

1 **Evaluation of commercially available high-throughput SARS-CoV-2 serological assays for**  
2 **serosurveillance and related applications**

3 Running title: Evaluation of high-throughput COVID-19 serology assays

4  
5 Mars Stone\*<sup>1,2</sup>, Eduard Grebe\*<sup>1,2,3</sup>, Hasan Sulaeman<sup>1</sup>, Clara Di Germanio<sup>1</sup>, Honey Dave<sup>1</sup>,  
6 Kathleen Kelly<sup>1</sup>, Brad Biggerstaff<sup>4</sup>, Bridgit O. Crews<sup>5</sup>, Nam Tran<sup>6</sup>, Keith R. Jerome<sup>7</sup>, Thomas N.  
7 Denny<sup>8</sup>, Boris Hogema<sup>9</sup>, Mark Destree<sup>10</sup>, Jefferson M. Jones<sup>11</sup>, Natalie Thornburg<sup>11</sup>, Graham  
8 Simmons<sup>1,2</sup>, Mel Krajden<sup>12</sup>, Steve Kleinman<sup>13</sup>, Larry J. Dumont<sup>1,14</sup>, Michael P. Busch<sup>1,2</sup>

9  
10 Author affiliations:

11 Vitalant Research Institute, San Francisco CA, USA (M. Stone, E. Grebe, H. Sulaeman, C. Di  
12 Germanio, H. Dave, K. Kelly, G. Simmons, L.J. Dumont, M.P. Busch); Department of  
13 Laboratory Medicine, University of California, San Francisco, CA, USA (M. Stone, E. Grebe, G.  
14 Simmons, M.P. Busch); SACEMA, Stellenbosch University, Stellenbosch, South Africa (E.  
15 Grebe); Centers for Disease Control and Prevention, Fort Collins, CO, USA (B. Biggerstaff)  
16 University of California Irvine Medical Center, Orange, CA, USA (B.O. Crews)  
17 Department of Pathology and Laboratory Medicine University of California, Davis, CA, USA  
18 (N. Tran); Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA,  
19 USA (K.R. Jerome); Duke Human Vaccine Institute, Duke University, Durham, NC, USA (T.  
20 Denny); Department of Blood-Borne Infections, Sanquin Research, Amsterdam, The  
21 Netherlands (B. Hogema); BloodWorks North West, Seattle, WA, USA (M. Destree)  
22 Centers for Disease Control and Prevention, COVID-19 Response Team, Atlanta GA, USA  
23 (J.M. Jones, N. Thornburg); British Columbia Centre for Disease Control, Vancouver, Canada

24 (M. Krajden); Department of Pathology and Laboratory Medicine, University of British  
25 Columbia, Vancouver, Canada (S. Kleinman); University of Colorado School of Medicine,  
26 Denver, CO, USA (L.J. Dumont).

27 \* These first authors contributed equally to this article.

28 Disclaimer: The findings and conclusions in this report are those of the authors and do not  
29 necessarily represent the official position of the Centers for Disease Control and Prevention.

30

31 Word count: 3390/3500

32 Abstract 150/150

33

34 Keywords: COVID-19 Antibody Testing, Serology Testing for COVID-19, High-Throughput  
35 Biological Assays, Seroprevalence

36

37

38 Corresponding author: Mars Stone, PhD., Vitalant Research Institute, San Francisco CA, USA,  
39 Department of Laboratory Medicine, University of California San Francisco, CA, USA. 270  
40 Masonic Ave, San Francisco, CA 94118. Phone (415) 354-1389, fax (415) 775-3859.

41

42

43

44

45

46

47 **Abstract**

48 SARS-CoV-2 serosurveys can estimate cumulative incidence for monitoring epidemics but  
49 require characterization of employed serological assays performance to inform testing algorithm  
50 development and interpretation of results. We conducted a multi-laboratory evaluation of 21  
51 commercial high-throughput SARS-CoV-2 serological assays using blinded panels of 1,000  
52 highly-characterized blood-donor specimens. Assays demonstrated a range of sensitivities (96%-  
53 63%), specificities (99%-96%) and precision (IIC 0.55-0.99). Durability of antibody detection in  
54 longitudinal samples was dependent on assay format and immunoglobulin target, with anti-spike,  
55 direct, or total Ig assays demonstrating more stable, or increasing reactivity over time than anti-  
56 nucleocapsid, indirect, or IgG assays. Assays with high sensitivity, specificity and durable  
57 antibody detection are ideal for serosurveillance. Less sensitive assays demonstrating waning  
58 reactivity are appropriate for other applications, including characterizing antibody responses  
59 after infection and vaccination, and detection of anamnestic boosting by reinfections and vaccine  
60 breakthrough infections. Assay performance must be evaluated in the context of the intended  
61 use.

62

63

64

65

66

67

68

69

## 70 **Introduction**

71           Serosurveillance for SARS-CoV-2 infection is critical to monitor the course of the  
72 evolving pandemic and local outbreaks, and informs infection fatality ratios, vaccine penetrance  
73 and the impact of mitigation measures, and levels of population immunity. Serosurveillance  
74 should be conducted with representative population sampling using well characterized  
75 serological assays selected based on their performance characteristics and optimized algorithms.  
76 The use of assays and algorithms that detect mild or asymptomatic infections are critical for  
77 accurately estimating cumulative incidence, and case- and death-to-infection ratios.

