

1 **Title: Performance of SARS-CoV-2 Serology tests: Are they good enough?**

2 Running title: Analytical performance of five SARS-CoV-2 antibody tests

3

4 **Authors** Isabelle Picc¹, Emma English², M Annette Thomas³, Samir Dervisevic⁴, William D Fraser^{1,5} and
5 W Garry John^{2,5}

6 **Affiliations** ¹BioAnalytical Facility, Faculty of Medicine, University of East Anglia, Norwich, UK.

7 ²Faculty of Medicine and Health, University of East Anglia, Norwich, UK. ³WEQAS, Cardiff and Vale

8 University Health Board, UK, ⁴Norfolk and Norwich University Hospitals, Department of Microbiology,

9 Norwich, UK. ⁵Norfolk and Norwich University Hospitals, Department of Clinical Biochemistry,

10 Norwich, UK.

11

12 **Corresponding author** Isabelle Picc, Bob Champion Research and Education Building, Rosalind
13 Franklin Road, Norwich NR4 7UQ, UK. Email: i.picc@uea.ac.uk

14

15 **Criteria for authorship** IP has provided substantial contributions to conception and design, data
16 acquisition, analysis and interpretation, to drafting the article and revising it. EE, AT, GJ, WDF and SD
17 provided substantial contributions to conception and design, analysis and interpretation of data and
18 revision of the article.

19

20 **Keywords**

21 SARS-CoV-2, serology, assay validation, specificity, sensitivity, cross-reactivity.

22

23 **List of Abbreviations**

24	LFIA	Lateral flow immunoassays
25	PHE	Public Health England
26	IgG	Immunoglobulin G
27	NNUH	Norwich and Norfolk University Hospital
28	QEH	Queen Elizabeth Hospital in King Lynn
29	EBV	Epstein Barr Virus
30	NOAR	Norfolk Arthritis Register
31	anti-CCP	Cyclic citrullinated peptide antibodies
32	N	Negative control
33	CR	Cross-reactivity
34	RA	Rheumatoid Arthritis
35	TSI	Thyroid stimulating immunoglobulin
36	P	SARS-CoV-2 Positive
37	EDI	Epitope Diagnostics Ltd
38	POCT	Point of Care Testing
39	OD	Optical density/absorbance
40	R	Threshold of positivity
41	CLSI	Clinical and Laboratory Standards Institute
42	CV	Coefficient of variation expressed as percentage

43 OD Optical density

44 RLU Relative Light Unit

45 S1/S2 Spike protein 1 and 2

46 WEQAS Wales External Quality Assessment Scheme (UK)

47 EQA External quality assessment

48

49 **Competing interest declaration**

50 The authors declare no competing financial, professional, or personal interests that might have influenced
51 the performance or presentation of the work described in this manuscript.

52

53

54 **Abstract:**

55 **Background:** In the emergency of the SARS-CoV-2 pandemic, great efforts were made to quickly
56 provide serology testing to the medical community however, these methods have been introduced into
57 clinical practice without the complete validation usually required by the regulatory organizations.

58 **Methods:** SARS-CoV-2 patient samples (n=43) were analysed alongside pre-pandemic control specimen
59 (n=50), confirmed respiratory infections (n=50), inflammatory polyarthritis (n=22) and positive for
60 thyroid stimulating immunoglobulin (n=30). Imprecision, diagnostic sensitivity and specificity and
61 concordance were evaluated on IgG serologic assays from EuroImmune, Epitepe Diagnostics (EDI),
62 Abbott Diagnostics and DiaSorin and a rapid IgG/IgM test from Healgen.

63 **Results:** EDI and EuroImmune imprecision was 0.02-14.0% CV. Abbott and DiaSorin imprecision (CV)
64 ranged from 5.2% - 8.1% and 8.2% - 9.6% respectively. Diagnostic sensitivity of the assays were 100%
65 (CI: 80-100%) for Abbott, EDI and EuroImmune and 95% (CI: 73-100%) for DiaSorin at ≥ 14 days post
66 PCR. Only the Abbott assay had a diagnostic specificity of 100% (CI: 91-100%). EuroImmune cross-
67 reacted in 3 non-SARS-CoV-2 respiratory infections and 2 controls. The DiaSorin displayed more false
68 negative results and cross-reacted in six cases across all conditions tested. EDI had one cross-reactive
69 sample. The Healgen rapid test showed excellent sensitivity and specificity. Overall, concordance of the
70 assays ranged from 76.1% to 97.9%.

71 **Conclusions:** Serological tests for SARS-CoV-2 showed good analytical performance. The head-to-head
72 analysis of samples revealed differences in results that may be linked to the use of nucleocapsid or spike
73 proteins. The point of care device tested demonstrated adequate performance for antibody detection.

