

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

A Serology Strategy for Epidemiological Studies Based on the Comparison of the Performance of Seven Different Test Systems

- The Representative COVID-19 Cohort Munich

Laura Olbrich, MD, ^{1,2,*} Noemi Castelletti, Dr, ^{1,3,*} Yannik Schälte, Msc, ^{4,5,*} Mercè Garí, PhD, ^{4,*} Peter Pütz, PhD, ^{4,6} Abhishek Bakuli, PhD, ¹ Michael Pritsch, MD, ¹ Inge Kroidl, MD, ^{1,2} Elmar Saathoff, PhD, ^{1,2} Jessica Michelle Guggenbuehl Noller, Msc, ¹ Volker Fingerle, MD, ^{2,7} Ronan Le Gleut, Msc, ^{4,8} Leonard Gilberg, ¹ Isabel Brand, ¹ Philine Falk, ¹ Alisa Markgraf, ¹ Flora Deák, ¹ Friedrich Riess, ¹ Dr, Max Diefenbach, Dr, ¹ Tabea Eser, ¹ Franz Weinauer, MD, ⁹ Silke Martin, PhD, ⁹ Ernst-Markus Quenzel, MD, ⁹ Marc Becker, MD, ^{10,11} Jürgen Durner, MD, ^{10,11} Philipp Girl, DVM, ^{2,12} Katharina Müller, DVM, ^{2,12} Katja Radon, PhD, ^{13,14,15} Christiane Fuchs, PhD, ^{4,5,6,8} Roman Wölfel, MD, ^{2,12,+} Jan Hasenauer, PhD, ^{4,5,16+} Michael Hoelscher, MD, ^{1,2,14,+} Andreas Wieser, MD ^{1,2,+,\$}

On behalf of the KoCo19-Study Team

Authors & Affiliations: please see separate sheet

¹ Division of Infectious Diseases and Tropical Medicine, University Hospital, LMU Munich, 80802 Munich, Germany

² German Center for Infection Research (DZIF), partner site Munich, Germany

³ Institute of Radiation Medicine, Helmholtz Zentrum München, 85764 Neuherberg, Germany

⁴ Institute of Computational Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

⁵ Center for Mathematics, Technische Universität München, 85748 Garching, Germany

⁶ Department of Business Administration and Economics, Bielefeld University, 33615 Bielefeld, Germany

⁷ Bavarian Health and Food Safety Authority (LGL)

⁸ Core Facility Statistical Consulting, Helmholtz Zentrum München, 85764 Neuherberg, Germany

⁹ BRK-Blutspendedienst, 80336 Munich, Germany

¹⁰ Department of Conservative Dentistry and Periodontology, University Hospital, LMU Munich Ludwig-Maximilians-University of Munich, Goethestr. 70, 80336 Munich, Germany

¹¹ Laboratory Becker and colleagues, Führichstr. 70, 81671 München, Germany

¹² Bundeswehr Institute of Microbiology, 80937 Munich, Germany

¹³ Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, 80336 Munich, Germany

¹⁴ Center for International Health (CIH), University Hospital, LMU Munich, 80336 Munich, Germany

¹⁵ Comprehensive Pneumology Center (CPC) Munich, German Center for Lung Research (DZL), 80337 Munich, Germany

¹⁶ Faculty of Mathematics and Natural Sciences, University of Bonn, 53113 Bonn, Germany

*Those authors contributed equally

+Those authors contributed equally

§ Corresponding author:

Andreas Wieser, MD

Division of Infectious Diseases and Tropical Medicine

University Hospital, LMU Munich

Leopoldstr. 5, 80802 Munich, Germany

Email: Wieser@mvp.lmu.de

Tel.: +49 89 2180 78 296

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

#KoCo19 study group: Emad Alamoudi, Jared Anderson, Valeria Baldassare, Maximilian Baumann, Marieke Behlen, Jessica Beyerl, Rebecca Böhnlein, Anna Brauer, Vera Britz, Jan Brugger, Friedrich Caroli, Lorenzo Contento, Alina Czwienzek, Emma Dech, Laura Dech, Jana Diekmannshemke, Anna Do, Gerhard Dobler, Ute Eberle, Judith Eckstein, Jonathan Frese, Stefanie Fischer, Felix Forster, Turid Frahnöw, Günter Fröschl, Marius Gasser, Sonja Gauder, Otto Geisenberger, Christof Geldmacher, Kristina Gillig, Elias Golschan, Vitus Grauvogl, Celina Halfmann, Tim Haselwarter, Arlett Heiber, Matthias Herrmann, Stefan Hillmann, Christian Hinske, Janna Hoefflin, Tim Hofberger, Michael Höfinger, Larissa Hofmann, Sacha Horn, Kristina Huber, Christian Janke, Ursula Kappl, Charlotte Kiani, Isabel Klugherz, Norah Kreider, Arne Kroidl, Magdalena Lang, Clemens Lang, Silvan Lange, Ekaterina Lapteva, Michael Laxy, Reiner Leidl, Felix Lindner, Paula Matcau, Rebecca Mayrhofer, Anna-Maria Mekota, Hannah Müller, Dafni Metaxa, Katharina Müller, Leonie Pattard, Claire Pleimelding, Michel Pletschette, Stephan Prückner, Kerstin Puchinger, Konstantin Pusl, Elba Raimúndez, Julius Raschka, Jakob Reich, Camila Rothe, Raquel Rubio-Acero, Nicole Schäfer, Paul Schandelmaier, Lara Schneider, Sophie Schultz, Mirjam Schunk, Lars Schwettmann, Heidi Seibold, Peter Sothmann, Paul Stapor, Jeni Tang, Fabian Theis, Verena Thiel, Sophie Thiesbrummel, Eva Thumser, Niklas Thur, Julian Ullrich, Julia Waibel, Claudia Wallrauch, Simon Winter, Julia Wolff, Pia Wullinger, Tobias Würfel, Patrick Wustrow, Houda Yaqine, Sabine Zange, Eleftheria Zeggini, Thomas Zimmermann, Lea Zuche

Author's contribution: MH is the principal investigator and obtained funds. AW, MH, JH and RW designed the study and experimental setup with input from CF, KR, MB and JD. The clinical work and study design of KoCo19 was led by LO, MP and IK, with support by JMGN, LG, PF, and AM. BSL-3 work and viral RT-PCRs measurements were planned, conducted and evaluated by MK, JD, PG, KM and RW. Serological testing was designed and conducted with AW, VF, PG, KM, MB, PG, KM, RW, FD, and TE, while SM, EMQ and FW contributed participants' samples. Data was cleaned and prepared by NC, MND, and SW. Statistical analysis was performed by NC, YS and MG, with support by PP, AB, ES, RLG, MD, CF and JH. The manuscript was primarily written by LO, NC and AW, with contributions from YS, MG, CF, JH and MH, and input of IK and MP. All authors revised and approved the final version of this manuscript.

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

1 **Abstract:**

2 **Background.** Serosurveys are essential to understand SARS-CoV-2 exposure and enable population-level
3 surveillance, but currently available tests need further in-depth evaluation. We aimed to identify testing-strategies
4 by comparing seven seroassays in a population-based cohort.

5 **Methods.** We analysed 6,658 samples consisting of true-positives (n=193), true-negatives (n=1,091), and
6 specimens of unknown status (n=5,374). For primary testing, we used Euroimmun-Anti-SARS-CoV-2-ELISA-IgA/IgG
7 and Roche-Elecsys-Anti-SARS-CoV-2; and virus-neutralisation, GeneScript®cPass™, VIRAMED-SARS-CoV-2-
8 ViraChip®, and Mikrogen-*recom*Line-SARS-CoV-2-IgG, including common-cold CoVs, for confirmatory testing.
9 Statistical modelling generated optimised assay cut-off-thresholds.

10 **Findings.** Sensitivity of Euroimmun-anti-S1-IgA was 64.8%, specificity 93.3%; for Euroimmun-anti-S1-IgG,
11 sensitivity was 77.2/79.8% (manufacturer's/optimised cut-offs), specificity 98.0/97.8%; Roche-anti-N sensitivity
12 was 85.5/88.6%, specificity 99.8/99.7%. In true-positives, mean and median titres remained stable for at least 90-
13 120 days after RT-PCR-positivity. Of true-positives with positive RT-PCR (<30 days), 6.7% did not mount detectable
14 seroresponses. Virus-neutralisation was 73.8% sensitive, 100.0% specific (1:10 dilution). Neutralisation surrogate
15 tests (GeneScript®cPass™, Mikrogen-*recom*Line-RBD) were >94.9% sensitive, >98.1% specific. Seasonality had
16 limited effects; cross-reactivity with common-cold CoVs 229E and NL63 in SARS-CoV-2 true-positives was
17 significant.

18 **Conclusion.** Optimised cut-offs improved test performances of several tests. Non-reactive serology in true-
19 positives was uncommon. For epidemiological purposes, confirmatory testing with virus-neutralisation may be
20 replaced with GeneScript®cPass™ or *recom*Line-RBD. Head-to-head comparisons given here aim to contribute to
21 the refinement of testing-strategies for individual and public health use.

22

23

24 **Keywords:** SARS-CoV-2, serology, KoCo19, virus neutralisation

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

25 Introduction

26 In December 2019, a cluster of atypical pneumonia of unknown origin was described in the region of Wuhan,
27 Hubei province, China. Subsequently, a novel coronavirus was identified as the causative pathogen: SARS-CoV-2
28 (Severe Acute Respiratory Syndrome Coronavirus 2).(1) As the virus spread rapidly across the globe, the Corona
29 Virus Disease 2019 (COVID-19) was declared a pandemic on March 12th, 2020.