78           More than >85 SARS-CoV-2 antibody (Ab) assays received FDA Emergency Use  
79 Authorization (EUA) as of August 19, 2021, ranging from point-of-care tests to fully automated  
80 high-throughput platforms [1]. These assays target different immunoglobulins (Ig) against viral  
81 antigens (full length Spike protein [S1/S2], subunit 1 [S1] and/or subunit 2 [S2] of Spike, the  
82 receptor binding domain [RBD] of Spike, or the nucleocapsid protein [N]). Detection methods  
83 include lateral flow assays [LFA], enzyme-linked immunosorbent assays [ELISA], and  
84 chemiluminescent immunoassay [CLIA], and detection of either total Ig, or selective IgG, IgM  
85 or IgA antibodies [1]. There are limited head-to-head evaluation data available for high-  
86 throughput SARS-CoV-2 serological assays and few large-scale studies that have focused on  
87 performance for serosurveillance applications. Comprehensive characterization of assay  
88 performance must include sensitivity, specificity, and durability of antibody detection over time  
89 since infection.

90           We conducted a standardized, multi-laboratory comparative assessment of 21 high-  
91 throughput, commercially available SARS-CoV-2 serological assays using blinded panels of  
92 1,000 highly characterized de-identified specimens including longitudinal and cross sectional

93 COVID-19 convalescent plasma (CCP) and pre-COVID control plasma specimens. Panels were  
94 distributed to experienced testing laboratories determined to be proficient by the manufacturers.  
95 Assays were selected to represent multiple formats and antigen targets. Data from this study will  
96 inform assay selection and development of testing algorithms to meet the optimal performance  
97 characteristics for primary screening and supplemental testing in US and global serosurveillance  
98 studies. The study also provides performance data relevant to other serological testing contexts  
99 that will allow clinicians, public health organizations, laboratorians, and emergency response  
100 planners to develop optimal algorithms for infection detection and confirmation, including  
101 vaccine breakthrough and recurrent infections.

## 102 **Methods**

### 103 *Assay selection, panel development and testing*

104 Key characteristics of the assays included, such as format and configuration, antigen  
105 composition, and immunoglobulin target, are summarized in Table 1. Uniquely blinded identical  
106 panels were distributed to experienced testing laboratories to determine performance  
107 characteristics including sensitivity, specificity, repeatability, dilutional performance, and  
108 durability of reactivity over time.

109 Plasma or serum specimens were obtained from apheresis plasma units or associated  
110 whole blood tubes from consenting CCP donors from March through November 2020.  
111 Specimens were shipped from collection sites and stored frozen at Vitalant Research Institute  
112 (VRI) until panel assembly and frozen distribution. All blood donors consented to use of de-  
113 identified, residual specimens for further research purposes. Consistent with the policies and  
114 guidance of the UCSF IRB, VRI self-certified that use of the de-identified CCP donations in this  
115 study does not meet the criteria for human subjects research. CDC investigators reviewed and

116 relied on this determination as consistent with applicable federal law and CDC policy (45 C.F.R.  
117 part 46, 21 C.F.R. part 56; 42 U.S.C. Sect. 241(d); 5 U.S.C. Sect. 552a; 44 U.S.C. Sect. 3501).  
118 Qualification for CCP donation required documentation of positive SARS-CoV-2 molecular or  
119 serologic test, complete resolution of symptoms 14-28 days prior to donation[2], and reactivity  
120 on the primary screening Ortho VITROS SARS-CoV-2 S total Ig (Ortho Clinical Diagnostics,  
121 Raritan, NJ) Ab assay and standard allogeneic blood donor qualification criteria[3].

122 To evaluate the waning of sensitivity over time, longitudinal specimens were included  
123 from 24 CCP donors who continued to qualify for CCP donation at each of 4-14 donations  
124 (median 9) over 79-126 days (median 95). A COVID-19 Seroconversion Panel consisted of 14  
125 time points from a single source plasma donor during the progression of a SARS-CoV-2  
126 infection over 87 days[4]. Fifteen CCP specimens were represented in 6 blinded replicates to  
127 evaluate precision. The dilution panel consisted of six 4-fold serial dilutions of specimens with a  
128 range of neat Ab titers [5]. The panel also included 24 apparent serosilent specimens from donors  
129 who initially qualified for CCP donation as having a positive molecular test but without evidence  
130 of seroconversion by the Ortho S total Ig assay. Specificity panel included pre-pandemic blood  
131 donor specimens derived from plasma components collected before end of year 2019 (N=432)  
132 and 27 donations collected in early 2020 that tested non-reactive on Ortho CoV2T and were non-  
133 neutralizing by pseudovirus neutralization assay [5].

#### 134 *Statistical Analysis*

135 All statistical analyses were performed using the R statistical programming language (v. 4.0.4,  
136 [6]) and using various packages, including the *binom* package for confidence intervals on  
137 proportions [7], the *glm2* package [8] for regression analysis and the *ggplot2* package [9] for  
138 plotting.