74

75

76 **Introduction**

77 The scientific community has had to rapidly develop and manufacture tests for the new SARS-CoV-2
78 pandemic at unprecedented speed, taking three months to develop assays that would ordinarily take three
79 years. Serology testing, that can identify those who have previously been exposed to the SARS-CoV-2
80 virus and have mounted an immune response, has been hailed as key to managing the pandemic however
81 controversy remains over both the accuracy and utility of serology testing in disease management.
82 Structural proteins, including the spike (essential for viral infection) and the nucleocapsid (important for
83 viral RNA transcription), are both potential targets for early detection of infection and known to elicit an
84 immune response in the host (1) with antibodies detectable within 20 days of disease onset (2–4).
85 Systematic reviews (5,6) challenged the diagnostic accuracy of serological tests, particularly when using
86 lateral flow immunoassays (LFIAs). Public Health England (PHE) showed only the Siemens and the
87 Roche Diagnostics assays met the minimum UK Medicines and Healthcare products Regulatory Agency
88 Target Product Profile criteria for sensitivity (7) after the threshold of positivity was adjusted to 0.128.
89 Assays from DiaSorin and Abbott Diagnostics (8) also provided acceptable diagnostic results. These
90 evaluations did not address cross-reactivity. To our knowledge little has been done regarding interference
91 from antibodies produced during other viral infection and autoimmune disorders. Additionally, with the
92 focus on diagnostic sensitivity and specificity, little has been done to evaluate the analytical accuracy,
93 which if poor, has the potential to negate all of these study findings. Indeed in the editorial, Duong and
94 colleagues clearly states that there is a need for critical independent evaluations of these tests, using the
95 same specimen panels(9).

96 This study provides a head-to-head evaluation of the diagnostic and analytical performance of four
97 commercially available IgG based serology assays for SARS-CoV-2 and a diagnostic accuracy study of
98 one point of care LFIA.

99

100 **Material and methods**

101 **Specimen collection and storage**

102 Patients were not involved in any part of the work. All samples were from archived specimens and were
103 fully anonymized before we accessed them. Therefore, our study is in accordance with the blanket Ethical
104 standards of University of East Anglia on de-identified samples for method development. Moreover,
105 using the UK NHS Research Ethics Committee decision toolkit ([http://www.hra-](http://www.hra-decisiontools.org.uk/ethics/)
106 [decisiontools.org.uk/ethics/](http://www.hra-decisiontools.org.uk/ethics/)) we confirmed that separate ethical review is not required for this study which
107 is in concordance with the Helsinki Declaration.

108 All serum samples were collected, anonymized, aliquoted and stored at -80°C until analysed. SARS-
109 CoV-2 PCR-positive patients (AusDiagnostics platform, Chesham, UK) were of both genders, age range
110 66 to 93 and hospitalized at the Norfolk and Norwich University Hospital (NNUH) or Queen Elizabeth
111 Hospital in King Lynn (QEH). Samples were taken 8-44 days after testing positive for SARS-CoV-2.
112 Negative control samples were collected in 2018 from patients with no history of infection or immune
113 disorder. Pre-pandemic samples from patients who had a range of confirmed respiratory infections
114 (including Influenza A, B and seasonal coronaviruses [Table 1]), samples collected from patients with
115 inflammatory polyarthritis positive for anti-cyclic citrullinated peptide antibodies (anti-CCP) along with
116 samples positive for thyroid stimulating immunoglobulin (TSI) were used to test the non-specific binding
117 of non-SARS-CoV-2 antibodies. These groups of samples are referred to as N (negative control), CR
118 (cross-reactivity), RA (Rheumatoid Arthritis), TSI (patients with thyroid stimulating immunoglobulin)
119 and P (SARS-CoV-2 Positive). A total of 195 individual serum samples (43 P, 50 N, 50 CR, 22 RA and
120 30 TSI) were analysed for SARS-CoV-2 IgG antibodies. For a subset of patients, samples were available
121 for a series of time-points thus allowing for a time course analysis (43 patients, 142 samples).

122

123 **Study design**

124 SARS-CoV-2 IgG immunoassays were from 1) Epitope Diagnostics Inc. (EDI, San Diego, CA, USA)
125 performed using the Agility ELISA automate (Dynex Technologies, Chantilly, VA, USA), 2) EuroImmun
126 UK ITC (UK) performed manually, 3) Abbott Diagnostics (Maidenhead, UK) on the Alinity™ i analyser
127 and 4) DiaSorin (London, UK) on the Liaison XL analyser. A subset of samples were also tested using
128 the point of care testing (POCT) device SARS-CoV-2 IgG/IgM rapid test from Healgen (Houston, TX,
129 USA). Due to a limited number of cassettes available, 49 samples from 27 P were analysed along with 3
130 N, 8 CR, 4 RA and 4 TSI. Cross-reactive and negative samples were primarily chosen from patient
131 samples proven positive for seasonal coronaviruses and influenza A or a false positive result in one or
132 more of the immunoassays. We focused on the IgG results in order to compare with the immunoassays.

133 Assays were performed by trained biomedical scientists using manufacturer's instructions. The SARS-
134 CoV-2 Abbott assay was performed in the clinical biochemistry department at NNUH and the other
135 SARS-CoV-2 assays were performed at the University of East Anglia. All other non- SARS-CoV-2
136 related tests were performed at the NNUH virology department. DiaSorin SARS-Cov-2 is a quantitative
137 assay and antibody concentrations are expressed in AU/mL. The Abbott, EDI and EuroImmun are
138 qualitative assays for which the result is calculated using the ratio of the sample optical density (OD)
139 against the negative or calibrator control (Supplemental Table). EuroImmun and DiaSorin assays detect
140 antibodies to, respectively, recombinant S1 and S1/S2 domains of the SARS-CoV-2 spike protein while
141 both the EDI and Abbott assay detect antibodies to the nucleocapsid. The POCT from Healgen is a solid
142 phase lateral flow immunochromatographic assay (LFIA) for detection of SARS-CoV-2 of IgG and IgM,
143 antigen not specified.

