

1 Side by side comparison of three fully automated SARS-CoV-2 2 antibody assays with a focus on specificity

3 *Running head: Specificity of automated Anti-SARS-CoV-2 assays*

4
5 Thomas Perkmann¹, Nicole Perkmann-Nagele¹, Marie-Kathrin Breyer², Robab Breyer-Kohansal², Otto C Burghuber³,
6 Sylvia Hartl^{2,3}, Daniel Aletaha⁴, Daniela Sieghart⁴, Peter Quehenberger¹, Rodrig Marculescu¹, Patrick Mucher¹, Robert
7 Strassl¹, Oswald F Wagner¹, Christoph J Binder¹, and Helmuth Haslacher^{1*}

8
9 ¹Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

10 ²Department of Respiratory and Critical Care Medicine and Ludwig Boltzmann Institute for COPD and Respiratory Epidemiology, Otto Wagner
11 Hospital, Vienna, Austria

12 ³Sigmund Freud University, Medical School and Ludwig Boltzmann Institute for COPD and Respiratory Epidemiology, Vienna, Austria

13 ⁴Divison of Rheumatology, Department of Medicine III, Medical University of Vienna, Vienna, Austria

14

15 *Please address all correspondence to:

16 Helmuth Haslacher, MD PhD MSc BSc BA

17 Medical University of Vienna, Department of Laboratory Medicine

18 Waehringer Guertel 18-20, A-1090 Vienna

19 Phone: +43 1 40400 53190

20 Fax: +43 1 40495 15547

21 E-Mail: helmuth.haslacher@meduniwien.ac.at

22

23 Key words:

24 SARS-CoV-2; serology; specificity; laboratory automation; positive predictive value;
25 seroprevalence;

26

27 List of abbreviations:

28 COVID-19	Coronavirus disease 2019
29 SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
30 RT-PCR	Reverse transcriptase-polymerase chain reaction
31 ECLIA	Electrochemiluminescence assay
32 95% CI	95% confidence interval
33 CMIA	Chemiluminescent microparticle immunoassay
34 NPA	Negative percentage agreement
35 CLIA	Chemiluminescence immunoassay
36 ROC	Receiver-operating-characteristics
37 AUC, AUROC	Area under the (ROC-)curve
38 LOD	Limit of detection
39 PPV	Positive predictive value
40 NPV	Negative predictive value
41 nAbs	Neutralizing antibodies
42 EUA	Emergency Use Authorization

43 **ABSTRACT**

44 **Background:** In the context of the COVID-19 pandemic, numerous new serological
45 test systems for the detection of anti-SARS-CoV-2 antibodies have become available
46 quickly. However, the clinical performance of many of them is still insufficiently
47 described. Therefore we compared three commercial, CE-marked, SARS-CoV-2
48 antibody assays side by side.

49 **Methods:** We included a total of 1,154 specimens from pre-COVID-19 times and 65
50 samples from COVID-19 patients (≥ 14 days after symptom onset) to evaluate the test
51 performance of SARS-CoV-2 serological assays by Abbott, Roche, and DiaSorin.

52 **Results:** All three assays presented with high specificities: 99.2% (98.6-99.7) for
53 Abbott, 99.7% (99.2-100.0) for Roche, and 98.3% (97.3-98.9) for DiaSorin. In contrast to
54 the manufacturers' specifications, sensitivities only ranged from 83.1% to 89.2%.
55 Although the three methods were in good agreement (Cohen's Kappa 0.71-0.87),
56 McNemar's test revealed significant differences between results obtained from Roche
57 and DiaSorin. However, at low seroprevalences, the minor differences in specificity
58 resulted in profound discrepancies of positive predictability at 1% seroprevalence: 52.3%
59 (36.2-67.9), 77.6% (52.8-91.5), and 32.6% (23.6-43.1) for Abbott, Roche, and DiaSorin,
60 respectively.

61

62 **Conclusion:** We find diagnostically relevant differences in specificities for the anti-
63 SARS-CoV-2 antibody assays by Abbott, Roche, and DiaSorin that have a significant
64 impact on the positive predictability of these tests.

65

66 **Introduction**

67 COVID-19 is a new disease caused by Severe Acute Respiratory Syndrome
68 Coronavirus 2 (SARS-CoV-2), which was first described by Chinese scientists in early
69 January 2020 (1). On March 11, the WHO officially declared the novel SARS-CoV-2
70 infections a pandemic, which has now spread rapidly across the entire globe, with
71 almost 6.5 million confirmed cases and over 375,000 confirmed deaths (2). COVID-19 is
72 characterized by a broad spectrum of individual disease courses, ranging from
73 asymptomatic infections to the most severe cases requiring intensive medical care (3).
74 The reliable detection of infected persons and, subsequently, their isolation is essential
75 for the effort to prevent the spread of the SARS-CoV-2 virus quickly and efficiently.
76 Therefore, reverse transcriptase-polymerase chain reaction (RT-PCR) testing is required
77 for direct detection of the pathogen. Unfortunately, RT-PCR testing does not always give
78 a clear answer to whether the SARS-CoV-2 infection is currently present or not (4,5).
79 On the other hand, serological testing for SARS-CoV-2 specific antibodies can be used
80 as an additional diagnostic tool in case of suspected false-negative RT-PCR results (6)
81 or for individual determination of antibody levels. Moreover, cross-sectional serological
82 studies provide essential epidemiological information to allow a correct estimation of the
83 spread of the disease within a population (7,8). The first commercially available
84 serological SARS-CoV-2 tests, mostly standard ELISA tests or lateral flow rapid tests,
85 have not always proved to be sufficiently specific and sensitive (9,10). Recently, the first
86 tests for fully automated large-scale laboratory analyzers have been launched. The
87 present evaluation aims to compare three of these test systems manufactured by Abbott
88 (11), Roche (12), and DiaSorin (13), with particular emphasis on specificity, which is

89 crucial for an adequate positive predictive value given the current low seroprevalence
90 worldwide.

