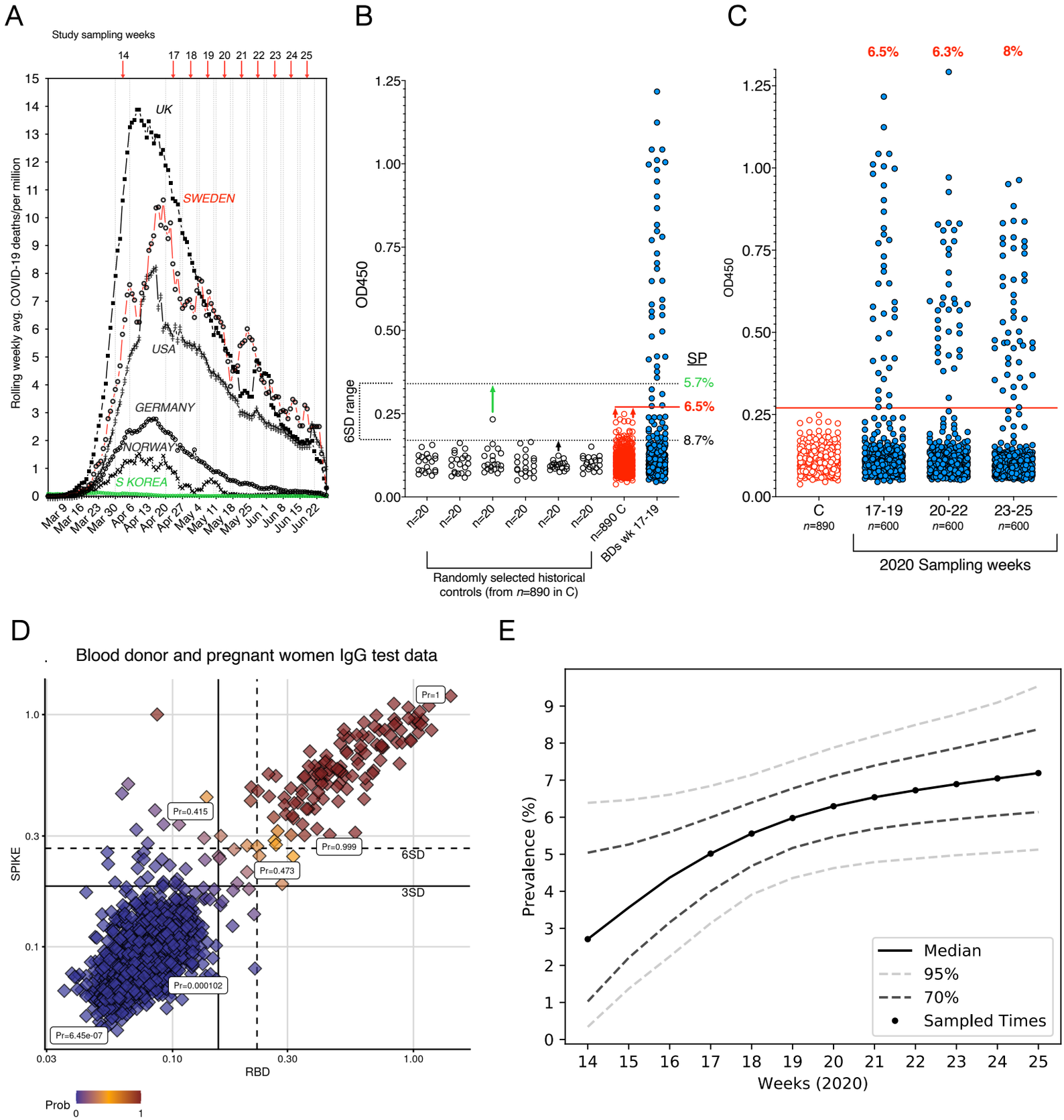

Figure 2: Probability-based seroprevalence in Stockholm during the initial outbreak