144

145 **Imprecision**

146 As the results are expressed with a values correlating with the amount of antibody detectable, imprecision
147 was assessed using a Clinical and Laboratory Standards Institute (CLSI) EP-15 based protocol on the

148 automated clinical laboratory analysers protocol (34). Positive and negative patient pools and/or controls
149 of different concentrations were prepared and frozen as aliquots and assayed as 5 replicates per day on 5
150 different days. For the plate based assays, inter- and intra-assay CVs were calculated. Intra-assay was
151 determined using the CV of the optical density (OD) of duplicated samples. Inter-assay was determined
152 using the CV obtained from the sample pool and the kit positive control across the plates.

153

154 **Statistics**

155 Using IBM SPSS Statistics 25.0.0.1, Mann-Whitney and Cohen's Kappa tests were used to compare OD
156 results between groups and to determine the concordance between the assays, respectively. Analysis of
157 EP15 was performed using EP evaluator. Variation was estimated on calculated values (R) or response
158 (Relative Light Unit, RLU) as intra and inter-assay coefficient of variation (CV). Graphical
159 representations were conducted with GraphPad Prism version 8.0 (GraphPad Software, Inc., USA).
160 Throughout the tables, figures, and legends, the following terminology is used to show statistical
161 significance: *P<0.05; **P<0.01 and ***P<0.001.

162

163 **Results**

164 **Imprecision**

165 *Abbott* EP15 and was performed on two Alinity analysers (Table 2). Overall, negative pool imprecision
166 was CV=8.1% and 6.8% on equipment 1 and 2 respectively. Positive pool imprecisions were CV=2.3%
167 and 1.1% respectively.

168 *DiaSorin* EP15 imprecision was estimated based on response intensity (RLU). Positive control
169 imprecision was between 8.2% and 13.8% (Table 2). The negative quality control material results were

170 consistently below the lower limit of detection of 3.8AU/mL and the negative pool concentration was
171 consistently below 10AU/mL, the resulting calculated imprecision was therefore expectedly elevated.

172 **EDI and EuroImmun** Intra-assay imprecision on duplicate samples (Table 2) was on average CV=
173 3.3±3.8% and 6.1±6.7% respectively. Inter-assay imprecision of EDI was CV=14.2% for the kit positive
174 pool and 16.5% for the negative pool. Baseline OD varied between the plates increasing the inter-assay
175 variations, however, the ratio positive/cut-off was on average 1.43 ±0.16 (CV =11.1%). Inter-assay of
176 EuroImmun was evaluated using the positive kit QC, the calibrator and the negative kit control.
177 Coefficient of variation were CV = 12.9%, 9.5% and 3.7% respectively.

178

179 **Specificity and sensitivity**

180 A total of 43 individual P were analysed for SARS-CoV-2 IgG antibodies. Of these, twenty had samples
181 taken at least 14 days after a positive PCR result ($P \geq 14$) and 23 were taken prior ($P < 14$). All $P \geq 14$ had
182 detectable antibodies in the EDI, EuroImmun, Abbott and Healgen assays. However, one sample returned
183 a negative result using the DiaSorin assay. These results suggest a true positive rate of 100% with EDI,
184 EuroImmun, Abbott and Healgen assays and 95% for the DiaSorin assay.

185 Amongst the 23 $P < 14$ samples, antibodies were detected for 65% (Abbott & EuroImmun), 61% (EDI) and
186 43% (DiaSorin) of the samples. Two samples R were close to the threshold in EDI and Abbott (EDI: 0.8
187 and Abbott 1.9; EDI: 1.0 and Abbott 0.8) resulting in one being positive in one assay and negative in the
188 other (and vice-versa).

189 All 50 N were negative on the Abbott and EDI. Two samples were positive and 48 were negative on the
190 EuroImmun (although 2 were equivocal). Two false positive samples were also observed on the DiaSorin,
191 one being positive on both DiaSorin and EuroImmun assays.

192 The IgG kits showed a very good diagnostic ability to differentiate between P and N (Table 3). Overall,
193 EuroImmun and DiaSorin showed lower sensitivity and specificity than EDI and Abbott. Sensitivity
194 ranged between 81-100% on all time points for EDI, EuroImmun and Abbott. DiaSorin sensitivity was
195 71% on all time points and 95% for $P \geq 14$. Specificity was consistently 100% for the Abbott while it
196 ranged between 92 to 100% for the other assays.

197

198 **Cross-Reactivity**

199 There were no SARS-CoV-2 IgG positive results from patients with non-SARS-CoV-2 infection (CR,
200 $n=50$, including seasonal flu ($n=7$)), anti-CCP positive (RA, $n=22$) nor TSI positive ($n=30$) using the
201 Abbott and the EDI assays. Overall, DiaSorin showed the highest (4%) cross-reactivity (2CR, 1 RA and 1
202 TSI), followed by EuroImmun (3% - 3CR) and EDI (1% - 1 TSI). The Mann-Whitney test showed that on
203 the EDI only, the R value of samples used to test cross-reactivity (RA and TSI) was significantly
204 elevated, however only one sample was falsely positive for SARS-CoV-2 (Figure 1).

205 Any sample that gave a false positive result in any of the immunoassays was also tested on the Healgen
206 POCT and none were IgG positive. However, a very weak signal could be detected on one TSI sample
207 and one sample from a patient with seasonal flu. Because of the very small number of sample tested,
208 specificity calculation was not performed for the rapid test.