30 Direct detection of viral nucleic acids or the virus itself in bodily fluids is considered the reference standard of
31 acute infection; primarily identified through the use of nasopharyngeal swabs or other respiratory probes.(2)
32 Additionally, serodiagnostics are valuable in order firstly to identify past infections, asymptomatic or symptomatic,
33 and secondly to elucidate transmission dynamics within the population; both being highly relevant to inform
34 evidence-based political decision making.(3, 4)

35 Several serological test systems have been introduced since the beginning of the pandemic.(5) Most target one of
36 two specific viral structures: parts of the trimeric CoV spike (S1-2) complex or the nucleocapsid (N) envelope
37 protein.(6) While the receptor binding domain (RBD) of S1 folds out to bind to the angiotensin-converting enzyme
38 2 (ACE2) receptor, the N-protein is involved in viral assembly and replication.(7) To select the most suitable system
39 for different applications, independent evaluations and comparisons of their sensitivities and specificities on
40 different cohorts are of utmost importance; yet, as Cheng et al. state, validations of these tests are often poorly
41 described and evaluations show several shortcomings.(7) Firstly, as most validation studies have been performed
42 with severely sick subjects, proposed cut-off values are ambiguous.(4, 8) Those individuals will likely have higher
43 titres than asymptomatic subjects,(6, 9-12) possibly resulting in a less reliable performance of the test system, or
44 the potential need for different cut-off values in these populations.(13) Secondly, it remains unclear whether the
45 antibodies are serocorrelates of protection, highlighting the gap in knowledge of the underlying biology.(3, 7) And
46 lastly, cross-reactivity with other CoVs has been reported, but its effects on the test readouts remain
47 uncertain.(14-16)

48 Here we present the serological testing systems applied within the Representative COVID-19 Cohort Munich
49 (KoCo19), a prospective sero-incidence study initiated in Munich, Germany, in April 2020.(17, 18) Three
50 independent primary tests (Euroimmun Anti-S1-SARS-CoV-2-ELISA-IgA & -IgG and Elecsys Anti-SARS-CoV-2 Roche
51 N pan-Ig) and a number of confirmatory tests (direct viral neutralisation, GeneScript[®]cPass[™], Mikrogen-
52 *recom*Line-RBD IgG line immunoassay, VIRAMED-SARS-CoV-2-ViraChip[®] microarray) were assessed in a head-to-
53 head cross-comparison. The tests were conducted on samples from (i.) RT-PCR positive individuals (true-positives),
54 (ii.) blood donors from the pre-COVID-19 era (true-negatives), and (iii.) subjects with unknown disease status
55 (unknown). Thus, we were able to generate reliable performance estimates for both primary and confirmatory
56 tests by using true-positive and true-negative individuals from the aforementioned cohorts and to subsequently
57 derive optimised cut-offs for our cohort.

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

58 **Methods**

59 **Study design and participants.** In total, we included 6,658 samples, derived from a set of SARS-CoV-2 RT-PCR
60 positives (“true-positives”, n=193), individuals from historical cohorts, blood donors without any indication of
61 SARS-CoV-2 infection (“true-negatives”, n=1,091), and specimen of unknown status (n=5,374); details on the
62 cohort characteristics, including collection-timepoints, can be found in the appendix (p.1; p.4 supplemental table
63 1).

64 The study was approved by the Ethics Commission of the Faculty of Medicine at LMU Munich (20-275-V) and the
65 protocol is available online (www.koco19.de).¹⁷

66

67 **Laboratory Assays.** All described analyses were performed using EDTA-plasma samples (appendix pp.1 for further
68 details on assays performed, and p.7 supplemental table 3 for details on platforms and units applied).

69 Euroimmun Anti-SARS-CoV-2-ELISA anti-S1 IgA/IgG (hereafter called EI-S1-IgG, EI-S1-IgA; Euroimmun, Lübeck,
70 Germany) test kits were used according to the manufacturer's instructions. Measurement values were obtained
71 using the quotient of the optical density measurement provided by the manufacturer's software. We evaluated
72 Elecsys Anti-SARS-CoV-2 Roche anti-N pan-Ig (hereafter called Ro-N-Ig; Roche, Mannheim, Germany) in
73 accordance with the manufacturer's guidelines. Values reported are the Cut-Off-Index (COI) of the individual
74 samples. Operative replicates of the same samples were performed to assess reliability of primary assay
75 performance.

76 For confirmatory testing, we conducted micro-virus neutralisation assays (NT) as described previously,⁽¹⁹⁾ with
77 the exception that confluent cells were incubated instead of adding cells following neutralisation reaction
78 (appendix pp.1). We classified samples with a titre <1:10 as "NT-negative" and samples with a titre ≥1:10 as "NT-
79 positive". The dilution steps indicated are <10, 10, 20, 40, and >80.

80 SARS-CoV-2 surrogate virus neutralisation test (GS-cPass; GenScript®, Piscataway, New Jersey, USA) was used to
81 measure binding inhibition, according to the manufacturer's instructions. The inhibition was calculated in
82 percentages, ranging from -30 to 100.

83 For SARS-CoV-2 ViraChip® microarray (VIRAMED Biotech AG, Planegg, Germany; hereafter named VC-N-
84 IgA/IgM/IgG; VC-S1-IgA/IgM/IgG; VC-S2-IgA/IgM/IgG) execution followed the manufacturer's instructions. We
85 obtained measurement values by the automated ELISA-processor in arbitrary units.

86 As outlined by the manufacturer, we conducted the *recom*Line SARS-CoV-2 IgG line immunoassay (MG-S1, MG-N,
87 MG-RBD; Mikrogen, Neuried, Germany), values below the cut-off of 1 were categorized “negative” without
88 quantitative information. Common-cold CoV-2 targets of 229E, NL63, OC43, and HKU1 were included in the
89 aforementioned assay.

90

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

91 **Statistical Analysis.** Prior to analysis, we cleaned and locked the data. For the analyses and visualisation, we used
92 the software R, version 4.0.2. Only one sample per individual was included in the statistical analyses; in case of
93 individuals with multiple blood samples, we only considered the sample with the most complete dataset. For
94 multiple measurements with complete datasets, we only included the first measurement; for operational
95 replicates we used the latest. We subsequently carried out sensitivity and specificity analyses for true-negative
96 and true-positive samples over all the tests performed.

97 We report square roots R of coefficients of determination for association among continuous variables. For paired
98 sample comparisons, we applied Wilcoxon-sign-rank tests, whereas for multiple group comparisons we applied
99 Kruskal-Wallis tests, followed by post-hoc Dunn tests using the Benjamini-Yekutieli adjustment for pairwise
100 comparisons.(20)

101 Using true-positives and true-negatives, we determined optimised cut-off thresholds and their confidence
102 intervals by a nonparametric bootstrap. In a similar way, we trained random forest and support vector machine
103 classifiers. We calculated estimates for sensitivities, specificities, and overall prediction accuracies for all
104 considered cut-off values and classifiers. This calculation was done on out-of-sample observations to avoid
105 overfitting and thus overoptimistic performance measures. Details on the algorithms are outlined in the appendix
106 (pp.2).

107

108 **Data & Code sharing.** Data are accessible subject to data protection regulations upon reasonable request to the
109 corresponding author. To facilitate reproducibility and reuse, the code used to perform the analyses and generate
110 the figures was made available on GitHub (https://github.com/koco19/lab_epi) and has been uploaded to
111 ZENODO (<http://doi.org/10.5281/zenodo.4300922> , DOI 10.5281/zenodo.4300922) for long-term storage.

112

113 **Role of funding source.** The funders had no role in study design, data collection, data analysis, data interpretation,
114 writing or submission for publication of this manuscript.

115

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

116 **Results**

117 We assessed SARS-CoV-2 antibodies in a total of 6,658 independent samples using EI-S1-IgA (n=6,657), EI-S1-IgG
118 (n=6,658), and Ro-N-Ig (n=6,636) (figure 1, appendix p.4 supplemental table 1). Supplemental table 2 presents an
119 overview of all tests performed (appendix p.5); table 1 outlines sensitivity and specificity of both primary and
120 confirmatory tests by applying both manufacturers' and optimised cut-offs.

121
122 **Performance of Primary Tests.** When we applied optimised cut-offs to the true-positive and true-negative
123 samples, EI-S1-IgA had a sensitivity of 64.8% and a specificity of 92.6%, while EI-S1-IgG had a sensitivity of 79.8%
124 and a specificity of 97.8% (table 1). The measurement value distribution of the EI-assays is depicted in figure 2A.
125 In the second half of the study period, a decrease in overall IgA-positivity was observed, while overall IgG-positivity
126 remained stable. Subsequent retesting of biobanked samples of the first weeks of sampling with kits of later
127 batches demonstrated overall lower IgA-positivity rates than in the first measurements. This was likely caused by
128 a batch effect, as retesting of samples from the second half of the sampling period did not reproduce these
129 findings; for the first period, it led to changes in classification from positive to negative in the operational replicates
130 (depicted as "positive-negative" in figure 2B and appendix p.8 supplemental figure 1A).

131 The sensitivity and specificity of Ro-N-Ig with optimised cut-offs were 88.6% and 99.7%, respectively (table 1). Re-
132 testing measurement values confirmed a correlation of R=1 for the quantitative measurement values without a
133 single classification change (0/423) (figure 2B, appendix p.8 supplemental figure 1B). For all three assays, few
134 samples tested false-negative (appendix p.26, supplemental figure 8).