91 **Materials and methods**

92 **Study design and patient cohorts**

93 This nonblinded prospective study aims at a detailed comparison of three automated
94 SARS-CoV-2 detection methods with a particular focus on specificity and positive
95 predictability. A total of 1,154 samples from three cohorts of patients/participants with
96 sampling dates before 01.01.2020 were used to test specificity. The samples derived
97 from three different collections: a cross-section of the Viennese population, LEAD study
98 (14), preselected for samples collected between November and April to enrich seasonal
99 infections (n=494); a collection of healthy voluntary donors (n=302; 269 individuals, 11
100 donors with a 4-fold repetition of the donation within a median period of 4.5 years [3.6-
101 5.5]); a disease-specific collection of samples from patients with rheumatic diseases
102 (n=358).

103 For estimation of test sensitivity, samples of 65 COVID-19 donors/patients with a
104 symptom onset to analysis time of ≥ 14 days (median time interval of 41 [28-49] days)
105 were evaluated in parallel on all three analysis platforms. For asymptomatic donors
106 (n=6), SARS-CoV-2 RT-PCR confirmation to analysis time was used instead. We
107 subjected only a single serum sample per patient to sensitivity analysis to avoid data
108 bias due to uncontrolled multiple measurement points of individual patients.

109 Supplementary Table 1 gives a comprehensive overview of characteristics and cohort-
110 specific inclusion and exclusion criteria; Supplementary Table 2, 3, and Supplementary
111 Fig. 1 provide additional descriptive statistics on donors/patients included in the cohorts.
112 All included participants gave written informed consent for donating their samples for
113 scientific purposes. From patients, only left-over material from diagnostic procedures
114 was used. The overall evaluation plan conformed with the Declaration of Helsinki as well
115 as with relevant regulatory requirements. It was reviewed and approved by the ethics
116 committee of the Medical University of Vienna (1424/2020).

117 **Biomaterials**

118 Used serum samples were either left-over materials from diagnostic procedures
119 (Department of Laboratory Medicine, Medical University of Vienna) or part of a sample
120 cohort processed and stored by the MedUni Wien Biobank. All pre-analytical processes
121 were carried out according to standard operating procedures in an ISO 9001:2008/2015-
122 certified (MedUni Wien Biobank, Department of Laboratory Medicine) and ISO
123 15189:2012-accredited (Department of Laboratory Medicine) environment. Standard
124 sample protocols were described previously (15).

125 **Antibody testing**

126 SARS-CoV-2 specific antibodies were measured according to the manufacturers'
127 instructions on three different automated platforms at the Department of Laboratory
128 Medicine of the Medical University of Vienna.

129 1. IgG antibodies against SARS-CoV-2 nucleocapsid (SARS-CoV-2 IgG) were quantified
130 employing a chemiluminescent microparticle immunoassay (CMIA) on the Abbott

131 ARCHITECT® i2000sr platform (Abbott Laboratories, Chicago, USA). For the cut-off of
132 ≥ 1.4 Index (S/C), the manufacturer gives a negative percentage agreement (NPA,
133 corresponding to specificity) of 99.60% (95% CI: 98.98 – 99.89) calculated from samples
134 collected before the COVID-19 outbreak and of 100.00% (95.07 – 100.00) in patients
135 with other respiratory illnesses. Regarding diagnostic sensitivity, 0.00% (0.00 – 60.24)
136 are reported <3 days after symptom onset, 25.00% (3.19 – 65.09) on days 3 – 7,
137 86.36% (65.09 – 97.09) on days 8 – 13, and 100.00% (95.89 – 100.00) from day 14 on
138 (11). According to the manufacturer, the assay is designed for the qualitative detection
139 of IgG antibodies to SARS-CoV-2.

140 2. The Elecsys® Anti-SARS-CoV-2 assay (Roche Diagnostics, Rotkreuz, Switzerland)
141 was applied on a Cobas e 801 modular analyzer. It detects total antibodies against the
142 SARS-CoV-2 nucleocapsid (N) antigen in a sandwich electrochemiluminescence assay
143 (ECLIA). For the suggested cut-off of ≥ 1 COI, the manufacturer reports specificities of
144 99.80% (95% CI: 99.58 – 99.92) for samples derived from diagnostic routine, 99.83%
145 (99.51 – 99.97) for blood donors, and 100% (91.19 – 100) for both a common cold panel
146 and a Coronavirus panel. Sensitivity was estimated as 65.5% (56.1 – 74.1) during days
147 0 – 6 post RT-PCR confirmation, 88.1% (77.1 – 95.1) from day 7 to day 13, and 100%
148 (88.1 – 100%) from day 14 on (12). According to the manufacturer, the system delivers
149 qualitative results, either being reactive or non-reactive for anti-SARS-CoV-2 antibodies.