139 *Sensitivity and Specificity*

140           Sensitivity was assessed in cross sectional CCP specimens. Because data on symptoms,  
141 clinical severity, hospitalization, and diagnostic test results (molecular or antigen) were not  
142 available, we defined “true positivity” according to three sets of criteria: (1) qualification as a  
143 CCP donor according to blood center policies, which required donors to provide evidence of a  
144 SARS-CoV-2 diagnosis, with symptomatic infection resolved at least 2 weeks prior to the first  
145 donation (N=191), [<https://www.fda.gov/media/141477/download>], (2) confirmation of  
146 detectable nAb by the Broad live virus neutralization assay [5] (N=154), or (3) reactive on  $\geq 3$   
147 evaluated binding Ab (bAb) tests (N=198). There is substantial overlap between the three  
148 definitions, with 149 specimens classified as positive by all three definitions, 34 by (1) and (3), 3  
149 by (2) and (3) and only 22 positive by only one of the definitions. The 24 purposely selected  
150 “serosilent” CCP specimens (see Table 2) were excluded from the sensitivity analysis based on  
151 criterion 1 above, while specimens from the longitudinal CCP donor cohort were excluded from  
152 all sensitivity analyses. Donors who continued to qualify for CCP donation may bias sensitivity  
153 estimates given they were required to have bAb reactivity for continued donation of CCP.

154           Specificity was assessed using pre-pandemic blood donor specimens (N=432) and 27  
155 seronegative early 2020 donations [5]. These 27 samples were not included in the primary  
156 analysis but were included in a secondary specificity analysis (N=459) (Appendix Figure 1).

157           All sensitivity and specificity estimates were based on reported qualitative interpretations  
158 of assay results. Results defined and reported by the manufacturer as “equivocal” were excluded  
159 from primary sensitivity and specificity estimates. A secondary analysis of sensitivity was  
160 conducted in which we considered results reported as equivocal by the testing lab as non-reactive  
161 (Appendix Figure 2). All 95% confidence intervals are Wilson score intervals.

162 *Repeatability and Assay Precision*

163           Coefficients of variation (CV, i.e., the ratio of the standard deviation across  
164 measurements of the six replicate specimens to the mean of the six measurements, expressed as a  
165 percentage) were computed for each of the replicate specimens (N=90). A limitation of this  
166 approach is that assays with narrower dynamic range produced very low or zero CVs for results  
167 at the upper limit of quantification. To adequately account for the impact of specimens with  
168 reactivity outside the measurement range, we excluded these specimens from the overall  
169 repeatability assessment for which we used the intraclass correlation coefficients (ICCs). The  
170 ICC expresses between-sample variance as a proportion of total variance in the tested replicate  
171 specimen. In the case of the Bio-Rad BioPlex assay (Bio-Rad, Hercules, CA), on-board dilutions  
172 were conducted by the testing lab and used to estimate reactivity in specimens where initial  
173 results were above the assay's limit of quantitation.

174 *Dilutional performance*

175           The dilution panel (N=55) allows comparative assessment of the linearity of observed vs.  
176 expected reactivity measurements above and below assay cutoffs. Expected reactivity is defined  
177 as the mean signal intensity measured over six replicates of the neat specimen divided by the  
178 dilution factor. These analyses are reported in supplemental materials.

179 *Durability of antibody detection*

180           We assessed both qualitative and quantitative durability of bAb detection in longitudinal  
181 CCP specimens (N=209 specimens from 27 donors). Documented dates of symptom onset,  
182 symptom resolution or nucleic acid test (NAT)-based diagnosis are not available for these  
183 donors, so all analyses are anchored to the index donation. These CCP donors first presented for  
184 donation early in the pandemic, typically within one month of symptom resolution [5].



185           Qualitative detection was assessed by estimating the proportion of specimens with  
186 detectable bAbs grouped in 30-day bins of time since index donations. To account for within-  
187 donor correlation, if a donor contributed more than one specimen in a particular time bin, the  
188 proportion of the donor's specimens that were reactive was added to the numerator for the bin,  
189 and only 1 to the denominator, so that the proportion detected reported is the proportion of  
190 donors whose bAbs could be detected in each bin.

191           Quantitative detection was assessed by fitting linear mixed effects regression models with  
192 time since index donation as the predictor. Average (fixed) slopes from models fit to log-  
193 transformed and rescaled signal intensities to a fixed range, and assay signal half-lives were  
194 estimated. (Appendix 1).

## 195 **Results**

196           When a true positive was defined by qualification as a CCP donor, the lowest assay  
197 sensitivity was 63.6% (95% CI: 56.3%-70.4%; EUROIMMUN IgA assay), and the highest was  
198 95.8% (95% CI: 92.0%-7.9%; Ortho VITROS Total Ig S assay; Figure 1, panel A). When a true  
199 positive was defined by confirmed neutralization by plaque reduction neutralization test (PRNT),  
200 the lowest assay sensitivity was 69.7% (95% CI: 61.7%-76.7%; EUROIMMUN IgA assay), and  
201 the highest was 98.7% (95% CI: 95.4%-99.6%; Ortho VITROS Total Ig S assay; Figure 1, panel  
202 B). Most assays (17/20) had sensitivities >80%, 12/20 had sensitivities >90%, and 7/20 had  
203 sensitivities >95%, by the first definition. No assays reached a sensitivity of 96% by CCP  
204 qualification criteria or 99% by detectable nAbs criteria. Assays with the lowest sensitivity were  
205 the Beckman Coulter Access SARS-CoV-2 IgG, Diazyme DZ-Lite SARS-CoV-2 IgG and  
206 EUROIMMUN IgA assays, with estimates below 80%. Figure 1, panel C shows similar patterns

207 to the first and second definitions of true positivity when true positivity was defined by the  
208 ‘operational standard’ of positivity based on bAb reactivity on three or more assays.