(**A**) Study sample collection intervals are shown alongside the rolling weekly average COVID-19 deaths per million inhabitants in relevant countries for comparison. (**B**) By randomly selecting six groups of 20 negative controls from a pool of 890 values (from n=295 individuals), the range of the 6SD cut-off from the lowest to the highest is shown /by the dotted line). BDs from wk17-19 shown. (**C**) Using a conventional 6SD cut-off, we found 6.2, 6.5 and 7.7% of healthy donors to be antibody-positive in weeks 17-19, 20-22 and 23-25, respectively. (**D**) The ensemble SVM-LDA learner trained on our SARS-CoV-2 PCR+ participant data illustrates the problem of weak responders and provides a multi-dimensional solution (tandem S and RBD) by providing individual datapoints with a likelihood of positivity, a feature with potential for clinical use. *ENS* identified 6.5% and 6.6% of blood donors and pregnant women, respectively, to be IgG-positive in the study. Conventional 3 and 6sd cut-offs are shown for comparison. (**E**) Seroprevalence estimates in the healthy donor cohorts through time using a cut-off-independent Bayesian approach.

Table 1 – Study patient, blood donor and pregnant women characteristics

PCR+ individuals (n=)		105
Females		44 (41.9%)
Males		61 (58.1%)
Median age (years)*		53.0 (49-61)
	Females	51.5 (48-56.2)
	Males	55.0 (49-63)
Non-hospitalized		53 (50.5%)
	Females, males	28, 25
Hospitalized patients		31 (29.5%)
	Females, males	12, 17
Intensive care (ICU) patients		21 (20%)
	Females, males	3, 17
Disease category unknown		2 (1.9%)
	Females, males	1, 1
SARS-CoV2+ PCR		105
Sample collection dates		February-June 2020
Blood donors (n=)		1,000
Median age (years)		N/A
Sex		N/A
SARS-CoV2+ PCR		N/A
Sample collection dates		Weeks 14-25 2020
Pregnant women (n=)		900
Age (years)		N/A
SARS-CoV2+ PCR		N/A
Sample collection dates		Weeks 17-25 2020
Historical blood donors (n=)		295
Median age (years)		N/A
Sex		N/A
SARS-CoV2+ PCR		No
Sample collection dates		Spring 2019
ECV+ donors (n=)		20
Median age (years)		N/A
Sex		N/A
SARS-CoV2+ PCR		No
Sample collection dates		Fall/Winter 2018-9