150 3. The LIAISON® SARS-CoV-2 S1/S2 IgG test detects IgG-antibodies against the
151 S1/S2 domains of the virus' spike protein in a chemiluminescence immunoassay (CLIA).
152 The test was applied to a LIAISON® XL Analyzer (DiaSorin S.p.A., Saluggia, Italy). The
153 manufacturer reports a diagnostic specificity of 98.5% (95% CI: 97.5 – 99.2) in blood
154 donors and 98.9% in presumably SARS-CoV-2 negative diagnostic routine samples

155 (94.0 – 99.8). Applying a cut-off >15.0 AU/mL (borderline results 12.0 – 15.0, require a
156 re-test algorithm), the test's sensitivity is reported time-dependently with 25.0% (14.6 –
157 39.4) ≤5 days after RT-PCR-confirmed diagnosis, 90.4% (79.4 – 95.8) from day 5 to day
158 15, and 97.4% (86.8 – 99.5) after >15 days (13). Samples that repeatedly tested
159 borderline were classified as positive. The manufacturer indicates to provide quantitative
160 measurement results on the system.

161 **Statistical analysis**

162 Unless stated otherwise, continuous data are given as median (quartile 1 – quartile 3).
163 Categorical data are given as counts and percentages. Diagnostic sensitivity and
164 specificity, as well as positive and negative predictive values, were calculated using
165 MedCalc software 19.2.1 (MedCalc Ltd., Ostend, Belgium). 95% confidence intervals
166 (CI) for sensitivity and specificity were calculated according to Clopper and Pearson
167 ("exact" method) with standard logit confidence intervals for the predictive values (16).
168 Receiver-Operating-Characteristic (ROC)-curve analysis was used to evaluate test
169 accuracy and compare the diagnostic performance of the three test systems, according
170 to DeLong et al. (17). Between-test agreements were assessed by interpretation of
171 Cohen's Kappa-statistics, and further evaluated with McNemar's tests. Statistical
172 significance was assumed at $p < 0.05$. Figures were produced with MedCalc software
173 19.2.1 and GraphPad Prism 8 (GraphPad Software, San Diego, USA).

174

175

176 **Results**

177 **Specificity**

178 To describe assay specificity, we used a total of 1,154 serum samples collected before
179 SARS-CoV-2 circulated in the population and which are, by definition, negative for
180 SARS-CoV-2 specific antibodies. The three different specificity cohorts A-C (described
181 in detail in Supplementary Table 1-3 and Supplementary Figure 1) presented with
182 different rates of false-positives (Table 1) - cohort C (cohort of rheumatic diseases)
183 showing the highest reactivities. We found in total 9, 3, and 20 false-positive samples for
184 Abbott, Roche, and DiaSorin, leading to an assay specificity of 99.2 (95%CI: 98.6-99.7),
185 99.7% (95%CI: 99.2-100.0), and 98.3% (95%CI: 97.3-98.9) respectively (Figure 1A-C).
186 Median and 90th percentile values of negative samples were 0.025 and 0.115 Index for
187 Abbott, 0.0815 and 0.0927 COI for Roche, and below LOD and 5.52 AU/ml for
188 DiaSorin. False-positive samples yielded median values of 2.21 Index (2.14-2.67) for
189 Abbott SARS-CoV-2 IgG (cut-off: ≥ 1.4 Index), 1.65 COI (1.47-1.72) for Roche Elecsys®
190 Anti-SARS-CoV-2 (cut-off: ≥ 1 COI), and 22.4 AU/ml (17.38-57.35) for DiaSorin
191 LIAISON® SARS-CoV-2 S1/S2 IgG (negative < 12.0 AU/ml, equivocal 12.0 AU/ml $\leq x$
192 < 15.0 AU/ml, positive ≥ 15.0 AU/ml).

193 **Sensitivity**

194 To estimate assay sensitivity, we used serum samples from 65 donors/patients at later
195 time points following SARS-CoV-2 infection, at least ≥ 14 days after symptom onset
196 (median interval of 41 [28-49] days). In this late phase, we assumed the majority of
197 donors/patients having reached prominent and constant levels of SARS-CoV-2 specific

198 antibodies. Surprisingly, we could find a relatively high percentage of samples that were
199 testing negative for SARS-CoV-2 antibodies: Abbott 10, Roche 7, and DiaSorin 11 false-
200 negatives leading to calculated sensitivities of 84.6% (73.6-92.4), 89.2% (79.1-95.6),
201 and 83.1 (71.3-91.2)%, respectively. Five serum samples were consistently tested
202 negative in all three assays despite being derived from individuals tested positive for
203 SARS-CoV-2 by RT-PCR. All seven false-negatives in the Roche test overlapped with
204 false-negatives in the Abbott test (both nucleocapsid-antigen based assays), whereas
205 DiaSorin was negative for an additional six serum samples exclusively (S1/S2-domain
206 antigen-based assay).

207 **PPV and NPV**

208 Although specificity and sensitivity are essential criteria for assessing the quality of a test
209 procedure, they have little informative value about the probability of a positive/negative
210 test result, indicating the presence/absence of SARS-CoV-2 specific antibodies without
211 taking prevalence into account. Therefore, a comparative overview for specificity,
212 sensitivity, as well as positive and negative predictive values at 1%, 5%, and 10%
213 SARS-CoV-2 antibody seroprevalence is shown in Figures 2A and 2B and summarized
214 in Table 2. While the differences between the test systems for varying seroprevalences
215 do not have a significant impact on NPV (range 98.1%-99.9%), the consequences for
216 PPV are pronounced. At seroprevalence rates of 10%, all three systems show
217 acceptable PPVs of 92.3%, 97.4%, and 84.2% for Abbott, Roche, and DiaSorin, but at
218 1% seroprevalence these drop to unsatisfactorily or even unacceptably low values of
219 52.3% (36.2-67.9), 77.6% (52.8-91.5), and 32.6% (23.6-43.1) for Abbott, Roche, and
220 DiaSorin.