209 Specificities, based on testing 432 pre-COVID-19 specimens, were high, with estimates  
210 ranging from 96.1% (95%CI: 93.8%-7.5%; Diazyme DZ-Lite assay) to 100% (95%CI: 99.1%-  
211 100%; Abbott IgG N, Bio-Rad BioPlex IgG, Bio-Rad Platelia Total Ig N, and Ortho VITROS  
212 Total Ig S assays. Most assays (13/20) had specificities above 99%, and 5/20 assays had  
213 specificities of 100% in this panel (Figure 2). Assays with poorer specificity tended to have  
214 poorer sensitivity, suggesting no tradeoff between sensitivity and specificity (Appendix Table 1  
215 and Appendix Figure 3). Specificity estimates that included the 27 specimens from 2020 are  
216 shown in Appendix Figure 1. Secondary sensitivity analysis with ‘equivocal’ results categorized  
217 as non-reactive is shown in Appendix Figure 2. The inclusion of these 2020 and equivocal  
218 samples had minimal impact on estimates of specificity and sensitivity respectively.

219 Durability of bAb detection was highly variable, with some assays reactive at all  
220 longitudinal timepoints, while others showed substantial declines in the proportion of reactive  
221 specimens over time (Figure 3). IgG assays and anti-N assays generally demonstrated more rapid  
222 seroreversion proportions compared to total Ig and anti-S assays. For example, the Abbott and  
223 EUROIMMUN IgG anti-N assays detected antibodies in <70% of specimens collected >90 days  
224 after index donation, while total Ig assays like the Ortho Vitros S total Ig and Roche Elecsys N  
225 total Ig assays detected antibodies in 100% of specimens at these timepoints. Given the relatively  
226 small number of donors in the cohort, the declining detection rates at later timepoints were  
227 generally not statistically distinguishable from sensitivity at earlier timepoints for these  
228 qualitative assays.

229 Regression models of quantitative signal intensity over time showed statistically  
230 significant declining reactivity in some assays. All anti-S total Ig (“direct” antigen sandwich  
231 format) assays showed stable or increasing reactivity, while all IgG assays showed declining  
232 reactivity over time (Figure 4, panel A). Anti-N assays showed more rapid waning than anti-S  
233 assays, with multivariable regression confirming that both assay format and antigen (Ag) target  
234 are important rate of waning predictors. Amongst assays that showed statistically significant  
235 declining reactivity, estimated half-lives varied from 41 to 574 days (median: 91 days)(Figure 4,  
236 panel B).

237 All assays included in the study showed ‘good’ or ‘excellent’ quantitative repeatability as  
238 reflected in the ICC, i.e., ICCs  $\geq 0.75$  and  $\geq 0.9$  [10], respectively, with the exception of the  
239 Wantai assay that had an ICC below 0.6 (Figure 5, Appendix Table 2). CVs were generally  
240  $< 10\%$  for low and medium titer blinded replicate specimens, and somewhat higher for high titer  
241 specimens, ranging from  $\sim 20\%$  to over  $100\%$  (Appendix Table 3). The Ortho VITROS anti-S  
242 and Roche Elecsys anti-N total Ig assays had notably low CVs on most replicate specimens  
243 (generally below  $10\%$ ).

244 Dilutional performance was generally good, with most assays demonstrating reasonable  
245 linearity in the relationship between expected and observed reactivity above the assay cutoff  
246 (Appendix Figure 5). Assays with greater dynamic ranges tended to show a linear dilutional  
247 response even below the cutoff. Most assays had a well-defined inflection point, representing a  
248 level of reactivity below which the dilutional response was not linear.

249 For most assays all 24 serosilent specimens were non-reactive, 7 assays had 1/24  
250 specimen reactive, 2 assays had 2/24 specimens reactive. The vast majority of apparent serosilent  
251 specimens were not detected using any of the serological assays included in this study (Appendix

252 Figure 6). For the single seroconversion series, most assays show seroconversion over the same  
253 two-week timeframe, providing little evidence of variable sensitivity relative to time of infection  
254 (Appendix Figure 7).

## 255 **Discussion**

256 This study characterized 21 commercial SARS-CoV-2 serological assays, supporting the  
257 development, validation, and implementation of testing algorithms for serosurveillance  
258 programs, including algorithms that can distinguish natural infection from vaccine induced  
259 seroreactivity.

260 The three most critical characteristics for assays used to conduct serosurveillance are 1)  
261 sensitivity including an assay's ability to detect antibodies following asymptomatic and mildly  
262 symptomatic infections potentially resulting in weak Ab responses [11-13], 2) specificity to  
263 minimize the impact of false positives on seroprevalence estimates, and 3) the ability to durably  
264 detect Ab responses for accurate estimation of cumulative infections. For serosurveillance in the  
265 context of widespread spike-based vaccine implementation, algorithms that combine assays with  
266 these characteristics and different Ag targets can differentiate natural infection from vaccine  
267 induced seroreactivity. The ability to detect reinfections and vaccine breakthrough infections  
268 based on development of anti-N reactivity, including those from variants,[14-18] requires  
269 quantitative assays with wide dynamic ranges, including the ability to extend the dynamic range  
270 through dilution, and good quantitative repeatability to enable detection of anamnestic boosting  
271 of humoral immune responses. However in regions where whole virus vaccines are in use,  
272 alternative algorithms should be considered.