135 For evaluation of primary test concordance, we excluded EI-S1-IgA due to inferior performance in sensitivity and
136 specificity. The concordance between EI-S1-IgG and Ro-N-Ig was 98.5% (6,538/6,636), where 4.0% (264/6,636) of
137 the samples were classified as positive and 94.5% (6,274/6,636) as negative unanimously. Of the remaining
138 samples (1.5%; 98/6,636), 56.1% (55/98) were classified as positive by EI-S1-IgG and negative by Ro-N-Ig (figure
139 4A), while the remaining 43.9% (43/98) were negative for EI-S1-IgG and positive for Ro-N-Ig. To clarify their true
140 serostatus, confirmatory testing was performed where possible (appendix p.26 supplemental figure 8).

141 To assess the dependence of result read-out on sampling timepoints, we considered the temporal distribution of
142 baseline titres over the sampling period (appendix pp.11 supplemental figure 2). Here, we considered the mean
143 and median sample values above and below cut-off separately and found them to be comparable over the whole
144 sampling period. Samples from blood donors were available from two distinct sample timepoints (before and after
145 the common cold season). We noticed discrete baseline titre increases in spring without significant changes in
146 overall positivity rate (appendix p.14 supplemental figure 3).

147 To assess the dynamics of the antibody levels in RT-PCR positive subjects, we analysed the measurement values
148 between the RT-PCR test and blood sampling for three time intervals: <30 days, 30-90 days, and >90 days. Across

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

149 the whole cohort, we did not observe an initial rise of EI-S1-IgA and EI-S1-IgG but found that EI-S1-IgA decreased
150 at later time points ($p=0.02$), while the distribution of EI-S1-IgG remained steady (figure 3). For Ro-N-Ig COI we
151 detected an increase in antibody levels ($p<0.001$), most notably between <30 and $30-90$ days ($p=0.02$).

152

153 **Performance of confirmatory tests.** A sample subset ($n=362$; figure 1) underwent confirmatory testing. The
154 overall confirmatory test performance is presented in figure 5 and table 1. The sensitivity was 73.8% for direct NT
155 (1:10 dilution), 96.3% for GS-cPass, and 94.9% for MG-RBD. All three tests had a specificity close to 100% (figure
156 5).

157 NT-titres in our cohort were low – mostly 1:10 – and few had high neutralisation titres of 1:80 or above (figure
158 5A). Measurement values for MG-S1 and MG-RBD were often similar, with few samples showing a reaction against
159 S1 (and N) but not RBD (3.7%; 5/134). For the VC-array, sensitivities of both VC-S1-IgG and VC-N-IgG were
160 improved markedly by optimising cut-offs, with gains of $>30\%$ (VC-N-IgG 39.8%/93.5%; VC-S1-IgG 65.8%/93.4%;
161 table 1, figure 5). Notably, a considerable number of samples (36.0%; 40/111) were reactive against S1 and N but
162 not S2 (figure 5D).

163 Figure 6 and supplemental figures 4, 5, and 10 (appendix pp.16) compare confirmatory tests. For surrogates of
164 viral neutralisation, 3.2% (3/95) were positive in NT and GS-cPass but not in MG-RBD (figure 6D).

165

166 **Associations of confirmatory and primary tests.** To examine pretest probability of assays following positive initial
167 testing, the measurement values of all primary and confirmatory tests were correlated (figure 7, appendix pp.19
168 supplementary figures 6, 7, 9). Overall, we observed good correlations, particularly for GS-cPass and MG-RBD with
169 EI-S1-IgG (figure 7B, C), and MG-N with Ro-N-Ig (figure 7H).

170 The categorical concordance for GS-cPass, MG-RBD, and MG-N with both Ro-N-Ig and EI-S1-IgG was similar (94%
171 or above), while the concordance of NT with both primary tests was lower (74%; figure 7A-H). Concordances were
172 improved by applying the optimised cut-offs, especially for VC-S1-IgG and VC-S2-IgG.

173 To derive an algorithm for an optimised testing-strategy, we also investigated potential gains when applying
174 multiple primary tests. Interestingly, combining different primary tests via machine learning techniques (random
175 forest and support vector machine) barely improved the performance beyond what was achieved by Ro-N-Ig alone
176 (table 1), neither could we demonstrate an added value of performing more than one confirmatory test.

177

178 **Common cold line blot.** Cross-reactivity between common cold CoV anti-N and SARS-CoV-2 anti-N was assessed
179 by line blot against 229E, NL63, OC43, and HKU1 for various samples subjected to confirmatory testing
180 ($n=273/362$; appendix p.27, supplemental figure 9). The distribution for 229E and NL63 showed a significant
181 association with SARS-CoV-2 infection and positivity ($p<0.001$ for both).

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

182 **Discussion**

183 We performed head-to-head comparison of seven seroassays for SARS-CoV-2 in a well-defined cohort with 6,658
184 samples. The overwhelming majority of infections could be detected by antibodies even more than three months
185 after infection. In addition, we generated optimised cut-offs which improved sensitivity in primary tests and
186 enhanced the performance of several confirmatory tests. We furthermore showed that surrogate tests, such as
187 GeneScript®cPass™ or RBD line-blot, should be considered instead of the infrastructurally challenging NT in large
188 scale studies, such as vaccine trials or epidemiological surveys. In addition, we observed limited seasonality effects
189 and a significant cross-reactivity with common-cold CoV 229E and NL63 in SARS-CoV-2 true-positive subjects.

190
191 We calculated diagnostic accuracy indicators for seroassays based on true-positive and true-negative samples and
192 derived optimised cut-offs for the evaluated assays. We subsequently applied these to a set of samples of
193 unknown infection status.(17) As reported previously, there is little diagnostic gain from EI-S1-IgA if EI-S1-IgG is
194 used on the same samples.(4) Our study suggests that Ro-N-Ig performs more reliably than EI-S1-IgA and EI-S1-
195 IgG; hence the application of the latter two is of questionable use for low-prevalence settings, considering their
196 lower specificity.

197 For seroprevalence studies it is essential to know how long the measured antibody response remains stable and
198 detectable. There have been different reports of a rapid decline in titre over time.(10, 21) In this analysis, we
199 observed similar declines for EI-S1-IgA titres, which were highest within the first 30 days but dropped significantly
200 thereafter. In contrast, for both EI-S1-IgG and Ro-N-Ig, average titres did not decrease over a period of more than
201 90 days.

202
203 We assessed confirmatory test performances by a subset of true-negative samples. The assays targeting the
204 highly-specific receptor binding domain (NT, GS-cPass, MG-RBD) are considered direct or surrogate markers for
205 viral neutralisation.(22) The cell-culture free tests performed particularly well with sensitivities of 96.3% for GS-
206 cPass and 94.9% for MG-RBD, using the manufacturer's thresholds. In contrast, we improved the sensitivity for
207 the VC-array markedly by applying optimised cut-offs, with a gain of >50% for VC-N-IgG (39.8%-93.5%) and almost
208 30% for VC-S1-IgG (65.8-93.4%).

209 To generate reliable data for surrogate neutralisation markers (GS-cPass, MG-RBD) feasible for epidemiological
210 purposes, true-positive samples with low neutralising activity are preferable. Our cohort consists mainly of oligo-
211 or asymptomatic COVID-19 subjects – NT-titres observed were generally low as expected – suggesting an NT
212 dilution decrease from 1:10 to 1:5 might improve sensitivity in similar settings. As MG-RBD, GS-cPass, and NT
213 performed similarly and correlated well, suggesting an interchangeability of one another, for example for certain
214 epidemiological questions. NT requires complex BSL-3-laboratory infrastructures (virus-culture) which currently

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

215 represents a critical bottleneck, while GC-cPass and MG-RBD can be performed under much less stringent safety
216 requirements. While NT is a direct representation of viral neutralisation, GS-cPass assesses the antibody-mediated
217 inhibition of ACE2-interaction with SARS-CoV-2-S1-RBD and is therefore a cell-free neutralisation surrogate
218 marker.(23) Both assays are independent of antibody-subclasses. In contrast, MG-RBD solely detects the RBD-IgG-
219 antibody interaction. The discrepant results in our sample set (figure 6D) suggest either a different neutralisation
220 mode than binding to the RBD as used in MG-RBD, or neutralisation due to other subclasses than IgG.

221

222 The influence of previous or acute infections with common-cold CoV and a resulting cross-reactivity in SARS-CoV-
223 2 assays has been discussed previously.(15, 16) Biologically, it is impacted by the target used, wherein whole virus-
224 or nucleocapsid proteins are particularly prone to cross-reactivity; this also applies for SARS-CoV-2.(6, 14, 15, 24)
225 Comparing true-negative samples from before and after the common cold season, we observed a marginal
226 seasonal increase in the measurement values of EI-S1-IgA, EI-S1-IgG, and Ro-N-Ig (supplemental figure 3). There
227 was a clear association between SARS-CoV-2 RT-PCR positivity and reactivity against 229E and NL63 N-protein;
228 this suggests that SARS-CoV-2 infected individuals develop a cross-reactivity in the assay against two of the four
229 tested CoV strains. Alternatively, in subjects with previous common cold, this could be explained by a cross-
230 stimulation of pre-existing cells specific to the respective common cold coronavirus strain. Similar findings were
231 described recently in a systematic review.(6)

232

233 Using the best-performing primary tests (EI-S1-IgG and Ro-N-Ig), we observed a relevant number of discordantly
234 classified samples (1.5%; 98/6,636; appendix p. 27 supplemental figure 8). Constellations of Ro-N-Ig negative and
235 EI-S1-IgG positive were observed frequently (56.1%; 55/98), with many samples being true-negatives (41.8%;
236 23/55), but none being true-positive (0%; 0/55). The overwhelming majority of those with unknown infection
237 status were negative in NT and GS-cPass (87.5%; 28/32). As the testing principle of EI-S1-IgG is based on S1-
238 reactivity, positive virus neutralisation or a neutralisation surrogate can be expected if the reaction is specific,
239 suggesting unspecific cross-reactivity in this group.