209

210 **Time course analysis**

211 We analysed 1 to 13 data points for 43 P. We observed an increase of the signal for presence of IgG over
212 time going from negativity to positivity and reaching a plateau (Figure 2). Sigmoid curve-fitting indicated
213 a time from PCR to seroconversion at 9.8 days (95% CI 10.7-13.7), 10.2 (95% CI 8.5-11.8), 12.2 days
214 (95% CI 10.7-13.7) and 10.4 days (95% CI 7.9-12.9) for EDI, Abbott, DiaSorin and EuroImmun assays

215 respectively. Note that due to a limited number of EuroImmun tests available, we only had measurements
216 for 56 (of 142) data points. One data point was missing for Abbott and 4 were missing for DiaSorin due to
217 insufficient sample volume.

218 We tested 48 samples from 27 P patients using the Healgen rapid test. Ninety four percent (n=45)
219 displayed a positive test for IgG. Samples showed positive results with POCT from day 7 post PCR
220 although these were still negative in the other immunoassays (SARS-CoV-2 positive at day 12).

221

222 **Assay Concordance**

223 Abbott and EDI had the greatest concordance with Cohen's Kappa of 0.957 and 97.9% agreement
224 between the all results (Table 4). DiaSorin was the most different, with agreements below 95%. The
225 Healgen POCT concordance with the other assays was low (below 90%) but reflect a limited number of
226 samples and may not be representative. Modifying the threshold to 0.8 for EDI would allow the detection
227 of 2 more P<14 without increasing the rate of false positive. No change in threshold in the other assay
228 would reclassify any results without dramatically affecting the specificity to either have a high rate of
229 false positive or false negative.

230

231 **Discussion**

232 **Statement of principal findings**

233 In this head to head study we demonstrated the good performance of four commercially available
234 serologic assays for SARS-CoV-2 and one POCT. Abbott, Epitepe Diagnostics Ltd and EuroImmun
235 demonstrated higher sensitivity and specificity than the DiaSorin assay on the same specimens. The
236 Abbott assay showed no cross-reactivity to any other potential interfering substances tested while EDI,
237 EuroImmun and DiaSorin cross-reacted in 1%, 3% and 4% of the sample tested. However, no assay

238 cross-reacted with Influenza A and B or other coronaviruses. The analytical performance was deemed
239 acceptable although it varied considerably between the different methods.

240

241 It is estimated that there are nearly 300 different SARS-CoV-2 antibody tests in development globally
242 ranging from POCT through to assays on large clinical laboratory analysers. Whilst data is accruing on
243 the sensitivity and specificity of a number of these assays (5,6) there are still many with little or no
244 published, independent performance evaluations. Whilst there is a focus on the diagnostic accuracy of
245 these tests, much less is understood about the analytical performance of these devices such as imprecision
246 and cross reactivity with common respiratory illnesses or immunoassay interferences. Without this
247 knowledge the sensitivity and specificity data is brought into question and it is important that the
248 limitations of assay are fully understood before applying the results in clinical practice. The Food and
249 Drug Administration and European Medicines Agency acceptance criteria for biological assays typically
250 define the required between-run and within-run precision as $CV \leq 15\%$ for positive samples and $\leq 20\%$ for
251 samples at the lower limit of quantification (10,11). All immunoassays passed the criteria for positive
252 samples.

253 Published median seroconversion time for IgG is around 14 days post symptoms (12–14). As we did not
254 have access to symptom onset for most patients, we used PCR day to date the samples, before and after
255 day 14. All samples post day 14 were positive in all assay except DiaSorin, which returned one false
256 negative (day 39). Positivity prior to day 14 was consistent between EDI, EuroImmun and Abbott. These
257 results are differing from those published by PHE who observed more false negative results in the Abbott
258 than the DiaSorin (92.7% sensitivity vs 95% sensitivity, respectively) (8). We estimated seroconversion
259 post PCR positivity to be between 9 and 12 days on these assays. Although we couldn't do a full
260 comparison of the POCT with the immunoassays, 100% of the $P \geq 14$ samples were IgG positive. More
261 samples were also positive with POCT prior day 14 than in the other assays.

262 In regard to the POCT, our study showed excellent sensitivity and specificity. We observed no false
263 negative results on P_≥14 after a positive SARS-CoV-2 PCR and more samples were IgG positive P<14
264 than the other immunoassays. Two potential false positive were detected (including seasonal flu) but the
265 signal was very weak and confirmation would be necessary. The results of systematic reviews on point-of
266 care serological tests for SARS-CoV-2 suggest discontinuing the use of the devices due to low sensitivity
267 (5). Our results tend to reveal a different pattern however we only performed a limited number of tests.
268 We chose 50 samples collected in 2018 from patients with no known infection as negative controls. Both
269 the EDI and the Abbott showed 100% specificity. However, EuroImmun and DiaSorin produced false
270 positives (n=4 and 2, respectively). Only one of these samples was common between both assays. PHE
271 also showed lower specificity of the DiaSorin assay (vs Abbott). We analysed 50 samples from patients
272 (pre-pandemic) presenting with respiratory infection. Among those 7 had the seasonal flu, 8 had influenza
273 A., other viruses included EBV, Varicellazoster virus, parainfluenza, Adenovirus. EDI and Abbott
274 showed 100% specificity with no false positive; however, we observed 3 positive results with the
275 EuroImmun, two of these also being positive with the DiaSorin. These samples were from patients with
276 EBV (n=1) and RSV (n=2). Our results on EuroImmun differ slightly from a previous evaluation (15),
277 where specificity of the assay was excellent as early as 4 days after positive PCR and only 2 of 28
278 samples showed borderline cross-reactivity to common human coronaviruses. None of the assays showed
279 cross-reactivity either to the seasonal CoV flu or to Influenza A. Although it is based on a small number
280 of sample (n=7 for each), it brings confidence that assays will be able to discriminate SARS-CoV-2
281 antibodies during the next seasonal flu. Tang *et al.*, showed similar results on 5 patients using EuroImmun
282 and Abbott Assay (16). A great variety of endogenous substances such as polyreactive antibodies or
283 autoantibodies, can interfere with the reaction between analyte and reagent antibodies in immunoassays.
284 Assays for SARS-CoV-2 are no exception. Manufacturers, and evaluation studies to date, offer a limited
285 insight into cross-reactivity of other antibodies in particular to other SARS-CoV antibodies(17–21). A
286 small independent study showed no cross-reactivity was seen for patients with Influenza A (n=2),