Table 1

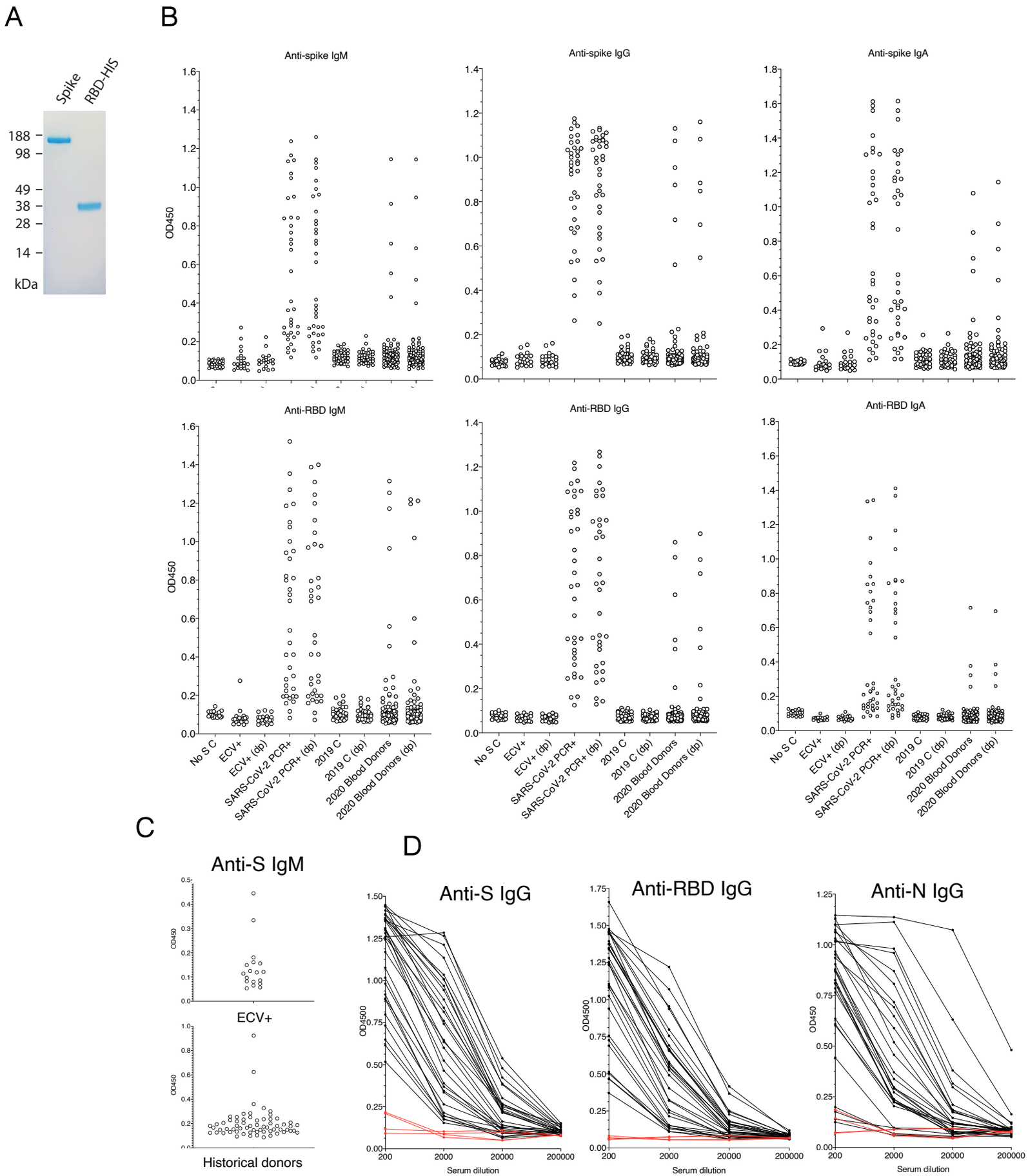

Figure S1: Anti-SARS-CoV-2 ELISA protocol development

(**A**) Trimeric S and RBD were expressed in 293F cells and purified as described. (**B**) A subset of PCR+ individuals and negative controls were used to validate the assays for the three isotypes, these were: individuals with confirmed SARS CoV-2 infection (*n*=36, from a range of disease severity categories); blood donors from the spring of 2020 (*n*=100); two sets of negative controls, blood donors from the spring of 2019 (*n*=75) and individuals PCR+ for endemic coronaviruses (ECV+) (*n*=20) (**C**) Two ECV+ donors, K2 and K4, showed reproducible IgM binding to S, please refer to their increased responses to S and nucleocapsid in S2. Testing another subset of historical controls (*n*=75) for a similar observation, two additional individuals were found to show IgM binding to S. (**D**) Serial dilutions of PCR+ participant serum are shown (in a representative sample, *n*=40) for anti-S and anti-RBD IgG.