240 In contrast, discordant serology with Ro-N-Ig positive and EI-S1-IgG negative was less frequent (43.9%; 43/98),
241 mostly from true-positives >40 days after positive RT-PCR (16.3%; 16/98). This is strongly suggestive of an inability
242 of these individuals to mount a detectable response against S1 during the natural infection. Late or lacking
243 seroconversion has been described previously, mostly in oligo- or asymptomatic subjects.(25, 26) Reports
244 speculated about vastly varying proportions of subjects unable to mount an antibody response detectable by
245 commonly used assays.(21, 27, 28) In a systematic review by Huang et al., the median detection time across
246 different antibodies against SARS-CoV-2 was 11 days, similar to SARS-CoV-1;(6) however, serological kinetics vary
247 across the severity gradients of symptoms.(6, 9-11) Additionally, misclassification of subjects might have occurred

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

248 due to RT-PCR-sample mix-up, unspecific reactivity, or contamination in the PCR-test; plausible scenarios in the
249 current pandemic situation where clinical and laboratory infrastructures are overburdened with high sample
250 volumes. In oligo- or asymptomatic subjects, a positive RT-PCR is often the only evidence of infection with SARS-
251 CoV-2 and therefore a misclassification might not be detected.

252 In our cohort of mostly oligosymptomatic true-positive subjects, 11.4% (22/193) were not detected in the primary
253 serological tests. Of these, 36.4% (8/22) were below 30 days after RT-PCR. A total of 6.0% (7/116) showed no
254 evidence of seroconversion in the group >90 days after RT-PCR. For one subject, the RT-PCR date was unknown.
255 Overall, in our dataset of samples >30 days after positive RT-PCR, a modest 8.1% (13/160) remained negative.

256
257 In addition to identifying suitable tests dependent upon the approach chosen, applying optimised cut-offs might
258 be a tool to enhance test performance.(4, 29) For EI-S1-IgG, Meyer et al. proposed optimised thresholds for both
259 lower and higher cut-offs.(4) However, a seroprevalence study in Geneva compared both recommended and
260 optimised cut-offs, but did not see any qualitative change.(30) The authors proposed evaluating the
261 manufacturer's cut-off before routine testing, and highlighted the dilemma of securing both rule-in and rule-out
262 properties to mitigate the risk of incorrect classification in a situation with highly-dynamic pre-test probabilities.(4)
263 For both EI-S1-IgG and Ro-N-Ig we demonstrated that sensitivity improved using optimised cut-offs, while
264 specificity remained similar. Even though these changes seem minimal, they might translate into a high number
265 of incorrect diagnoses when testing is performed on a large scale: this is especially pertinent in low-prevalence
266 settings, as a particularly high specificity is crucial to achieve a high positive predictive value. It may also be
267 preferable to have a more sensitive cut-off for a primary test and confirm the positives with a highly-specific
268 secondary test system in low prevalence settings.(17)

269
270 Our study has several of limitations. Firstly, the samples are mostly derived from subjects with mild to no
271 symptoms and therefore conclusions drawn here might best apply to an epidemiological rather than a diagnostic
272 approach. Further evaluation using specimens from severely ill subjects or from high-prevalence settings will need
273 to be conducted and compared to the enclosed results.

274 Secondly, the dataset used to estimate optimised cut-offs and performance of the tests is not an unbiased random
275 sample, as all were sampled in the city of Munich, Germany. It cannot be ruled out that blood donors are generally
276 healthier than the general population. Furthermore, not all confirmatory tests were performed on all samples;
277 only a subset, namely those with positive results in at least one primary test and a dedicated negative/positive
278 cohort, were tested using these systems. Finally, we did not have information on underlying health conditions
279 known to increase the quantity of polyclonal antibodies.

280

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

281 In conclusion, our study provides a head-to-head cross-comparison of tests and can be used as a resource to
282 enable the refinement of testing-strategies for individual and public-health use. We also investigated the
283 correlations of the different test systems in-depth, considering confounders such as seasonality and titres against
284 common-cold CoV strains. By combining a diagnostic accuracy approach in a well-defined sample set with true-
285 positive as well as true-negative specimens and extrapolating the established findings to samples derived from a
286 population-based seroprevalence cohort, we were able to identify reliable test systems for our Prospective COVID-
287 19 Cohort Munich (KoCo19).(17)

288
289

290 **Acknowledgements**

291 We wholeheartedly thank all study participants for their trust, time, data, and samples. We are grateful for the
292 financial support of the Bavarian State Ministry of Science and the Arts, the University Hospital of LMU Munich,
293 the Helmholtz Centre Munich, the University of Bonn, the University of Bielefeld, and the German Ministry for
294 Education and Research. This study would also not have been possible without the passionate contribution of the
295 staff of the Division of Infectious Diseases and Tropical Medicine at the University Hospital, LMU Munich,
296 Helmholtz Centre Munich, Bundeswehr Institute of Microbiology, as well as all medical students involved. We
297 thank Judith Eckstein for outstanding support regarding public relations. We thank the teams from the press
298 offices of LMU, University Hospital of LMU Munich, and of Helmholtz Centre Munich. We thank the KoCo19
299 advisory board members Stefan Endres, Stephanie Jacobs, Bernhard Liebl, Michael Mihatsch, Matthias Tschöp,
300 Manfred Wildner, and Andreas Zapf. We thank Accenture for the development of the KoCo19 web-based survey
301 application. We are grateful to the Statistical Office of the City of Munich, Germany, for providing statistical data
302 on the Munich general population. We thank Helmut Küchenhoff for a critical review of an earlier version of the
303 manuscript and Jared Anderson for English language corrections. We are grateful to the Munich police for their
304 support in the fieldwork and for the Munich Surgical Imaging GmbH, Cisco Systems, and the graphic/photo/IT
305 infrastructure departments at the University Hospital of LMU Munich provided support during video production
306 and online events. For their assistance with our fieldwork, we thank the BMW Group for providing free cars as a
307 part of their campaign “BMW hilft Helfenden.” We would also like to thank Mercedes-Benz Munich who supported
308 the project infrastructure with Mercedes-Benz Rent. Lastly, we thank the IT Infrastructure Department of the LMU
309 University Hospital Munich. MG acknowledges the support from the Joachim Herz Foundation through the Add-
310 on Fellowship for Interdisciplinary Science.

311
312
313

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

314 **Declaration of Interests**

315 AW and MH report personal fees and non-financial support from Roche Diagnostics, LO reports non-financial
316 support from Roche Diagnostics. AW, MH and LO report non-financial support from Euroimmun, non-financial
317 support from Viramed, non-financial support from Mikrogen. AW, MH, LO report grants, non-financial support
318 and other from German Center for Infection Research DZIF, grants and non-financial support from Government
319 of Bavaria, non-financial support from BMW, non-financial support from Munich Police, non-financial support and
320 other from Accenture. JH reports grants from German Federal Ministry of Education and Research, during the
321 conduct of the study. MH and AW report personal fees and non-financial support from Dr.Box-Betrobox, non-
322 financial support from Dr. Becker MVZ during the conduct of the study. In addition, MH, AW, MB have a patent
323 on a sample system for sputum diagnostics of SARS-CoV-2 pending. AW is involved in other different patents and
324 companies not in relation with the serology of SARS-CoV-2. AW reports personal fees and other from Haeraeus
325 Sensors, non-financial support from Bruker Daltonics, all of which are outside the submitted work, and non-related
326 to SARS-CoV-2. MB is an authorized representative partner of Dr. Becker MVZ.

327

328 **Funding**

329 Bavarian State Ministry of Science and the Arts, University Hospital, LMU Munich, Helmholtz Centre Munich,
330 University of Bonn, University of Bielefeld, German Ministry for Education and Research (proj. nr.: 01KI20271).
331 Euroimmun, Roche, Mikrogen, Viramed provided kits and machines for analyses at discounted rates.