287 Influenza B (n=2) and other coronaviruses (n= 5) (16). Samples with potentially interfering antibodies did
288 not cross-react in the Abbott Diagnostics assay, and a limited number cross-reacted in the other assays.
289 None of these samples was common between the different assays and modification of the various
290 threshold would not improve performance of any assay.

291 Successful attempts to treat SARS-CoV-2 patients with blood from convalescent individuals suggest
292 antibodies against SARS-CoV-2 may have the ability to confer protective immunity to the disease (22–
293 27). Spike proteins are the most likely target for neutralizing antibodies are displayed on the surface of the
294 virus whereas the nucleocapsid is contained within the viral envelope(28,29). Antibodies against the
295 nucleocapsid have been shown to appear first (30,31), followed by the production of antibodies against
296 the spike protein (12,13). Therefore, assays based on the nucleocapsid detection appear to be more
297 sensitive early on in the disease recovery but presence of anti-S1/S2 antibodies may indicate presence of
298 neutralizing antibodies. Both the EuroImmun and the DiaSorin are targeted the spike protein of SARS-
299 Cov-2 while the EDI and Abbott are targeted to the nucleocapsid protein of the virus. We observe a
300 highest specificity of both nucleocapsid assays (EDI and Abbott, 100% (91-100%)) compared to the two
301 spike assays (DiaSorin (96% (85-99%)) and EuroImmun (92% (79-97%)). Although the EuroImmun
302 assay had the same sensitivity (all time points to PCR) as the EDI and Abbott, the DiaSorin assay was
303 less sensitive (71% (73-100%) vs (81% (66-91%)), potentially supporting this hypothesis.

304 Overall, the assays had high concordance, DiaSorin being the least identical to the others, with higher
305 false negative and false positive, and lower performance. This is in accordance with the high false
306 positive rate observed by Boukli et al. (32) with the DiaSorin Liaison SARS-CoV-2 IgG assay on patients
307 with non-SARS-CoV-2 acute infections. The same samples were analysed on the different platforms and
308 therefore the direct comparison is possible. However, one needs to consider the potential variance in
309 antigen as the Wuhan strain has evolved as geographic spread has occurred between the different regions
310 of the globe (GISAID) (33) and it is possible that these differences will not be seen on a different set of
311 samples. Harmonization of the assays is necessary but will be near impossible with such variation

312 between assay designs (spike vs nucleocapsid). The Wales External Quality Assessment Scheme
313 (WEQAS, UK, <https://www.wegas.com/>) is now offering a SARS-CoV-2 antibody external quality
314 assessment (EQA) program for laboratories which will reduce uncertainty associated with different
315 methods.

316

317 **Conclusion and policy implications**

318 The role of serology testing in the management of people with SARS-CoV-2 infection will remain
319 controversial until we have clear data that enables an understanding of how production of IgG relates to
320 immunity over time and whether or not the presence or absence of antibodies can inform risk of future
321 infection. Whilst the clinical utility of serology tested is debated, it is important that the diagnostic and
322 analytical performance of these tests is understood and adequate for need so that there can be confidence
323 in the results when a meaningful clinical use is determined. Without high quality analytical testing the
324 clinical application of serology testing in the future is not viable.

325 This study examines the performance of four commercially available serologic assays for SARS-CoV-2
326 in a head to head study. Our study demonstrated good analytical performance for all of the assays,
327 however we observed Abbott, EDI and EuroImmun demonstrated higher sensitivity and specificity than
328 the DiaSorin assay in this study. Whilst a full evaluation was not possible the P14+ samples from the
329 main study were used in a sub analysis using the Healgen POCT device which showed 100% specificity,
330 this contradicts earlier studies (5,6) and indicates that the evolution of the quality of POC devices has
331 been rapid and some may now demonstrate adequate performance for antibody detection.

332 Assays showed 0-4% cross-reactivity, however none with Influenza viruses. This may give increase
333 confidence of the test during the seasonal flu period. We observed differences between the assay
334 responses with DiaSorin being the most different from the other three. We hypothesize that these
335 differences may be linked to the design of the assay themselves (spike glycoprotein or nucleocapsid) and

336 the timeline of production of antibodies for either antigen. We also suggested the possibility that the
337 antigen plasticity and the antigen used when the manufacturer set up the test may influence the sensitivity
338 of the CoV-2 assays. These findings highlight the importance of following the evolution of the antibody
339 production and evolution of the virus over time. But it also highlights how harmonization of the assays
340 will be complex.