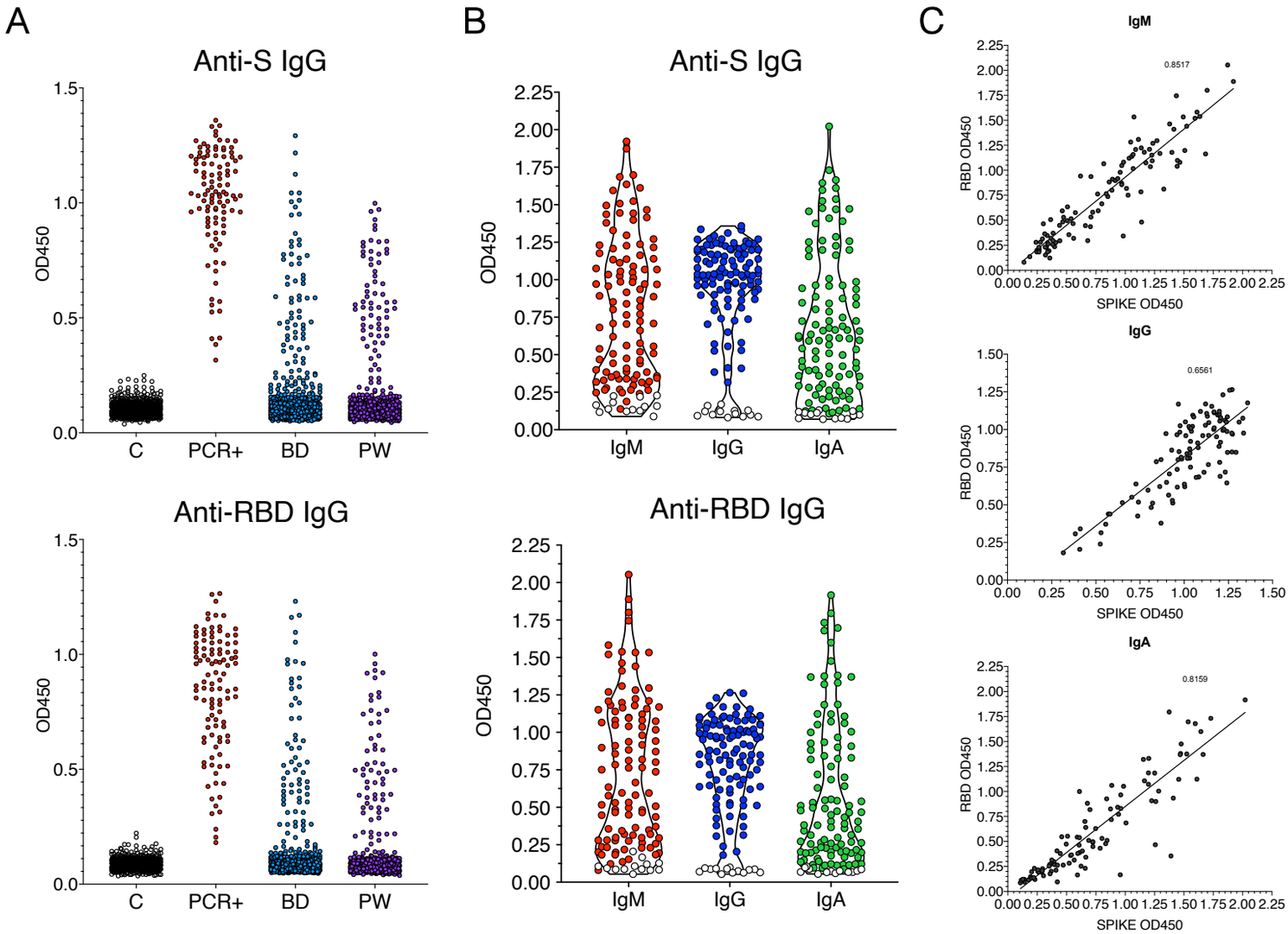

Figure S2: Antibody levels in PACR+ individuals and healthy participants phenotypes

(**A**) Raw OD450 anti-S and -RBD IgG responses are shown for SARS-CoV-2 PCR+ individuals and healthy donors. BD: blood donors PW: pregnant women. (**B**) Raw OD450 isotypic responses in PCR+ individuals against S and RBD. A limited number of negative controls are depicted by clear circles. (**C**) Anti-S vs -RBD responses in PCR+ individuals are highly correlated.