221 **ROC Curve Analysis**

222 As shown in Figure 3A-C, all three ROC curves presented with areas under the curves
223 (AUC) above 0.97 (Abbott: 0.994 [95% CI: 0.987-0.997], Roche: 0.989 [0.981-0.994],
224 DiaSorin: 0.977 [0.967-0.985]). Comparison of ROC-AUCs, according to DeLong et al.,
225 did not reveal significant differences (Differences: Abbott/Roche $p=0.487$,
226 Abbott/DiaSorin $p=0.112$, Roche/DiaSorin $p=0.395$). In the next step, we aimed to
227 assess whether modifying the cut-off values could improve the explanatory power of the
228 ROC-curves.

229 Cut-offs associated with the Youden's index (maximum sum of sensitivity and specificity)
230 of >0.42 , >0.355 , and >8.76 for Abbott, Roche, and DiaSorin lowered the PPV
231 considerably, being as low as 26.4 (20.3-33.6), 62.1 (43.9-77.4), and 24.8 (18.9-31.9) at
232 1% seroprevalence for Abbott, Roche, and DiaSorin (Supplementary Table 4).

233 **Between-test agreement/disagreement**

234 Correlation analysis of measurement values between the different platforms showed
235 only moderate to weak concordance. The Pearson correlation coefficient was $r=0.66$
236 ($p<0.001$), for both Abbott/DiaSorin (both IgG-assays) and Abbott/Roche (both
237 nucleocapsid-antigen based assays). In contrast, Roche/DiaSorin, with a coefficient of
238 $r=0.25$ and a hardly reached significance of 0.044, could only show a very weak
239 correlation.

240 Therefore, the test systems' agreements were studied in a pairwise fashion applying
241 inter-rater agreement statistics (Cohen's Kappa). The agreement between Abbott and
242 Roche was very good (0.87 [0.81-0.94]). Agreement between Abbott and DiaSorin, and
243 DiaSorin and Roche was good: 0.71 (0.62-0.80), and 0.76 (0.67-0.84), respectively

244 (Table 3). Despite a good overall inter-rater agreement, significant differences could be
245 shown using McNemar's test for DiaSorin and Roche (Supplementary Table 5).

246 **Discussion**

247 To the best of our knowledge, this is the first side-by-side comparison of three fully
248 automated SARS-CoV-2 antibody tests applying more than 1,200 distinct donor/patient
249 samples. We identified significant differences between two of the three systems,
250 especially regarding positive predictability at the expectable low prevalence rates.
251 SARS-CoV-2 is a new virus closely related to the betacoronaviruses SARS-CoV and
252 MERS. Like SARS-CoV-2, those highly virulent pathogens cause severe respiratory
253 syndromes, often with lethal outcome (18). In contrast, infections with other members of
254 the coronavirus family usually present with mild colds, including 229E, OC43, NL63, and
255 HKU1 (19). Compared to SARS-CoV (which is no longer circulating), cross-reactivity
256 between SARS-CoV-2 and endemic seasonal coronaviruses is low. To date, with few
257 exceptions (24), no accumulation of cross-reactivities between anti-SARS-CoV-2
258 antibodies and seasonal coronavirus antibodies has been found. We have therefore
259 refrained from screening a coronavirus panel for possible cross-reactivity.

260

261 *Specificity*

262 To best describe the specificity of a serological test, it is essential to have a reliable
263 reference, i.e., to ensure that the samples used are negative for the target analyte. For
264 SARS-CoV-2, this means using serum/plasma samples obtained before the first
265 appearance of the new virus. Therefore, we have compiled large pre-COVID-19 cohorts,
266 which have the following characteristics: A) samples of an age and sex-controlled

267 population-based cohort of more than 11,000 participants (LEAD-Study) (14), randomly
268 chosen from Vienna and surrounding areas (n=494). B) samples of healthy voluntary
269 donors (n=302), which are typically used at our Department for the evaluation of new
270 assays, and C) samples of a disease-specific collection of patients with rheumatic
271 diseases including rheumatoid arthritis and systemic lupus erythematoses (n=358),
272 known to have a high prevalence of autoantibodies and other atypical immune activities,
273 enhancing the potential of interference with serological testing. We found several false-
274 positives in the rheumatological cohort (n=13), and to a lesser extent in the other two
275 cohorts (n=9 in the healthy donor cohort and n=10 in the LEAD study). Notably, an
276 overlap of samples tested false-positive in the different systems did not typically occur,
277 and only one of 32 samples tested positive in more than one assay (Abbott and
278 DiaSorin, one sample from the LEAD study). Since these two test systems use different
279 antigens (nucleocapsid vs. S1/S2 proteins) but the same detection method (IgG), this
280 false-positive reaction is likely associated with interference of the IgG measurement.
281 Calculated specificities are strongly dependent on the spectrum and the size of a
282 selected specificity cohort. If we calculated the specificities of each cohort separately,
283 we would be able to report variable specificities: cohort A (Abbott 99.2%, Roche 100%,
284 DiaSorin 98.8%), cohort B (Abbott 99%, Roche 99.7%, DiaSorin 98.3%), and cohort C
285 (Abbott 99.4%, Roche 99.4%, DiaSorin 97.5%). Roche would range from ideal 100%
286 down to 99.4%, the same level as the best result for Abbott, and DiaSorin would be
287 nearly as good as the worst Abbott specificity or show a 2.5% difference to the best
288 Roche value. This would have an enormous impact on prevalence dependent
289 parameters like PPV. A recent evaluation of the DiaSorin LIAISON® SARS-CoV-2
290 S1/S2 IgG assay with 1,140 pre-COVID-19 samples reported a specificity of 98.5% (20),