273 The impact of particular performance characteristics on interpretation of serosurveillance  
274 data is context dependent. Ideal assays for serosurveillance applications (eg total Ig assays

275 demonstrating excellent specificity, sensitivity and durability of Ab detection) may not be a  
276 viable option and other factors such as cost, logistics and regulatory process may influence assay  
277 availability and selection, particularly in resource constrained settings. However, if assay  
278 performance has been robustly characterized, statistical adjustments can be made in the  
279 estimation of seroprevalence.

280         It is common practice for assay manufacturers to determine sensitivity based on timing of  
281 seroconversion relative to diagnostic testing or clinical disease. Because this clinical diagnostic  
282 definition of sensitivity may not be the most relevant criterion in cases where there is a high rate  
283 of mildly symptomatic or asymptomatic infections, alternate definitions of true positivity should  
284 be considered. Thus, in this study focused on serosurveillance applications, we used multiple  
285 definitions to assess sensitivity. These definitions allowed us to assess sensitivity in practical  
286 serosurveillance contexts. Of particular note, the inclusion of all CCP donors results in lower  
287 sensitivity estimates consequent to inclusion of serosilent infection cases, whereas the  
288 requirement for neutralization activity excludes those cases resulting in higher sensitivity  
289 estimates.

290         The ability to detect past infections long after the resolution of symptoms is key to  
291 accurately estimate cumulative incidence of infections based on seroreactivity rates; otherwise,  
292 complex and unvalidated adjustments for seroreversion may be required[17, 19, 20]. To evaluate  
293 the durability of humoral immunity after natural infection and vaccination, and to detect  
294 anamnestic boosting of Abs following reinfection or vaccine breakthrough, it is necessary to  
295 detect both changes in quantitative signal intensity and qualitative results over time, and this  
296 requires assays with wide dynamic range and quantitative precision. Seroreactivity levels on  
297 such assays may still plateau at the upper limit of quantitation. Dilution of specimens, which

298 many platforms can perform automatically, extend dynamic ranges enabling quantitation of high  
299 titer specimens as demonstrated by the Bio-Rad BioPlex assay. Furthermore, rates of waning  
300 immunity are difficult to assess using assays with narrow dynamic ranges that constrain  
301 detection of declining reactivity, which may persist at the upper limit of quantification. Although  
302 qualitative seroreversion was observed over the timescale evaluated in some assays including  
303 ones with narrow dynamic range (Figure 3), further studies are required to assess the durability  
304 of detection over longer timescales. Quantitation of very low-level reactivity is possible in assays  
305 demonstrating linearity of dilutional performance below the manufacturer defined thresholds for  
306 reactivity, which are generally set to maintain high specificity. Quantitation of high-level  
307 reactivity requires assays with a wide dynamic range, or testing of dilutions to extend the  
308 measurement range, which is more practical on platforms that support on-board dilutions.

309 We observed that all anti-S total Ig (“direct” antigen sandwich format) assays showed  
310 stable or increasing reactivity, while all but one IgG assays showed declining reactivity over time  
311 presumably due to continued maturation of Ab affinity and/or avidity resulting in increasing  
312 signal intensity in these assays [21-23]. Anti-N assays showed more rapid waning than anti-S  
313 assays, with multivariable regression confirming that both assay format and Ag target are  
314 important rate of waning predictors.

315 The relatively stable detection of neutralizing activity up to four months post index  
316 donation demonstrates that in the cross-sectional CCP sample set used in the sensitivity analysis,  
317 any waning of nAb titers was very unlikely to have taken place by the time specimens were  
318 collected and would therefore not have biased sensitivity analyses based on neutralizing activity.

319 Although there was sporadic reactivity in a few specimens from serosilent cases, most  
320 assays included in this evaluation tested non-reactive on all specimens. This corroborates the

321 findings of other studies [11, 24] indicating that some infected individuals do not develop a  
322 detectable systemic humoral immune response to SARS-CoV-2 infection.

323         The best performing assays for serosurveillance applications in this evaluation were high-  
324 throughput total Ig antigen sandwich format assays, as they met the three key performance  
325 criteria of durable Ab detection, sensitivity and specificity. The Ortho and Roche total Ig assays  
326 that target S and N antibodies, performed well and are currently employed in large scale  
327 serosurveillance studies in the US, Canada, the UK and other countries, including the CDC-  
328 nationwide blood donor seroprevalence study ([COVID Data Tracker](#)). The Wantai assay has  
329 been widely used in serosurveillance globally [25-27]; while this demonstrated lower specificity  
330 and reproducibility than the best performing assays, it performs adequately for serosurveillance  
331 with accounting for those limitations. Several other assays, including the Abbott IgG anti-N and  
332 EUROIMMUN IgG anti-S assays, have been employed in large-scale serosurveillance, but  
333 require adjustments for rapid waning and seroreversion to estimate cumulative incidence or  
334 attack rates, especially over longer periods and multiple epidemic waves. This study provides  
335 critical data that can be applied to adjust for waning in other studies.

336         This study has several limitations. Asymptomatic cases are underrepresented in the panel  
337 as CCP donors qualify based on recovery from symptomatic infection, potentially resulting in  
338 overestimation of sensitivity. The assessment of durability of bAb detection is based on CCP  
339 donations from donors whose continued qualification required ongoing Ortho VITROS Total Ig  
340 anti-S1 reactivity. Although these CCP donors do not have documented dates of NAT-positivity,  
341 symptom onset or resolution, the first donations were generally within 1-2 months of symptom  
342 resolution [5]. To address these limitations we developed approaches to adequately characterize

343 sensitivity and durability of reactivity. The number of specimens included in the dilutional series  
344 subpanels are not sufficient for robust assessment of endpoint dilutional sensitivity.