332

333

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

334 References

- 335 1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019
336 novel coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497-506.
- 337 2. Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova V, McMahon M, et al. A
338 serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat Med*. 2020;26(7):1033-6.
- 339 3. Chen X, Pan Z, Yue S, Yu F, Zhang J, Yang Y, et al. Disease severity dictates SARS-CoV-2-
340 specific neutralizing antibody responses in COVID-19. *Signal Transduction and Targeted Therapy*.
341 2020;5(1):180.
- 342 4. Meyer B, Torriani G, Yerly S, Mazza L, Calame A, Arm-Vernez I, et al. Validation of a commercially
343 available SARS-CoV-2 serological immunoassay. *Clinical Microbiology and Infection*. 2020;26(10):1386-
344 94.
- 345 5. (FIND) FfIND. SARS-CoV-2 diagnostic pipeline 2020 [Available from:
346 <https://www.finddx.org/covid-19/pipeline/>.
- 347 6. Huang AT, Garcia-Carreras B, Hitchings MDT, Yang B, Katzelnick LC, Rattigan SM, et al. A
348 systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and
349 association with severity. *Nature Communications*. 2020;11(1):4704.
- 350 7. Cheng MP, Yansouni CP, Basta NE, Desjardins M, Kanjilal S, Paquette K, et al. Serodiagnostics
351 for Severe Acute Respiratory Syndrome-Related Coronavirus 2 : A Narrative Review. *Ann Intern Med*.
352 2020;173(6):450-60.
- 353 8. Nuccetelli M, Pieri M, Grelli S, Ciotti M, Miano R, Andreoni M, et al. SARS-CoV-2 infection
354 serology: a useful tool to overcome lockdown? *Cell Death Discovery*. 2020;6(1):38.
- 355 9. Long Q-X, Tang X-J, Shi Q-L, Li Q, Deng H-J, Yuan J, et al. Clinical and immunological
356 assessment of asymptomatic SARS-CoV-2 infections. *Nature medicine*. 2020:1-5.
- 357 10. Seow J, Graham C, Merrick B, Acors S, Steel KJA, Hemmings O, et al. Longitudinal evaluation
358 and decline of antibody responses in SARS-CoV-2 infection. *medRxiv*. 2020:2020.07.09.20148429.
- 359 11. Eyre DW, Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, et al. Stringent
360 thresholds for SARS-CoV-2 IgG assays result in under-detection of cases reporting loss of taste/smell.
361 *medRxiv*. 2020:2020.07.21.20159038.
- 362 12. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in
363 patients of novel coronavirus disease 2019. *Clin Infect Dis*. 2020.
- 364 13. Ainsworth M, Andersson M, Auckland K, Baillie JK, Barnes E, Beer S, et al. Performance
365 characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. *The
366 Lancet Infectious Diseases*.
- 367 14. Meyer B, Drosten C, Müller MA. Serological assays for emerging coronaviruses: challenges and
368 pitfalls. *Virus research*. 2014;194:175-83.
- 369 15. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell
370 Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals.
371 *Cell*. 2020;181(7):1489-501.e15.
- 372 16. Klompus S, Leviatan S, Vogl T, Kalka I, Godneva A, Shinar E, et al. Cross-reactive antibody
373 responses against SARS-CoV-2 and seasonal common cold coronaviruses. *medRxiv*.
374 2020:2020.09.01.20182220.
- 375 17. Pritsch M, et al. Prevalence and risk factors of infection in the representative COVID-19 cohort
376 Munich. *Adjacent manuscript*.
- 377 18. Radon K, Saathoff E, Pritsch M, Guggenbühl Noller JM, Kroidl I, Olbrich L, et al. Protocol of a
378 population-based prospective COVID-19 cohort study Munich, Germany (KoCo19). *BMC Public Health*.
379 2020;20(1):1036.
- 380 19. Haselmann V, Özçürümez MK, Klawonn F, Ast V, Gerhards C, Eichner R, et al. Results of the
381 first pilot external quality assessment (EQA) scheme for anti-SARS-CoV2-antibody testing. *Clinical
382 chemistry and laboratory medicine*. 2020.
- 383 20. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under
384 dependency. *Annals of statistics*. 2001:1165-88.

Disclaimer: this is a pre-print manuscript and has not been peer-reviewed yet

- 385 21. Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-
386 2 in patients with COVID-19. *Nat Med.* 2020;26(6):845-8.
- 387 22. Chi X, Liu X, Wang C, Zhang X, Li X, Hou J, et al. Humanized single domain antibodies neutralize
388 SARS-CoV-2 by targeting the spike receptor binding domain. *Nature Communications.* 2020;11(1):4528.
- 389 23. Tan CW, Chia WN, Qin X, Liu P, Chen MIC, Tiu C, et al. A SARS-CoV-2 surrogate virus
390 neutralization test based on antibody-mediated blockage of ACE2–spike protein–protein interaction.
391 *Nature Biotechnology.* 2020;38(9):1073-8.
- 392 24. Braun J, Loyal L, Frentsch M, Wendisch D, Georg P, Kurth F, et al. SARS-CoV-2-reactive T cells
393 in healthy donors and patients with COVID-19. *Nature.* 2020.
- 394 25. Staines HM, Kirwan DE, Clark DJ, Adams ER, Augustin Y, Byrne RL, et al. Dynamics of IgG
395 seroconversion and pathophysiology of COVID-19 infections. *medRxiv.* 2020:2020.06.07.20124636.
- 396 26. Tuaille E, Bolloré K, Pisoni A, Debiec S, Renault C, Marie S, et al. Detection of SARS-CoV-2
397 antibodies using commercial assays and seroconversion patterns in hospitalized patients. *The Journal*
398 *of infection.* 2020;81(2):e39-e45.
- 399 27. Gallais F, Velay A, Wendling M-J, Nazon C, Partisani M, Sibilia J, et al. Intrafamilial Exposure to
400 SARS-CoV-2 Induces Cellular Immune Response without Seroconversion. *medRxiv.*
401 2020:2020.06.21.20132449.
- 402 28. Okba NM, Müller MA, Li W, Wang C, GeurtsvanKessel CH, Corman VM, et al. Severe acute
403 respiratory syndrome coronavirus 2– specific antibody responses in coronavirus disease patients.
404 *Emerging infectious diseases.* 2020;26(7):1478-88.
- 405 29. Plebani M, Padoan A, Negrini D, Carpinteri B, Sciacovelli L. Diagnostic performances and
406 thresholds: the key to harmonization in serological SARS-CoV-2 assays? *medRxiv.*
407 2020:2020.05.22.20106328.
- 408 30. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al. Seroprevalence of
409 anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study.
410 *Lancet.* 2020.
411

Sample composition True pos. / true neg.	Test	Manuf.'s cut-off	Optimised cut-off [CI]	Sensitivity [%] (Manuf.'s / Optim. cut-off) **	Specificity [%] (Manuf.'s / Optim. cut- off) **	Overall accuracy [%] (Manuf.'s / Optim. cut- off) **
193 / 1073	EI-S1-IgA	1.100	1.085 [0.855; 1.705]	64.77 / 64.77	93.29 / 92.64	88.94 / 88.39
193 / 1073	EI-S1-IgG	1.100	1.015 [0.850; 1.395]	77.20 / 79.79	98.04 / 97.76	94.87 / 95.02
193 / 1073	Ro-N-Ig	1.000	0.422 [0.295; 0.527]	85.49 / 88.60	99.81 / 99.72	97.63 / 98.03
107 / 106	NT	-	10.000*	- / 73.83	- / 100.00	- / 86.85
108 / 106	GS-cPass	20.000	20.538 [13.768; 24.241]	96.30 / 96.30	100.00 / 99.06	98.13 / 97.66
108 / 111	VC-N-IgG	100.000	18.500 [13.500; 23.000]	39.81 / 93.52	99.10 / 91.89	69.86 / 92.69
108 / 111	VC-S1-IgG	100.000	10.000 [10.000; 10.000]	65.74 / 95.37	100.00 / 100.00	83.11 / 97.72
108 / 111	VC-S2-IgG	100.000	10.000 [10.000; 10.000]	17.59 / 63.89	100.00 / 99.10	59.36 / 81.74
78 / 106	MG-N	1.000	1.000 [1.000; 1.600]	94.87 / 94.87	98.11 / 98.11	96.74 / 96.74
78 / 106	MG-RBD	1.000	1.000 [1.000; 1.000]	94.87 / 94.87	100.00 / 100.00	97.83 / 97.83
78 / 106	MG-S1	1.000	1.000 [1.000; 1.000]	96.15 / 96.15	100.00 / 100.00	98.37 / 98.37
193 / 1073	Random Forest	-	-	88.60	99.81	98.10
193 / 1073	Support Vector Machine	-	-	84.46	99.91	97.47

Table 1: Manufacturer's and optimised cut-off, sensitivity, specificity and accuracy

Evaluation of diagnostic accuracy of primary tests was conducted with samples from true-positives (n=193) and true-negatives (n=1073); subsequently, optimised cut-offs were applied to the KoCo19-cohort (see Methods).

* For NT, dilutions starting at 1:10 were used; see Methods.

** The random forest and the support vector machine combine all three primary tests, the accuracy measures thus do not relate to specific cut-offs.