291 nearly perfectly matching the specificity of 98.3% we found when calculating the average
292 of all three cohorts. In contrast, another recent study reported a specificity of 100% for
293 DiaSorin. However, the authors used only n=81 samples for specificity testing (21).
294 Similarly, a further evaluation comparing all three SARS-CoV-2 tests by Abbott, Roche,
295 and DiaSorin found quite different specificities, namely 100%, 98%, and 96.9% for
296 Abbott, Roche, and DiaSorin, respectively. Again, the specificity cohort was very small
297 (n=100, and n=98 for DiaSorin) (22). This underlines the importance of selecting
298 adequately sized testing cohorts to obtain reliable and comparable results. In summary,
299 the specificities of 99.2%, 99.7%, and 98.3% found in the present study are very close to
300 the values given by the manufacturers of 99.6%, 99.8%, and 98.5% for Abbott, Roche,
301 and Diasorin, which were also established on large collectives.

302 *Sensitivity*

303 The COVID-19 positive cohort used in this study for the estimation of sensitivities is
304 relatively small (n=65). However, it has three distinctive features:

- 305 1. each patient/donor is represented in the collective with only one serum sample,
306 avoiding bias of the data by multiple measurements of the same individuals,
- 307 2. the median time of blood sampling was 41 days after onset of symptoms and thus
308 in the plateau phase of antibody formation, and
- 309 3. 80% of the cohort were non-hospitalized COVID-19 patients (two-thirds of them
310 with mild symptoms), and only 20% were intensive care patients.

311 As sensitivity within the first 14 days after symptom onset is highly variable for most
312 SARS-CoV-2 antibody assays but becomes better >14 days (23,24), we expected high
313 sensitivities for all tested assays in the plateau phase of antibody formation. Surprisingly
314 we found multiple RT-PCR confirmed COVID-19 patients displaying very low antibody

315 titers that did not surpass the respective assay-specific cut-offs and therefore were
316 considered negative. Five samples were negative in all three assays: all were RT-PCR
317 confirmed cases, 4/5 non hospitalized (42-51 days after symptom onset), two with mild
318 and the other two with moderate symptoms, and symptom duration of <1 week for all.
319 None of these patients had a known immune dysfunction or other severe diseases. One
320 patient was an ICU patient with an underlying hematological disease, and the sample
321 was taken at day 15 after symptom onset. This patient mounted a partial antibody
322 response starting from day 21 after symptom onset becoming positive in the Abbott
323 assay (2.21 Index) and reaching positivity on day 30 also in the DiaSorin assay (29.1
324 AU/ml). At this late point, Abbott measured 4.19 Index, whereas Roche remained (since
325 day 21) at a level around 0.2 COI. This example illustrates that although the vast
326 majority of patients show compatible and plausible results in different test systems,
327 single cases can display a complex picture of time-courses and reactivities in specific
328 assay systems that are still poorly understood. Another interesting observation was that
329 in six patients with positive detection of SARS-CoV-2 specific antibodies in the Roche
330 and Abbott test, DiaSorin failed to detect antibodies. This observation, combined with
331 the claim that the detection of S1/S2 protein-specific antibodies is equivalent to the
332 detection of neutralizing antibodies (nAbs) (20), raises the fundamental question of
333 whether nAbs are detectable in all patients with confirmed COVID-19 infection. There is
334 evidence that neither the assumption of equivalence of nAbs and S1/S2 protein-specific
335 antibodies (25) nor the assumption that all COVID-19 patients produce measurable titers
336 of nAbs is universally valid (26). A clear answer to the question of whether antibody
337 measurements against nucleocapsid- or spike protein-associated antigens are more
338 sensitive and specific, and how these behave in relation to nAbs assays (the postulated

339 gold standard in terms of sensitivity and specificity) is not possible based on the data
340 currently available.

341 *PPV, NPV, ROC-Analysis, test agreement*

342 Specificity and sensitivity alone are not sufficient to judge the performance of a
343 diagnostic test; prevalence-dependent accuracy measures like PPV and NPV are
344 necessary, and especially PPV, in times of low prevalence (27). For most regions
345 affected by the pandemic, the prevalence of SARS-CoV-2 antibody-positive individuals
346 is unknown but can be estimated to be below 5%. Therefore, for all SARS-CoV-2 EUA
347 approved antibody tests, the FDA compares the performance of the assays based on a
348 5% seroprevalence (28). At this rate, the results presented here show PPV values of
349 85.1% (74.7-91.7), 94.8% (85-98), and 71.6% (61.7-79.8) for Abbott, Roche, and
350 DiaSorin, respectively. The PPV values between Roche and DiaSorin differ so clearly
351 that not even the 95% CI intervals overlap. Therefore, we must assume that these two
352 assays differ significantly from each other in terms of positive predictability. Using these
353 two tests at lower seroprevalences, such as 1%, leads to an even more pronounced
354 difference between Roche and DiaSorin (77.6% vs. 32.6%) and an unacceptable low
355 PPV of 32.6% (23.6-43.1) for DiaSorin.

356 Although the area under the curves (0.994, 0.989, and 0.977 for Abbott, Roche, and
357 DiaSorin) did not differ significantly from each other (sensitivity cohort size was too
358 small), modeling of the cut-offs according to Youden's index revealed interesting
359 insights: only Roche could increase the sensitivity without losing specificity dramatically
360 (Sensitivity: 89.2% → 98.5%; Specificity: 99.7% → 99.4%; cut-off: >0.355 COI). In
361 contrast, DiaSorin at the suggested cut-off of >8.76 AU/ml (similar to (20)) increased the
362 sensitivity from 83.1% to 90.8% but worsened the specificity from 98.3 to 97.2%. In line

363 with this, despite a good overall agreement between Roche and DiaSorin results
364 (Cohen's Kappa 0.76 [0.67-0.84]), the McNemar's test still showed significant
365 differences, indicating disagreement (in particular in false-positives) more often than
366 expected by chance.