341

342

343

344 **Acknowledgements**

345 The authors would like to thank the staff at the Norfolk and Norwich University Hospital and the Queen
346 Elizabeth Hospital, Kings Lynn who collected the samples from SARS-CoV-2 patients, in particular
347 Christopher Jeanes, and the Norfolk Arthritis Register (NOAR) for kindly providing historical samples of
348 patients with inflammatory polyarthritis. We would also like to express our gratitude to Myra Del Rosario
349 and Simon Clements who analysed the samples on the Abbott Alinity, and Christopher McDonnell and
350 Reenesh Prakash who procured EuroImmun kits. We finally wish to thank the Norfolk, Suffolk, Essex
351 and Hertfordshire Freemasons for their generous financial support to purchase the Agility system.

352

353

354 **References**

- 355 1. Chang MS, Lu YT, Ho ST, Wu CC, Wei TY, Chen CJ, et al. Antibody detection of SARS-CoV
356 spike and nucleocapsid protein. *Biochem Biophys Res Commun.* 2004;314(4):931–6.
- 357 2. Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mueller MA, et al. Clinical
358 presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a
359 travel-associated transmission cluster. *medRxiv.* 2020;2020.03.05.20030502.
- 360 3. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-
361 CoV-2 in patients with COVID-19. *Nat Med.* 2020;26(6):845–8.
- 362 4. Okba NMA, Müller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, et al. Severe Acute
363 Respiratory Syndrome Coronavirus 2-Specific Antibody Responses in Coronavirus Disease
364 Patients. *Emerg Infect Dis.* 2020;26(7):1478–88.
- 365 5. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests
366 for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev.*
367 2020;2.
- 368 6. Lisboa Bastos M, Tavaziva G, Abidi SK, Campbell JR, Haraoui L-PP, Johnston JC, et al.
369 Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *Bmj*
370 [Internet]. 2020;370:m2516. Available from: <http://www.bmj.com/lookup/doi/10.1136/bmj.m2516>
- 371 7. UK Medicines and Healthcare products Regulatory Agency (MHRA). Target Product Profile:
372 enzyme Immunoassay (EIA) Antibody tests to help determine if people have antibodies to SARS-
373 CoV-2 [Internet]. [cited 2020 Aug 10]. Available from:
374 [https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-](https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-enzyme-immunoassay-eia-antibody-tests-to-help-determine-if-people-have-antibodies-to-sars-cov-2)
375 [work/target-product-profile-enzyme-immunoassay-eia-antibody-tests-to-help-determine-if-people-](https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-enzyme-immunoassay-eia-antibody-tests-to-help-determine-if-people-have-antibodies-to-sars-cov-2)
376 [have-antibodies-to-sars-cov-2](https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-enzyme-immunoassay-eia-antibody-tests-to-help-determine-if-people-have-antibodies-to-sars-cov-2)
- 377 8. England PH, Down P. Evaluation of sensitivity and specificity of four commercially available
378 SARS-CoV-2 antibody immunoassays About Public Health England. 2020;(July).
- 379 9. Duong YT, Wright CG, Justman J. Antibody testing for coronavirus disease 2019: Not ready for
380 prime time. Vol. 370, *The BMJ.* 2020.
- 381 10. Kaza M, Karaźniewicz-Łada M, Kosicka K, Siemiątkowska A, Rudzki PJ. Bioanalytical method
382 validation: new FDA guidance vs. EMA guideline. Better or worse? *J Pharm Biomed Anal.*
383 2019;165:381–5.
- 384 11. Center for Drug Evaluation and Research. Bioanalytical Method Validation Guidance for Industry
385 Bioanalytical Method Validation. FDA Guid Ind [Internet]. 2018;(May):1–22. Available from:
386 <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>
387 [/orhttp://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/](http://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm)
388 [default.htm](http://www.fda.gov/AnimalVeterinary/GuidanceComplianceEnforcement/GuidanceforIndustry/default.htm)
- 389 12. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling Early Humoral Response to