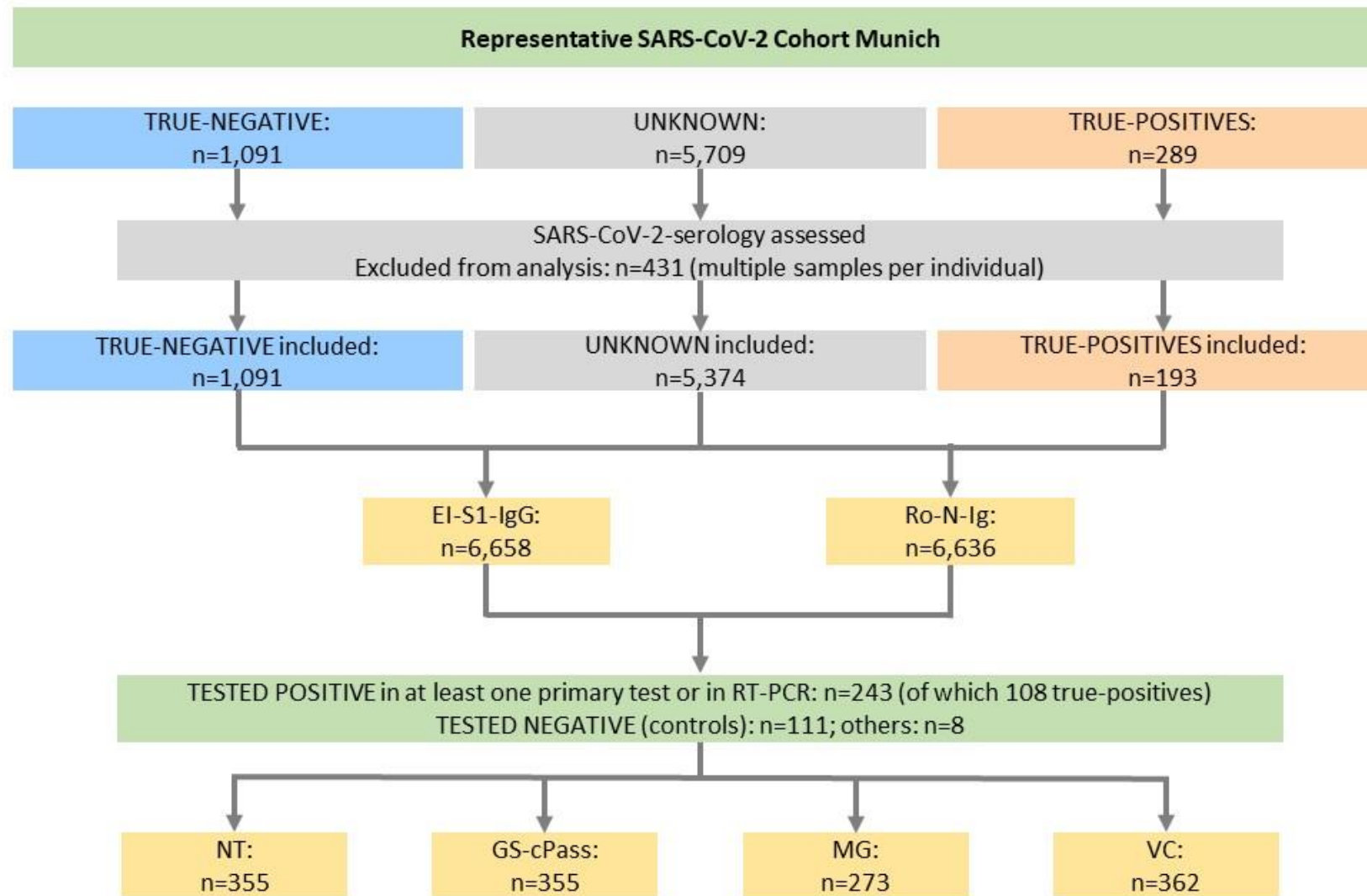
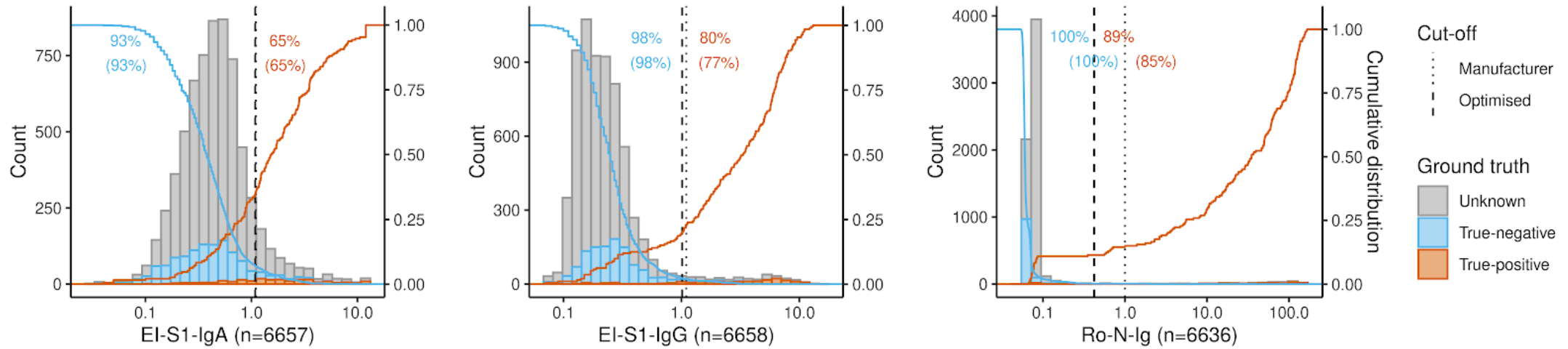


Figure 1: Cohort composition and characterisation of the study participants.

True-positives were defined as subjects with a positive RT-PCR; true-negatives as blood donors from the pre-COVID-19 era. In addition, we included individuals recruited into the Representative COVID-19 Cohort Munich (KoCo19), 100 of which were considered as true-negatives. For each participant, a single sample was used for statistical analysis. Among individuals with longitudinal measurements, the blood sample with the most complete dataset was retained. For similar datasets, the earliest measurement was considered. For operational replicates, the latest measurement was used.

A Distributions of primary tests



B Reliability of the primary tests on multiple measurements

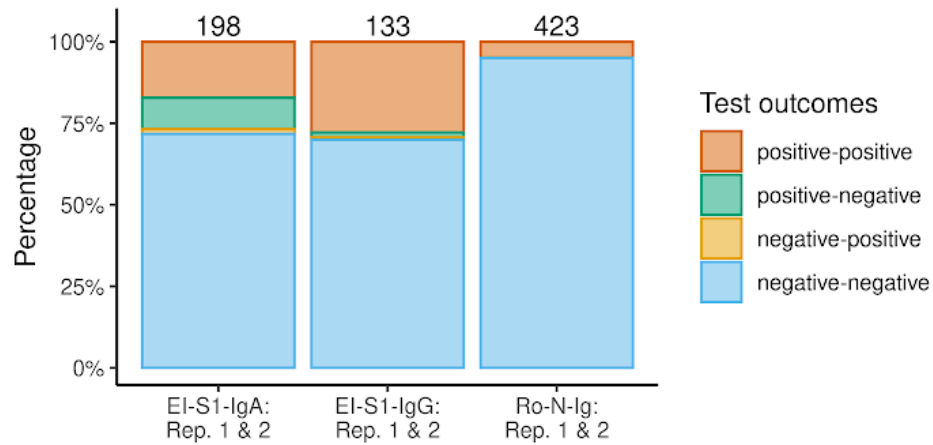


Figure 2: Performance of primary tests .

Results of primary tests for true-negatives (*blue*), true-positives (*orange*), and individuals with unknown SARS-CoV-2 status (*grey*).

(A) Performance of EI-S1-IgA (*left*), EI-S1-IgG (*centre*), and Ro-N-Ig (*right*). Dotted lines mark the manufacturer's cut-off value (between indeterminate and positive for EI, and between negative and positive in Ro) and dashed lines mark the optimised cut-off value as determined in this study. Orange and blue solid lines represent the percentage of test results for true-positives and true-negatives above (*blue*) or below (*orange*) the value on the x-axes, respectively. Orange and blue numbers give the percentage of true-positives and true-negatives, which were correctly detected by the test (without brackets: manufacturers' cut-offs; within brackets: optimised cut-offs). Raw values for EI-S1-IgA show a slightly asymmetric but unimodal distribution for the overall population. The EI-S1-IgG raw values show a bimodal distribution. Ro-N-Ig raw values/COI demonstrate a narrow distribution with the bulk of values in the range COI 0.1 and below, whereas a clearly separated second population above COI 10 was observed. For EI-S1-IgG and Ro-N-Ig, the cut-offs separate the blue and orange subpopulations more reliably than for EI-S1-IgA.

(B) Consistency of the primary tests on operational replicates (manufacturer's cut-offs). The changes from positive to negative status in EI-S1-IgA was most likely caused by a batch effect. Reliability of the other primary tests was higher.

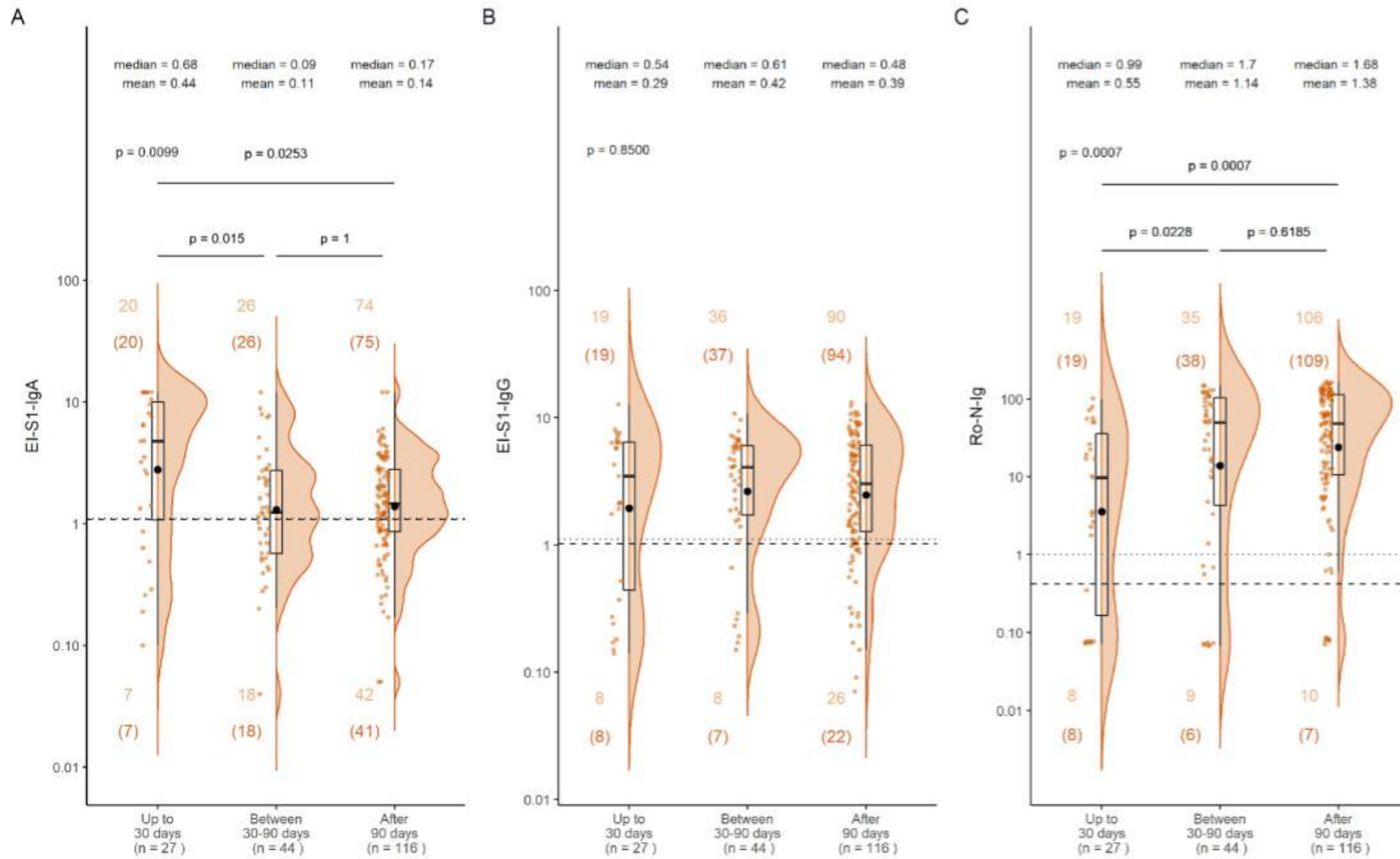


Figure 3: Time-dependence in primary tests for RT-PCR true-positives.