345         This study provides a standardized, comparative assessment of 21 SARS-CoV-2 Ab  
346 assays from major commercial manufacturers and allows for identification of optimal assays and  
347 testing algorithms for serosurveillance applications in various contexts. These results also  
348 provide performance data applicable to other serological testing use cases relevant to clinicians,  
349 public health organizations, laboratorians, and emergency response planners.

350

351

352

353 Acknowledgements: We thank the manufacturers for participation and/or provision of reagents:  
354 Ortho Clinical Diagnostics, Euroimmun, Roche Diagnostics, Diasorin, Siemens, Abbott, Bio-  
355 Rad, Quotient, Diazyme, Beckman Coulter, Wantai; the VRI Research Operations Core and Core  
356 Immunology Labs, and the testing laboratories.

357 **Funding source:** This work was supported by research contracts from the Centers for Disease  
358 Control and Prevention (CDC Contract 75D30120C08170).

359

360



361 **Tables and Figures**

Manufacturer	Assay	Ig Target	Antigen	Assay Format	Reported units	Testing Lab
<b>Ortho</b>	VITROS Immunodiagnostic Products Anti-SARS-CoV-2 Total Ig	Total Ig	S1	Double-antigen sandwich CLIA	S/CO	Vitalant Research Institute
	VITROS Immunodiagnostic Products Anti-SARS-CoV-2 IgG	IgG	S1	Double-antigen sandwich CLIA	S/CO	CTS
<b>EUROIMMUN</b>	Anti-SARS-CoV-2 NCP ELISA	IgG	N	Indirect IgG EIA	S/CO	Advent Health
	Anti-SARS-CoV-2 ELISA	IgG	S1	Antigen sandwich ELISA	S/CO	
	Anti-SARS-CoV-2 QuantiVac ELISA	IgG	S1	Antigen sandwich ELISA	RU/ml	
	Anti-SARS-CoV-2 ELISA	IgA	S1	Antigen sandwich ELISA	S/CO	
<b>Roche</b>	Elecsys Anti-SARS-CoV-2 N on cobas	Total Ig	N	Double-antigen sandwich CLIA	COI	University of California, Davis
	Elecsys Anti-SARS-CoV-2 S on cobas	Total Ig	S1/S2/RBD	Double-antigen sandwich CLIA	U/ml	
<b>DiaSorin</b>	LIAISON 28 SARS-CoV-2 TrimericS IgG	IgG	TrimericS	IgG magnetic particle CLIA	AU/ml	British Columbia Centers for Disease Control and Prevention
<b>Siemens</b>	ADVIA Centaur SARS-CoV-2 Total Ig	Total Ig	S1/RBD	Ag sandwich CLIA	S/CO	
	ADVIA Centaur SARS-CoV-2 IgG	IgG	S1/RBD	Ag sandwich CLIA	Index	
<b>Abbott</b>	SARS-CoV-2 IgG N on ARCHITECT	IgG	N	CMIA	AU/ml	Duke Human Vaccine institute
	SARS-CoV-2 IgG S1 on ARCHITECT	IgG	S1	CMIA	S/CO	
	SARS-CoV-2 IgG N on Alinity	IgG	N	CMIA	S/CO	Fred Hutchinson Cancer Research Center
	SARS-CoV-2 IgG S1 on Alinity	IgG	S1	CMIA	AU/ml	
<b>Bio-Rad</b>	Platelia SARS-CoV-2 Total Ab (Evolis)	Total Ig	N	One-step antigen capture	S/CO	BloodWorks NorthWest
	BioPlex 2200 SARS-CoV-2 IgG Panel	IgG	RBD, S1, S2, N	Multiplexed microbeads two-step assay	S/CO	
<b>Quotient</b>	MosaiQ COVID-19 Antibody Microarray	IgM/IgG	S1/S2	Array	Qual only	
<b>Diazyme</b>	DZ-Lite SARS CoV-2	IgG	N & S1/S2	IgG microbead CLIA	S/CO	University of California, Irvine
<b>Beckman Coulter</b>	Access SARS-CoV-2 IgG	IgG	S1 RBD	IgG 2-step paramagnetic particle CLIA	S/CO	
<b>Wantai</b>	SARS-CoV-2 Total Ig	Total Ig	S1 RBD	Total Ig sandwich ELISA	S/CO	Sanquin

362 **Table 1.** Serological assays and key characteristics. S, spike protein; RBD, receptor binding  
363 domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig, immunoglobulin; S/CO, signal to cutoff  
364 ratio; EIA, enzyme immunoassay; ELISA, enzyme-linked immunosorbent assay;RU, relative  
365 units, AU, arbitrary units; CLIA, Chemiluminescent immunoassay; CMIA, Chemiluminescent  
366 microparticle immunoassay.

367 Current US regulatory status can be obtained on the FDA website at:

368 [https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-  
authorizations-medical-devices/in-vitro-diagnostics-euas-serology-and-other-adaptive-immune-  
response-tests-sars-cov-2](https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-<br/>369 authorizations-medical-devices/in-vitro-diagnostics-euas-serology-and-other-adaptive-immune-<br/>370 response-tests-sars-cov-2) .