367 The strength of this study is the side by side evaluation of three assays with a large
368 number of negative samples to give reliable and comparable specificity data (no missing
369 data). Limitations are the moderate numbers of positive samples. Moreover, obtained
370 sensitivities cannot easily be compared to other studies because of the unique feature of
371 our COVID-19 cohort, including 80% non-hospitalized patients with mainly mild
372 symptoms. The latter is highly relevant for a potential use of antibody tests to assess
373 seroprevalence in large populations.

374 **Conclusion**

375 We find diagnostically relevant differences in specificities for the anti-SARS-CoV-2
376 antibody assays by Abbott, Roche, and DiaSorin that have a significant impact on the
377 positive predictability of these tests. We conclude that low seroprevalences require an
378 unusually high specificity for SARS-CoV-2 antibody tests, which pushes some test
379 systems to their limits earlier than others. Therefore, the choice of the test must depend
380 on the respective seroprevalence, and strategies such as confirmation of possible false-
381 positive test results with additional testing must be considered.

382 **Acknowledgements**

383 We sincerely thank Marika Gerdov, Susanne Keim, Karin Mildner, Elisabeth Ponweiser,
384 Manuela Repl, Ilse Steiner, Christine Thun, Martina Trella, for excellent technical
385 assistance. Finally, we want to thank all the donors of the various study cohorts. Without
386 their voluntary participation in the establishment of the biobanks, this study would not
387 have been possible.

388

389

390 **References**

391

- 392 1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from
393 Patients with Pneumonia in China, 2019. *N Engl J Med. Massachusetts Medical*
394 *Society*; 2020;382:727–33.
- 395 2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19
396 in real time. *The Lancet Infectious Diseases. Elsevier*; 2020;20:533–4.
- 397 3. Chen Q, Zheng Z, Zhang C, Zhang X, Wu H, Wang J, et al. Clinical characteristics
398 of 145 patients with corona virus disease 2019 (COVID-19) in Taizhou, Zhejiang,
399 China. *Infection. Springer Berlin Heidelberg*; 2020;92:401–9.
- 400 4. Hase R, Kurita T, Muranaka E, Sasazawa H, Mito H, Yano Y. A case of imported
401 COVID-19 diagnosed by PCR-positive lower respiratory specimen but with PCR-
402 negative throat swabs. *Infect Dis (Lond). Taylor & Francis*; 2020;52:423–6.
- 403 5. Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y, et al. Positive rate of RT-PCR detection of
404 SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan
405 to Feb 2020. *Clin Chim Acta. 2020*;505:172–5.
- 406 6. Yong G, Yi Y, Tuantuan L, Xiaowu W, Xiuyong L, Ang L, et al. Evaluation of the
407 auxiliary diagnostic value of antibody assays for the detection of novel coronavirus
408 (SARS-CoV-2). *Journal of Medical Virology. 2020*;:jmv.25919.

- 409 7. Farnsworth CW, Anderson NW. SARS-CoV-2 Serology: Much Hype, Little Data.
410 Clin Chem. 2020.
- 411 8. Theel ES, Slev P, Wheeler S, Couturier MR, Wong SJ, Kadkhoda K. The Role of
412 Antibody Testing for SARS-CoV-2: Is There One? J Clin Microbiol. 2020.
- 413 9. Infantino M, Grossi V, Lari B, Bambi R, Perri A, Manneschi M, et al. Diagnostic
414 accuracy of an automated chemiluminescent immunoassay for anti-SARS-CoV-2
415 IgM and IgG antibodies: an Italian experience. Journal of Medical Virology.
416 2020;;jmv.25932.
- 417 10. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Test
418 performance evaluation of SARS-CoV-2 serological assays. medRxiv. Cold Spring
419 Harbor Laboratory Press; 2020;;2020.04.25.20074856.
- 420 11. Abbott Diagnostics. SARS-CoV-2 IgG For Use With ARCHITECT, Revised April
421 2020. 2020.
- 422 12. Roche Diagnostics. Elecsys Anti-SARS-CoV-2 package insert 2020-04, V1.0.
423 2020.
- 424 13. DiaSorin S p A. LIAISON®SARS-CoV-2 S1/S2 IgG package insert 2020-04.
425 2020.
- 426 14. Breyer-Kohansal R, Hartl S, Burghuber OC, Urban M, Schrott A, Agusti A, et al.
427 The LEAD (Lung, Heart, Social, Body) Study: Objectives, Methodology, and
428 External Validity of the Population-Based Cohort Study. J Epidemiol. Japan
429 Epidemiological Association; 2019;29:315–24.