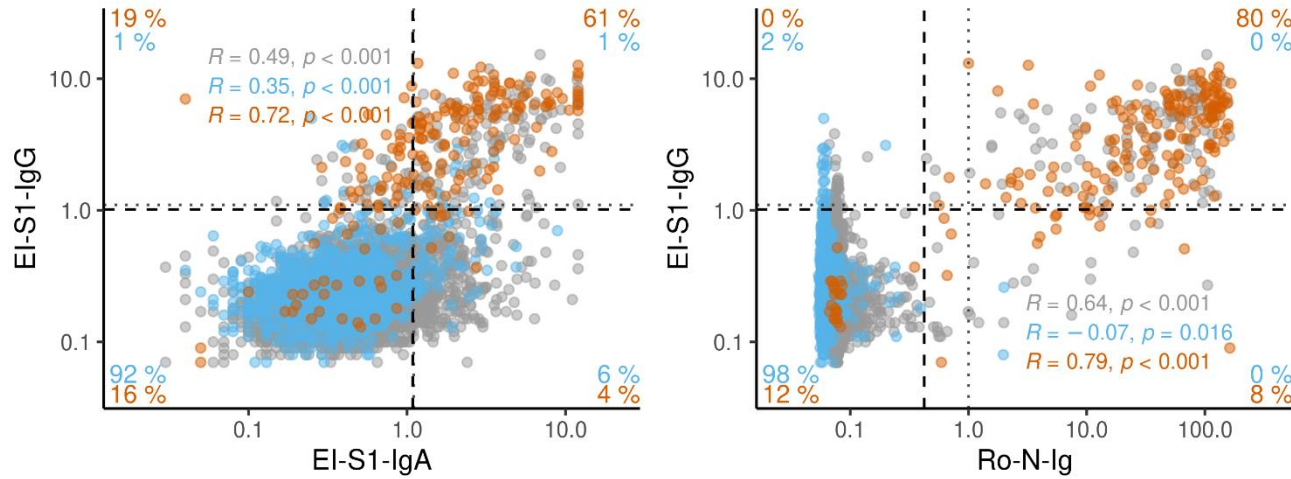
Titre values of the 187 true-positives with available data on time between RT-PCR and blood sampling for (A) EI-S1-IgA, (B) EI-S1-IgG, and (C) Ro-N-Ig. The read-outs were categorized according to the time after positive RT-PCR (<30 days, 30-90 days, and >90 days). Plots show the individual read-out (orange dots), a density estimate (orange area), the 25-,50- and 75-percentiles (black box), and the mean (black dot). Counts n refer to the number of observations above/below manufacturer's and optimised cut-off for each of the temporal groups (without brackets: manufacturers' cut-offs; within brackets: optimised cut-offs). Pairwise differences are considered only after adjusting for multiple testing.

(A) EI-S1-IgA values were highest on average in the category of <30 days and significantly decline thereafter.

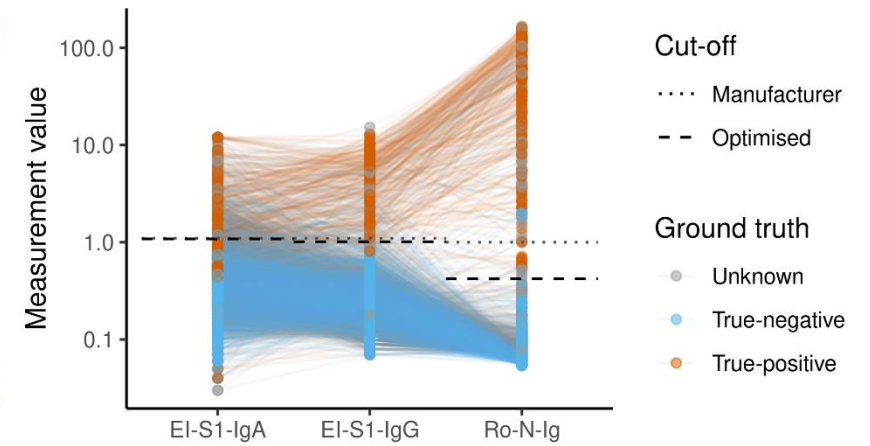
(B) EI-S1-IgG values were widespread in the group <30 days, and on average do not decline significantly >90 days.

(C) Similarly, Ro-N-Ig does not decline during the study period; however, raw values increase over time.

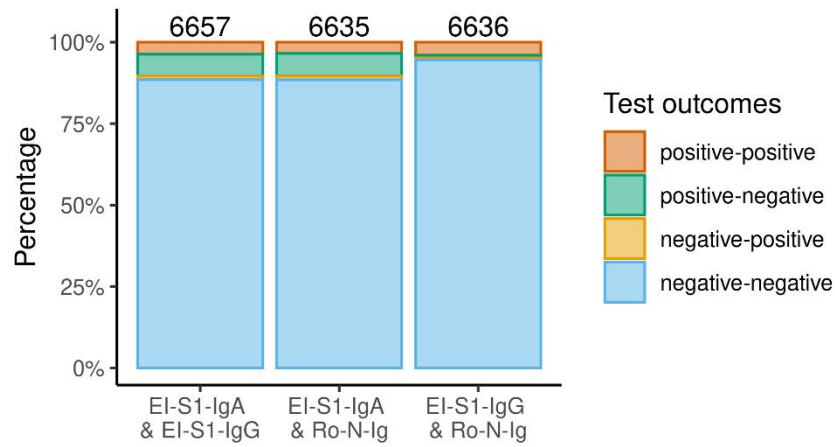
A Performance comparison of primary tests



B Parallel coordinate plot of primary tests



C Concordance of primary tests



D Concordance of primary tests and ground truth

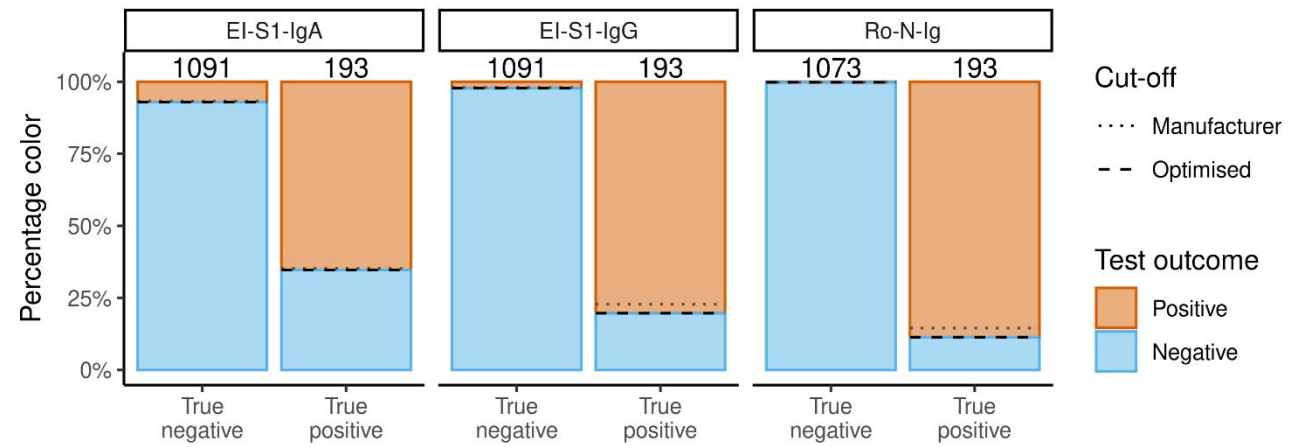


Figure 4: Comparison of primary tests.

Results of primary tests compared to ground truth for true-negatives (*blue*), true-positives (*orange*), and individuals with unknown SARS-CoV-2 status (*grey*). The dotted lines represent the manufacturer's cut-offs, the dashed lines the optimised cut-offs defined within this study.