371

372

373

374

375

<i>Group</i>	<i>Description</i>	<i>Number of specimens</i>
<i>Sensitivity sub-panel</i>		
	130 cross sectional CCP donor plasma specimens.	185
<i>Specificity sub-panel</i>		
	Pre-pandemic blood donor specimens collected prior to 2020 and demonstrated to be anti-SARS-CoV-2 negative by RVP neutralization testing.	459
<i>Ab persistence sub-panel</i>		
	Longitudinal specimens from 26 donors with at $\geq 4$ CCP donations 84-150 days post index donation.	209
<i>Seroconversion sub-panel</i>		
	Longitudinal specimens from a single source plasma donor with acute SARS-CoV-2 infection	14
<i>Dilutional Performance sub-panel</i>		
	Serial dilutions of five specimens from Sensitivity sub-panel; neat (6 replicates), 1:40, 1:80, 1:160, 1:320, 1:640 analogous to nAb testing	55
<i>Serosilent cases</i>		
	Individual CCP donors non-reactive by S and N anti-SARS-CoV-2 Total Ig	27
<i>Repeatability sub-panel</i>		
	Six blinded replicates each of 15 CCP specimens	90

376

377 **Table 2.** Composition of the assessment panel. CCP, COVID convalescent plasma, S, spike  
 378 protein; RBD, receptor binding domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig,  
 379 immunoglobulin; RVP, reporter viral particle; bAb, neutralizing antibodies.

380

381

382 **Figure 1.** Sensitivity of SARS-CoV-2 serological assays using three definitions of a “true  
383 positive.”

384 Panel A: Positivity defined by qualification as CCP donor (excluding purposely selected  
385 serosilents)

386 Panel B: Positivity defined by neutralizing activity measured by Broad PRNT

387 Panel C: Positivity defined by ‘operational standard’ (3 or more bAb assays reactive).

388 Dots indicate point estimates and bars indicate Wilson score 95% confidence intervals.

389 *The Ortho VITROS IgG S assay is included only in panel B, because the assay required use of*  
390 *serum for testing, thus only specimens with available serum and neutralizing data were tested .*

391 S, spike protein; RBD, receptor binding domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig,  
392 immunoglobulin; PRNT, plaque reduction neutralization test. See Table 1 for assay details.

393

394 **Figure 2.** Specificity of SARS-CoV-2 serological assays in pre-COVID-19 negative control  
395 specimens.\*

396 \*Dots indicate point estimates and bars indicate Wilson score 95 confidence intervals. S, spike  
397 protein; RBD, receptor binding domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig,  
398 immunoglobulin. See Table 1 for assay details.

399

400 **Figure 3:** Proportion of donors with detectable SARS-CoV-2 antibodies in the longitudinal  
401 COVID-19 convalescent plasma (CCP) donor cohort with donations sorted into time bins  
402 relative to index CCP donation. Time bin labels on x axis are denoted with square brackets to  
403 indicate inclusive boundaries and round brackets to indicate exclusive boundaries. Donors that  
404 contributed more than one donation in a time bin contributed the fractional proportion reactive to

405 the numerator and 1 to the denominator for estimation of proportion reactive in the time bin.  
406 Symbols indicate point estimates of proportion reactive and bars indicate 95% confidence  
407 intervals (Wilson score). See Table 1 for assay details. S, spike protein; RBD, receptor binding  
408 domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig, immunoglobulin; PRNT, plaque  
409 reduction neutralization test.

410 \*Ortho VITROS Total Ig anti-S reactivity was required for qualification of continued CCP  
411 donation, and therefore shows 100% detection in all time bins by definition.

412

413 **Figure 4.** Durability of antibody detection as assessed by mixed effects regression modelling.

414 Panel A: Average ('fixed') slopes from linear mixed effects regression models with donor  
415 random effects, fit to rescaled and log-transformed quantitative assay signal.

416 Panel B: Assay signal half-lives post index donation for assays demonstrating rapid waning of  
417 seroreactivity over time (upper bound on half-life less than 220 days), estimated based on linear  
418 mixed effects regression models.

419 S, spike protein; RBD, receptor binding domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig,  
420 immunoglobulin. See Table 1 for assay details.

421

422 **Figure 5.** Intraclass correlation coefficients based on blinded replicate sample testing, reflecting

423 the proportion of total variance that is between-sample rather than within-sample variability. S,  
424 spike protein; RBD, receptor binding domain; N, nucleocapsid; Ag, antigen; Ab, antibody; Ig,  
425 immunoglobulin. See Table 1 for assay details.

426 \* Results falling outside the primary measurement range excluded. \*\* On-board dilutions were  
427 used to estimate reactivity in specimens where initial results fell outside the primary

428 measurement range. *Horizontal dotted lines show conventional (although arbitrary) thresholds*  
429 *for “moderate” (0.5), “good” (0.75) and “excellent” (0.9) repeatability [10].*