- 430 15. Haslacher H, Gerner M, Hofer P, Jurkowitsch A, Hainfellner J, Kain R, et al. Usage
431 Data and Scientific Impact of the Prospectively Established Fluid Bioresources at
432 the Hospital-Based MedUni Wien Biobank. *Biopres Biobanking*; 2018;16:477–82.
- 433 16. Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an
434 emphasis to case–control studies. *Statistics in Medicine*. John Wiley & Sons, Ltd;
435 2007;26:2170–83.
- 436 17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or
437 More Correlated Receiver Operating Characteristic Curves: A Nonparametric
438 Approach. *Biometrics*. 1988;44:837.
- 439 18. Rockx B, Kuiken T, Herfst S, Bestebroer T, Lamers MM, Oude Munnink BB, et al.
440 Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman
441 primate model. *Science*. American Association for the Advancement of Science;
442 2020;:eabb7314.
- 443 19. Weiss SR. Forty years with coronaviruses. *J Exp Med*. 2020;217.
- 444 20. Bonelli F, Sarasini A, Zierold C, Calleri M, Bonetti A, Vismara C, et al. Clinical And
445 Analytical Performance Of An Automated Serological Test That Identifies S1/S2
446 Neutralizing IgG In Covid-19 Patients Semiquantitatively. *bioRxiv*. Cold Spring
447 Harbor Laboratory; 2020;25:2000082–35.
- 448 21. Tré-Hardy M, Wilmet A, Beukinga I, Dogné J-M, Douxfils J, Blairon L. Validation of
449 a chemiluminescent assay for specific SARS-CoV-2 antibody. *Clinical Chemical
450 Laboratory Medicine*. 2020.

- 451 22. Ekelund O, Ekblom K, Somajo S, Pattison-Granberg J, Olsson K, Petersson A.
452 High-throughput immunoassays for SARS-CoV-2, considerable differences in
453 performance when comparing three methods. medRxiv. Cold Spring Harbor
454 Laboratory Press; 2020;:2020.05.22.20106294.
- 455 23. Tang MS, Hock KG, Logsdon NM, Hayes JE, Gronowski AM, Anderson NW, et al.
456 Clinical Performance of Two SARS-CoV-2 Serologic Assays. Clin Chem. 2020.
- 457 24. Bryan A, Pepper G, Wener MH, Fink SL, Morishima C, Chaudhary A, et al.
458 Performance Characteristics of the Abbott Architect SARS-CoV-2 IgG Assay and
459 Seroprevalence in Boise, Idaho. J Clin Microbiol. 2020;:1–19.
- 460 25. Jääskeläinen AJ, Kuivanen S, Kekäläinen E, Ahava MJ, Loginov R, Kallio-Kokko
461 H, et al. Performance of six SARS-CoV-2 immunoassays in comparison with
462 microneutralisation. medRxiv. Cold Spring Harbor Laboratory Press;
463 2020;:2020.05.18.20101618.
- 464 26. Wu F, Wang A, Liu M, Wang Q, Chen J, Xia S, et al. Neutralizing Antibody
465 Responses to SARS-CoV-2 in a COVID-19 Recovered Patient Cohort and Their
466 Implications. SSRN Journal. 2020.
- 467 27. Šimundić A-M. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC.
468 International Federation of Clinical Chemistry and Laboratory Medicine;
469 2009;19:203–11.
- 470 28. US Food Drug Administration. EUA Authorized Serology Test Performance
471 [Internet]. fda.gov. 2020 [cited 2020 Jun 2]. Available from:

472 [https://www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-](https://www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-authorized-serology-test-performance)
473 [authorized-serology-test-performance](https://www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-authorized-serology-test-performance)

474

475 **Tables**

476 **Table 1.** Numbers and percentages of false positive SARS-CoV-2 antibody reactivities
477 in three different specificity cohorts: Cohort A (LEAD-Study), Cohort B (Healthy donor
478 collective), and Cohort C (Rheumatic diseases cohort). χ^2 -tests for differences of
479 proportions: *... Roche vs. DiaSorin $p < 0.05$, &... Abbott vs. Diasorin $p < 0.05$

	COHORT A	COHORT B	COHORT C	TOTAL
	<i>n=494</i>	<i>n=302</i>	<i>n=358</i>	<i>n=1,154</i>
Abbott SARS-CoV-2 IgG	4 (0.8%)	3 (1.0%)	2 (0.6%) ^{&}	9 (0.8%)
Roche Elecsys® Anti-SARS-CoV-2	0 (0.0%)*	1 (0.3%)	2 (0.6%)*	3 (0.3%)
DiaSorin LIAISON® SARS-CoV-2 S1/S2 IgG	6 (1.2%)*	5 (1.7%)	9 (2.5%)* ^{&}	20 (1.7%)

480

481

482 **Table 2.** Values for Specificity, Sensitivity, Positive-Predictive-Value (PPV) and
 483 Negative-Predictive-Value (NPV) at 1%, 5% and 10% SARS-CoV-2 seroprevalence (SP)
 484 with 95% confidence intervals (95% CI).

485

	Abbott SARS-CoV-2 IgG		Roche Elecsys® Anti-SARS-CoV-2		DiaSorin LIAISON® SARS-CoV-2 S1/S2 IgG	
Statistic	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	84.6%	73.6-92.4	89.2%	79.1-95.6	83.1%	71.3-91.2
Specificity	99.2%	98.6-99.7	99.7%	99.2-100	98.3%	97.3-98.9
1% Seroprevalence						
<i>PPV</i>	52.3%	36.2-67.9	77.6%	52.8-91.5	32.6%	23.6-43.1
<i>NPV</i>	99.9%	99.7-99.9	99.9%	99.8-100	99.8%	99.7-99.9
5% Seroprevalence						
<i>PPV</i>	85.1%	74.7-91.7	94.8%	85.3-98.3	71.6%	61.7-79.8
<i>NPV</i>	99.2%	98.6-99.5	99.4%	98.9-99.7	99.1%	98.5-99.5
10 % Seroprevalence						
<i>PPV</i>	92.3%	86.2-95.9	97.4%	92.5-99.2	84.2%	77.3-89.3
<i>NPV</i>	98.3%	97.0-99.0	98.8%	97.6-99.4	98.1%	96.8-98.9

486

487

488 **Table 3.** Inter-rater agreement (Cohen's kappa) with linear weights. Value of K <0.20
489 poor agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80
490 good agreement, and 0.81-1.00 very good agreement.