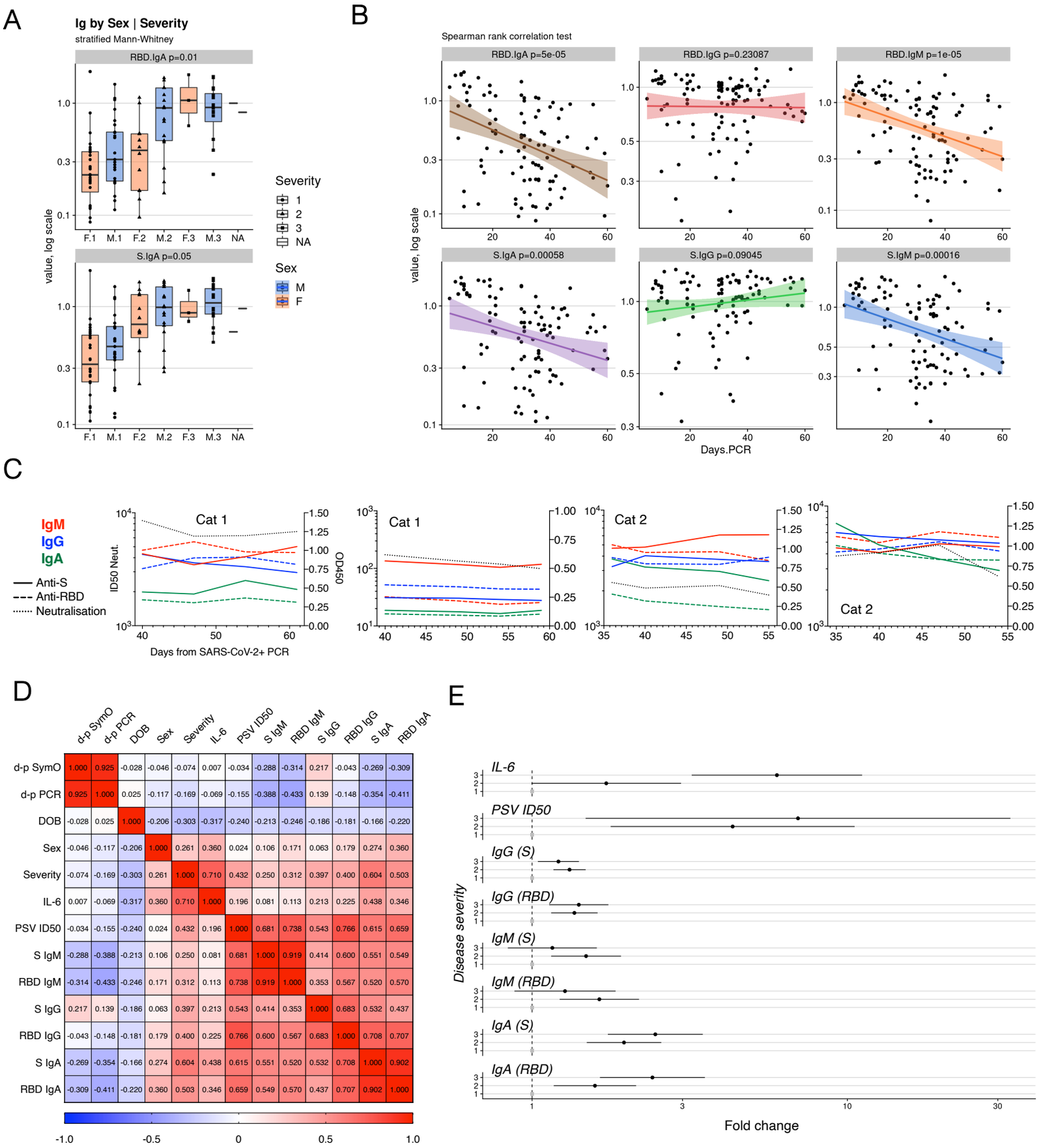

Figure S3: Antibody phenotypes in PCR+ individuals

(**A**) Antibody responses according to sex for anti-S and –RBD IgA. (**B**) IgM and IgA titers declined with time from first symptom/SARS-CoV-2+ PCR. IgG levels were maintained during this time. *P* values from a Spearman rank correlation test. (**C)** Longitudinal profiles of category 1 (mild and asymptomatic) and 2 (hospitalized patients) during seroconversion. (**D**) Spearman's rank correlation of PCR+ dataset features and antibody levels. DOB - *date of birth*; d-p SymO - *days post-symptom onset*; d-p PCR – *days post SARS-CoV-2+ PCR*; PSV ID50 – *neutralizing titer* (**E**) Adjusted fold-change for dataset features in PCR+ individuals compared to category 1. The effects of age (DOB), sex, days from PCR test were considered.