- 390 Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis.* 2020;(Xx Xxxx):1–8.
- 391 13. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in
392 patients of novel coronavirus disease 2019. *Clin Infect Dis.* 2020;
- 393 14. Burbelo PD, Riedo FX, Morishima C, Rawlings S, Smith D, Das S, et al. Sensitivity in Detection
394 of Antibodies to Nucleocapsid and Spike Proteins of Severe Acute Respiratory Syndrome
395 Coronavirus 2 in Patients With Coronavirus Disease 2019. *J Infect Dis.* 2020;
- 396 15. Beavis KG, Matushek SM, Abeleda APF, Bethel C, Hunt C, Gillen S, et al. Evaluation of the
397 EUROIMMUN Anti-SARS-CoV-2 ELISA Assay for detection of IgA and IgG antibodies. *J Clin*
398 *Viro.* 2020;129.
- 399 16. Tang MS, Hock KG, Logsdon NM, Hayes JE, Gronowski AM, Anderson NW, et al. Clinical
400 Performance of Two SARS-CoV-2 Serologic Assays. *Clin Chem.* 2020;
- 401 17. Zheng Z, Monteil VM, Maurer-Stroh S, Yew CW, Leong C, Mohd-Ismail NK, et al. Monoclonal
402 antibodies for the S2 subunit of spike of SARS-CoV-1 cross-react with the newly-emerged SARS-
403 CoV-2. *Euro Surveill.* 2020;
- 404 18. Patrick DM, Petric M, Skowronski DM, Guasparini R, Booth TF, Krajden M, et al. An outbreak of
405 human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus. *Can J*
406 *Infect Dis Med Microbiol.* 2006;17(6):330–6.
- 407 19. Lv H, Wu NC, Tsang OTY, Yuan M, Perera RAPM, Leung WS, et al. Cross-reactive Antibody
408 Response between SARS-CoV-2 and SARS-CoV Infections. *Cell Rep.* 2020;
- 409 20. Che XY, Qiu LW, Liao ZY, Wang Y Di, Wen K, Pan YX, et al. Antigenic cross-reactivity
410 between severe acute respiratory syndrome-associated coronavirus and human coronaviruses 229E
411 and OC43. *J Infect Dis.* 2005;191(12):2033–7.
- 412 21. Chan KH, Cheng VCC, Woo PCY, Lau SKP, Poon LLM, Guan Y, et al. Serological responses in
413 patients with severe acute respiratory syndrome coronavirus infection and cross-reactivity with
414 human coronaviruses 229E, OC43, and NL63. *Clin Diagn Lab Immunol.* 2005;12(11):1317–21.
- 415 22. Hegerova L, Gooley T, Sweerus KA, Maree CL, Bailey N, Bailey M, et al. Use of Convalescent
416 Plasma in Hospitalized Patients with Covid-19 - Case Series. *Blood.* 2020;
- 417 23. Duan K, Liu B, Li C, Zhang H, Yu T, Qu J, et al. Effectiveness of convalescent plasma therapy in
418 severe COVID-19 patients. *Proc Natl Acad Sci U S A.* 2020;117(17):9490–6.
- 419 24. Shen C, Wang Z, Zhao F, Yang Y, Li J, Yuan J, et al. Treatment of 5 Critically Ill Patients with
420 COVID-19 with Convalescent Plasma. *JAMA - J Am Med Assoc.* 2020;323(16):1582–9.
- 421 25. Zhang B, Liu S, Tan T, Huang W, Dong Y, Chen L, et al. Treatment With Convalescent Plasma
422 for Critically Ill Patients With Severe Acute Respiratory Syndrome Coronavirus 2 Infection.
423 *Chest.* 2020;158(1):e9–13.
- 424 26. Ahn JY, Sohn Y, Lee SH, Cho Y, Hyun JH, Baek YJ, et al. Use of convalescent plasma therapy in
425 two covid-19 patients with acute respiratory distress syndrome in Korea. *J Korean Med Sci.*

- 426 2020;35(14).
- 427 27. Ye M, Fu D, Ren Y, Wang F, Wang D, Zhang F, et al. Treatment with convalescent plasma for
428 COVID-19 patients in Wuhan, China. *J Med Virol*. 2020;
- 429 28. Wang C, Li W, Drabek D, Okba NMA, van Haperen R, Osterhaus ADME, et al. A human
430 monoclonal antibody blocking SARS-CoV-2 infection. *Nat Commun*. 2020;
- 431 29. Ni L, Ye F, Cheng ML, Feng Y, Deng YQ, Zhao H, et al. Detection of SARS-CoV-2-Specific
432 Humoral and Cellular Immunity in COVID-19 Convalescent Individuals. *Immunity*. 2020;
- 433 30. To KKW, Tsang OTY, Leung WS, Tam AR, Wu TC, Lung DC, et al. Temporal profiles of viral
434 load in posterior oropharyngeal saliva samples and serum antibody responses during infection by
435 SARS-CoV-2: an observational cohort study. *Lancet Infect Dis*. 2020;
- 436 31. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 viral load in upper
437 respiratory specimens of infected patients. *New England Journal of Medicine*. 2020.
- 438 32. Boukli N, Le Mene M, Schnuriger A, Cuervo NS, Laroche C, Morand-Joubert L, et al. High
439 incidence of false positive results in patients with other acute infections, using the LIAISON®
440 SARS-CoV-2 commercial chemiluminescent micro-particle immunoassay for detection of IgG
441 anti SARS-CoV-2 antibodies. *J Clin Microbiol*. 2020;
- 442 33. GISAID. Genomic epidemiology of hCoV-19 [Internet]. 2020 [cited 2020 Aug 18]. Available
443 from: <https://www.gisaid.org/epiflu-applications/next-hcov-19-app/>
- 444 34. Neil Carey R, Paul Durham A, Hauck WW, Kallner A, Kondratovich M V., Guy Middle J, et al.
445 EP15-A3 User Verification of Precision and Estimation of Bias; Approved Guideline—Third
446 Edition. *Clin Lab Stand Inst*. 2014;34(12):1–81.
- 447
- 448

449 **Tables**

450 Table 1: Respiratory infections tested for cross reactivity in the SARS-CoV-2 IgG Immunoassays.

Infection	No patients
Epstein-Barr virus	8
Influenza A virus	8
Respiratory syncytial virus	7
Seasonal Coronaviruses	7
Borrelia burgdorferii	4
Cytomegalovirus	3
Varicellazoster virus	3
Bordella Pertussis	2
Hepatitis B	2
Human immunodeficiency virus	2
Adenovirus	1
Mycoplasma	1
ParaInfluenza	1
Rhinovirus	1

451

452

453

454

455 Table 2: EP15 analysis on two Abbott Alinity, DiaSorin Liaison XL and ELISAs imprecision tests. For
 456 the DiaSorin, negative samples (QC or pools) results were typically below the limit of detection of 3.8
 457 AU/mL and variation was estimated on the response in relative light units (RLU).