430 References

- 431 1. U.S. Food & Drug Administration. In Vitro Diagnostics EUAs. Available at:  
432 [https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-](https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas)  
433 [use-authorizations-medical-devices/vitro-diagnostics-euas](https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas).
- 434 2. Goodhue Meyer E, Simmons G, Grebe E, et al. Selecting COVID-19 convalescent  
435 plasma for neutralizing antibody potency using a high-capacity SARS-CoV-2 antibody  
436 assay. *Transfusion* **2021**.
- 437 3. U.S. Food and Drug Administration. Investigational COVID-19 Convalescent Plasma;  
438 Guidance for Industry. Available at: [https://www.fda.gov/regulatory-information/search-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/investigational-covid-19-convalescent-plasma)  
439 [fda-guidance-documents/investigational-covid-19-convalescent-plasma](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/investigational-covid-19-convalescent-plasma).
- 440 4. Belda F, Lopez-Martinez M, Torres N, Cherenzia R, Crowley M. Available COVID-19  
441 serial seroconversion panel for validation of SARS-CoV-2 antibody assays. *Diagnostic*  
442 *Microbiology and Infectious Disease* **2021**; 100(2): 115340.
- 443 5. Di Germanio C, Simmons G, Kelly K, et al. SARS-CoV-2 Antibody persistence in  
444 COVID-19 convalescent plasma donors: Dependency on assay format and applicability  
445 to serosurveillance. *Transfusion* **2021**.
- 446 6. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria:  
447 R Foundation for Statistical Computing, **2021**.
- 448 7. Dorai-Raj S. binom: Binomial Confidence Intervals For Several Parameterizations, **2014**.
- 449 8. Marschner IC. glm2: Fitting generalized linear models with convergence problems. *The*  
450 *R Journal* **2011**; 3: 12-5.
- 451 9. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York,  
452 **2016**.
- 453 10. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – A discussion and  
454 demonstration of basic features. *PLoS One* **2019**; 14(7): e0219854.
- 455 11. Petersen LR, Sami S, Vuong N, et al. Lack of antibodies to SARS-CoV-2 in a large  
456 cohort of previously infected persons. *Clinical Infectious Diseases* **2020**; Epub ahead of  
457 print.
- 458 12. Ibarrondo FJ, Fulcher JA, Goodman-Meza D, et al. Rapid Decay of Anti-SARS-CoV-2  
459 Antibodies in Persons with Mild Covid-19. *New England Journal of Medicine* **2020**.
- 460 13. Peluso MJ, Takahashi S, Hakim J, et al. SARS-CoV-2 antibody magnitude and  
461 detectability are driven by disease severity, timing, and assay. *medRxiv* **2021**:  
462 2021.03.03.21251639.
- 463 14. Garcia-Beltran WF, Lam EC, St Denis K, et al. Multiple SARS-CoV-2 variants escape  
464 neutralization by vaccine-induced humoral immunity. *Cell* **2021**.
- 465 15. Gundlapalli AV, Salerno RM, Brooks JT, et al. SARS-CoV-2 Serologic Assay Needs for  
466 the Next Phase of the US COVID-19 Pandemic Response. *Open Forum Infect Dis* **2021**;  
467 8(1): ofaa555.
- 468 16. Moore JP, Offit PA. SARS-CoV-2 Vaccines and the Growing Threat of Viral Variants.  
469 *JAMA* **2021**; 325(9): 821-2.
- 470 17. Sabino EC, Buss LF, Carvalho MPS, et al. Resurgence of COVID-19 in Manaus, Brazil,  
471 despite high seroprevalence. *The Lancet* **2021**.
- 472 18. Babiker A, Marvil C, Waggoner JJ, Collins M, Piantadosi A. The Importance and  
473 Challenges of Identifying SARS-CoV-2 Reinfections. *Journal of clinical microbiology*  
474 **2020**.

- 475 19. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain  
476 (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet*  
477 **2020**; 396(10250): 535-44.
- 478 20. Shioda K, Lau MS, Kraay AN, et al. Estimating the cumulative incidence of SARS-CoV-  
479 2 infection and the infection fatality ratio in light of waning antibodies. *medRxiv* **2020**.
- 480 21. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-  
481 CoV-2 lineage in Manaus, Brazil. *Science* **2021**; 372(6544): 815-21.
- 482 22. Liu H, Wu NC, Yuan M, et al. Cross-Neutralization of a SARS-CoV-2 Antibody to a  
483 Functionally Conserved Site Is Mediated by Avidity. *Immunity* **2020**; 53(6): 1272-80.e5.
- 484 23. Benner SE, Patel EU, Laeyendecker O, et al. SARS-CoV-2 Antibody Avidity Responses  
485 in COVID-19 Patients and Convalescent Plasma Donors. *The Journal of infectious*  
486 *diseases* **2020**; 222(12): 1974-84.
- 487 24. Gallais F, Velay A, Nazon C, et al. Intrafamilial Exposure to SARS-CoV-2 Associated  
488 with Cellular Immune Response without Seroconversion, France. *Emerging Infectious*  
489 *Disease journal* **2021**; 27(1): 113.
- 490 25. Gudbjartsson DF, Norddahl GL, Melsted P, et al. Humoral Immune Response to SARS-  
491 CoV-2 in Iceland. *The New England journal of medicine* **2020**.
- 492 26. Slot E, Hogema BM, Reusken C, et al. Low SARS-CoV-2 seroprevalence in blood  
493 donors in the early COVID-19 epidemic in the Netherlands. *Nat Commun* **2020**; 11(1):  
494 5744.
- 495 27. Jespersen S, Mikkelsen S, Greve T, et al. SARS-CoV-2 seroprevalence survey among  
496 17,971 healthcare and administrative personnel at hospitals, pre-hospital services, and  
497 specialist practitioners in the Central Denmark Region. *Clin Infect Dis* **2020**.  
498

