491

	Abbott			Kappa (95% CI)
Roche	NEG	POS		0.87 (0.81-0.94)
NEG	1149	9	1158 (95.0%)	Standard Error
POS	6	55	61 (5.0%)	0.032
	1155 (94.7%)	64 (5.3%)	1219	

492

	Abbott			Kappa (95% CI)
DiaSorin	NEG	POS		0.71 (0.62-0.79)
NEG	1131	14	1145 (93.9%)	Standard Error
POS	24	50	74 (6.1%)	0.045
	1155 (94.7%)	64 (5.3%)	1219	

493

	DiaSorin			Kappa (95% CI)
Roche	NEG	POS		0.76 (0.67-0.84)
NEG	1136	22	1158 (95.0%)	Standard Error
POS	9	52	61 (5.0%)	0.042
	1145 (93.9%)	74 (6.1%)		

494

495

496 **Figure Captions**

497
498 **Figure 1:** Specificity was determined using 1,154 serum samples taken before the
499 circulation of SARS-CoV-2. For SARS-CoV-2 antibody tests Abbott SARS-CoV-2 IgG
500 **(A)**, Roche Elecsys® Anti-SARS-CoV-2 **(B)**, and DiaSorin LIAISON® SARS-CoV-2
501 S1/S2 IgG **(C)**, values of specificity samples are shown in rank order. Horizontal dotted
502 lines mark the respective cut-offs recommended by the manufacturer and, in the case of
503 DiaSorin, a gray zone for equivocal results. Vertical dotted lines indicate the median (*)
504 and the 90th percentile values (**). Median and 90th percentile values of negative
505 samples were 0.025 and 0.115 for Abbott, 0.0815 and 0.0927 for Roche, below LOD
506 and 5.52 for DiaSorin.

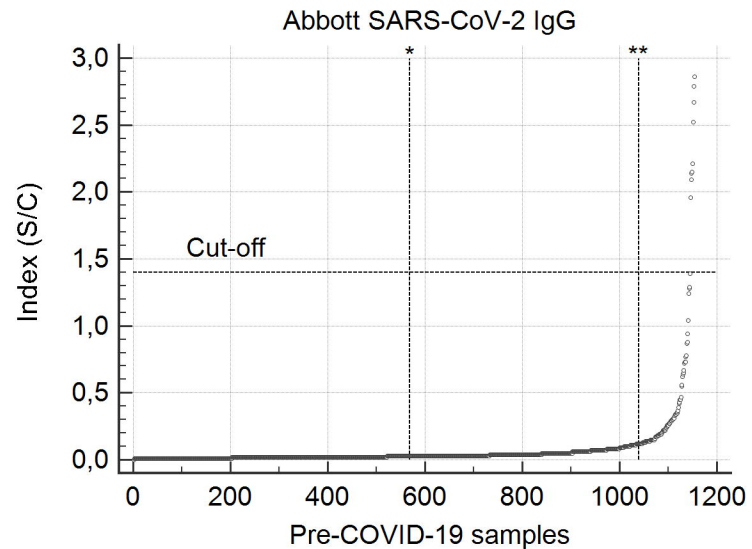
507
508 **Figure 2:** Quality criteria of the SARS-CoV-2 antibody tests Abbott SARS-CoV-2 IgG,
509 Roche Elecsys® Anti-SARS-CoV-2, and DiaSorin LIAISON® SARS-CoV-2 S1/S2 IgG.
510 Columns represent values of sensitivity, specificity **(A)**, PPV, and NPV **(B)** at 1%, 5%,
511 and 10% assumed seroprevalence; bars indicate the 95% CI.

512
513 **Figure 3:** Receiver Operating Characteristic (ROC) curves for Abbott **(A)**, Roche **(B)**,
514 and DiaSorin **(C)** are shown. The Area Under the Curve (AUC) indicates the test
515 accuracy, 95% confidence intervals are represented by gray dotted lines.

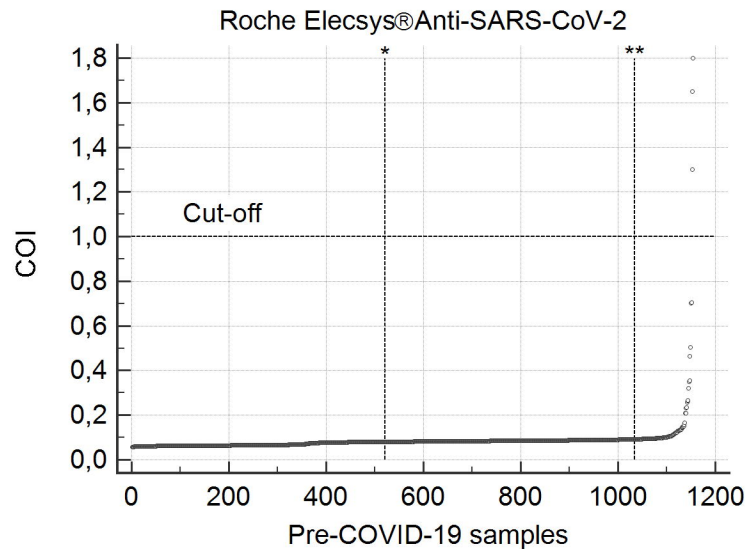
516

517

A Specificity = 99.2%



B Specificity = 99.7%



C Specificity = 98.3%