	Sample	n	Mean	Intra-assay imprecision		Inter-assay imprecision	
				SD	%CV	SD	%CV
ABBOTT	Alinity 1 (Neg)	25	0.136 (R)	0.011	8.1	0.011	8.1
	Alinity 1 (Pos)	25	7.254 (R)	0.167	2.3	0.170	2.3
	Alinity 2 (Neg)	25	0.143 (R)	0.007	5.2	0.010	6.8
	Alinity 2 (Pos)	25	7.242 (R)	0.081	1.1	0.082	1.1
DIASORIN	Kit Negative control	20	2457 (RLU)	1730	70.4	2860	116.4
	Level 1 (Neg pool)	25	6945 (RLU)	4003	57.6	5887	84.8
	Kit Positive control	25	58662 (RLU)	4815	8.2	5608	9.6
	Level 2 (Pool 1)	25	83236 (RLU)	11128	13.4	11128	13.4
	Level 3 (Pool 2)	25	410600 (RLU)	56802	13.8	56802	13.8
	Level 4 (pool 3)	25	557660 (RLU)	55667	10.0	58092	10.4
EDI	Kit Negative control	27	0.074 (OD)	-	-	0.014	10.9
	Kit positive control	9	0.482 (OD)	-	-	0.068	14.2
	Duplicate samples	308	-	3.8	3.3	-	-
EURO-IMMUN	Kit Negative control	3	0.074 (OD)	-	-	0.003	3.7
	Kit positive control	3	1.169 (OD)	-	-	0.15	2.9
	Calibrator	3	0.277 (OD)	-	-	0.027	9.5
	Duplicate samples	44	-	6.7	6.1	-	-

458

459 Table 3: Sensitivity of the assays was estimated on all time points and including only samples >14 days
 460 post PCR. Specificity was estimated on pre-2020 samples (N) from healthy individuals and patients with
 461 disorders that induce the production of potentially interfering substances. n/a=no equivocal range
 462 available.


463

		Assay	Total Tested	SARS-CoV-2 IgG Positive	SARS-CoV-2 IgG Negative	Equivocal result	Result (95%CI)
SENSITIVITY	SARS-CoV-2 Positive all time points	EDI	43	35	8	n/a	81 (66-91)
		EuroImmun	43	35	8	0	81 (66-91)
		Abbott	43	35	8	n/a	81 (66-91)
		DiaSorin	42	30	12	0	71 (55-84)
		Healgen	27	27	0	n/a	100 (84-100)
	SARS-CoV-2 Positive ≥14 days post PCR	EDI	20	20	0	n/a	100 (80-100)
		EuroImmun	20	20	0	0	100 (80-100)
		Abbott	20	20	0	n/a	100 (80-100)
		DiaSorin	20	19	1	0	95 (73-100)
		Healgen	20	20	0	n/a	100 (80-100)
SPECIFICITY	Pre-pandemic controls (N)	EDI	50	0	50	n/a	100 (91-100)
		EuroImmun	50	2	46	2	92 (79-97)
		Abbott	50	0	50	n/a	100 (91-100)
		DiaSorin	50	2	48	0	96 (85-99)
		Healgen	4	0	4	n/a	-
	Other Respiratory Infection (CR)	EDI	50	0	50	n/a	100 (91-100)
		EuroImmun	50	3	47	0	94 (82-98)
		Abbott	50	0	50	n/a	100 (91-100)
		DiaSorin	50	2	48	0	96 (85-99)
		Healgen	9	1	8	n/a	-
	Rheumatoid Arthritis (RA)	EDI	22	0	22	n/a	92 (72-99)
		EuroImmun	22	0	22	0	92 (72-99)
		Abbott	22	0	22	n/a	100 (82-100)
		DiaSorin	22	1	21	0	95 (75-100)
		Healgen	4	0	4	n/a	-
	Thyroid Disorder (TSI)	EDI	30	1	29	n/a	97 (81-100)
		EuroImmun	30	0	28	2	93 (76-99)
		Abbott	30	0	30	n/a	100 (85-100)
		DiaSorin	30	1	29	0	97 (81-100)
		Healgen	4	1	3	n/a	-

464

465 Table 4: Cohen’s Kappa concordance analysis of the assays and overall (all samples included) agreement
 466 of results given as %. Equivocal results were considered negative.

467



Abbott	Eurolmmun	DiaSorin	Healgen	Cohen's Kappa (\pm SD) % agreement (Equivocal = POS)
0.950 (0.019) 97.9%	0.907 (0.034) 96.6%	0.892 (0.027) 94.8%	0.745 (0.083) 88.2%	EDI
	0.892 (0.037) 96.1%	0.891 (0.027) 94.8%	0.777 (0.078) 89.6%	Abbott
		0.852 (0.043) 94.7%	0.631 (0.111) 82.7%	Eurolmmun
			0.485 (0.109) 76.1%	DiaSorin

468

469

470

471

472 **Figure Captions**

473 Figure 1: Box-plots of R values for each conditions (N, P, CR, RA and TSI) for the (A) EDI, (B) Abbott,
474 (C) EuroImmuno and (D) DiaSorin tests. Mann-Whitney analysis demonstrated a significant increase in the
475 R value for the positive samples. Mann-Whitney statistical significance * $p < 0.05$; ** $p < 0.01$ and
476 *** $p < 0.001$. Dotted line represents the positive cut-off for each assay.

477

478

479

480 Figure 2: Seropositivity in specimen with PCR positive relative to day of PCR. Dashed line represent the
481 cut-off ratio for each assay. Solid black line and dotted lines represent the 4 parameter logistic curve-fit of
482 the points with confidence interval. Time to PCR onset is calculated as curve inflection point.