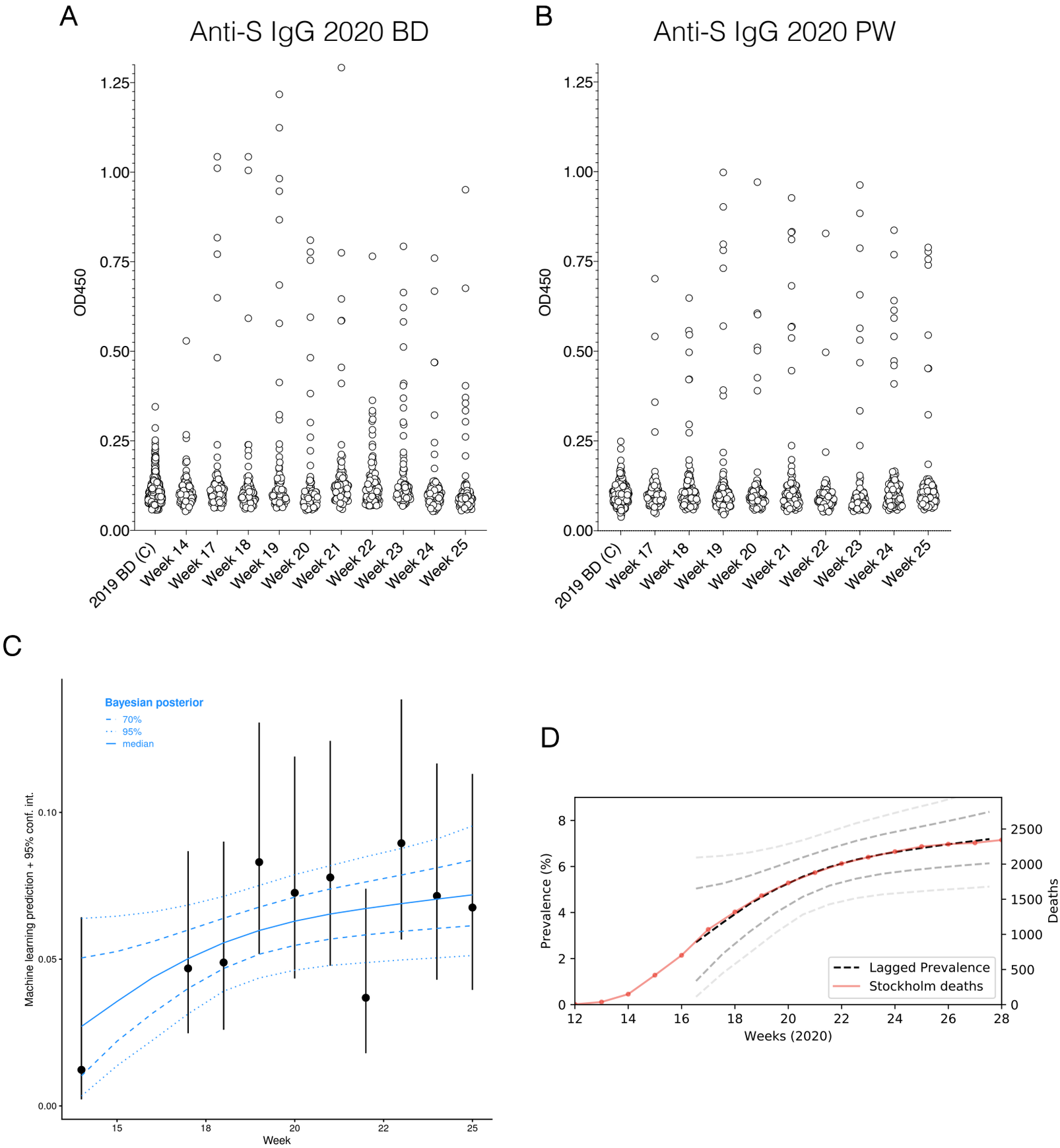

Figure S4: Healthy donor raw data and the comparison of cut-off-independent machine learning methods with our Bayesian framework

Raw anti-S IgG OD values for blood donors (A) and pregnant women (B) sampled during the pandemic according to sampling week. 890 control values (C) from 2019 Bds and run alongside test samples is shown. Raw data (inc. for anti-RBD) is available in the supplementary data table. (C) Overlay and agreement of SVM-LDA ensemble learner predictions with our Bayesian framework. (D) Seroprevalence estimates make a close approximation to Stockholm region deaths.