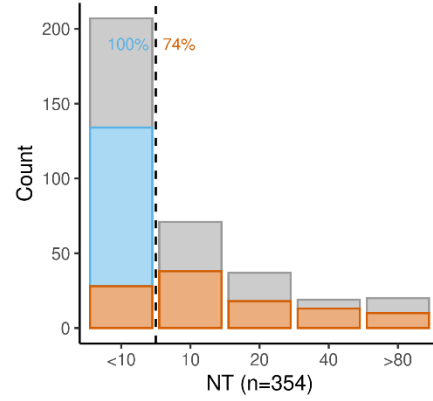
(A) Pairwise scatter plots for primary tests: EI-S1-IgA vs. EI-S1-IgG (left; n=6657), and Ro-N-Ig vs. EI-S1-IgG (right; n=6636). Percentages in orange indicate fractions of true-positives in the respective quadrant with respect to all true-positives; blue for true-negatives. Percentages were calculated using the optimised cut-off. EI-S1-IgA classified 65% of true-positives correctly and 7% of the true-negatives as positive. EI-S1-IgG classified 80% of the true-positives correctly and 2% of true-negatives incorrectly. 61% of true-positives were identified by both tests unanimously. Comparing Ro-N-Ig and EI-S1-IgG, 80% of true-positives were concordantly classified as positive by both tests, while 98% of the true-negatives were classified as negative. Correlation of true-positives between Ro-N-Ig and EI-S1-IgG was $R=0.79$. The fraction of true-negatives falsely classified as positives in Ro-N-Ig was below 1%.

(B) Parallel coordinate plot of the primary tests. Magnitude of titres of individual samples across the different assay used were similar. The true-negative samples presented as a group in Ro-N-Ig. Only few samples display as high EI-S1-IgG and low Ro-N-Ig; on the contrary, most samples with low EI-S1-IgA titres, have higher titres in EI-S1-IgG and even higher Ro-N-Ig, presenting separately from the negative population.

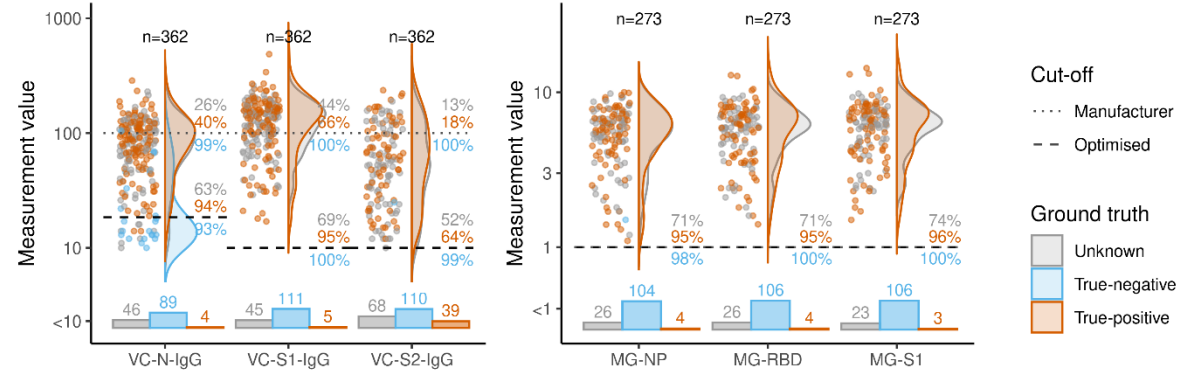
(C) Concordance of primary tests (based on manufacturer's cut-offs). Numbers of paired samples are indicated above the bars. In the two columns on the left, the unspecific reactivity of EI-S1-IgA is represented in green; the concordance of EI-S1-IgG and Ro-N-Ig was pronounced, with 98 (1.5%) discordant results (right bar).

(D) Test results of the primary tests (based on optimised cut-offs). Highest concordance was seen in Ro-N-Ig, lowest in EI-S1-IgA. The proportion of false-negatives in the true-positive cohort was 34.7%, 19.7%, and 11.4%, respectively. Applying optimised (dashed line) and manufacturer's (dotted line) cut-off results in a reduction of false-negatives in EI-S1-IgG and Ro-N-Ig for the optimised cut-off.

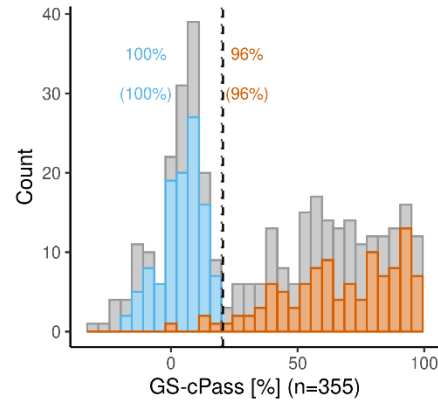
A Distribution of NT



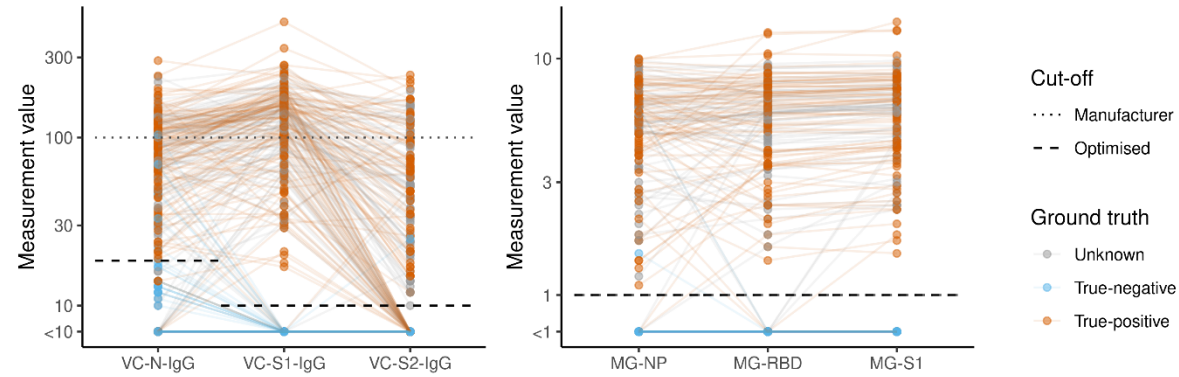
B Distribution of VC-IgG and MG



C Distribution of GS-cPass



D Parallel coordinate plots for VC-IgG and MG



E Concordance of confirmatory tests and ground truth

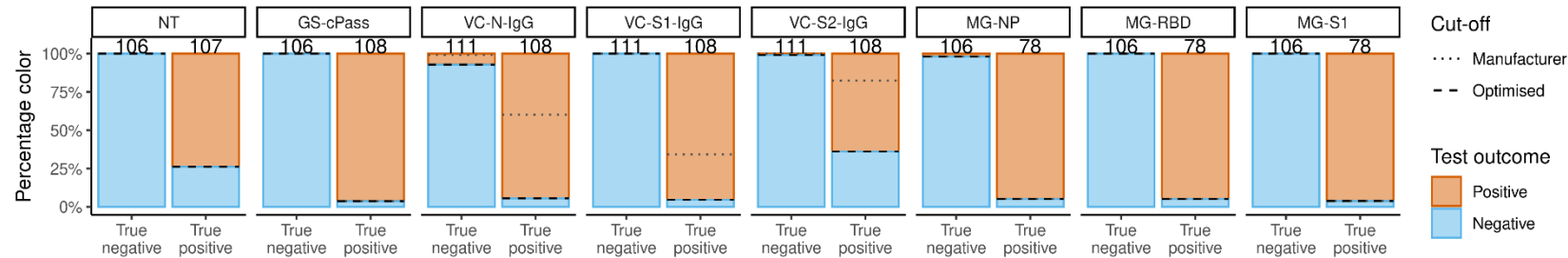


Figure 5: Confirmatory tests

Results of confirmatory tests compared to ground truth for true-negatives (*blue*), true-positives (*orange*), and individuals with unknown SARS-CoV-2 status (*grey*). Black dotted and dashed lines represent the manufacturer's and the optimised cut-offs, respectively. Orange/blue numbers indicate percentages of true-positives/-negatives correctly detected by the test.

(A,C) Distribution of results of NT and GS-cPass. At 1:10 dilution, NT endpoint was categorical with a specificity of 100% in this cohort; sensitivity was 73.8%. GC-cPass manufacturer's cut-off yielded a specificity of 99.1% and a sensitivity of 96%. Adjustments of the cut-off did not improve the performance of the test systems (shown in parentheses).

(B) Distribution of results of the VC-array (left) and the MG-line blot (right). Bar charts below violin plots represent information for the categorical part of the tests. Grey numbers give the percentages of positive samples with unknown SARS-CoV-2 as determined by the manufacturer's and optimised cut-offs. Percentages were calculated over the total number of samples of unknown SARS-CoV-2 status with available test results. In the VC-array, S1-IgG / N-IgG performed best as confirmatory tests with 95.4% / 93.5% sensitivity, and 100% / 91.9% specificity, respectively, applying optimised cut-offs. With manufacturer's cut-offs, sensitivity was 65.7% / 39.8%, respectively. With the optimised cut-off, VC-S2-IgG had a sensitivity of 63.9% and a specificity of 99.1%. Performance of VC-S2-IgA and VC-S2-IgM are presented for reference in supplemental figure 4 (appendix p.17), demonstrating a potential use in diagnostics settings. MG-N, -S1 and -RBD had sensitivities of 94.9% or higher, and specificities of 98.1% or higher, no optimisation of cut-offs could not be optimised here.

(D) Parallel coordinate plot of VC-Array and MG. Many subjects do not show S2 reaction although strong reactivity against N and S1 can be detected. Comparing individual subjects in NP, RBD and S1 reaction indicates that several individuals developed strong N and S1 reactions whereas RBD was not bound. General concordance between S1 and RBD was, however, observed.

(E) Description of concordance between results comparing each confirmatory test with the ground truth. The colour coding is based on the optimised cut-off.

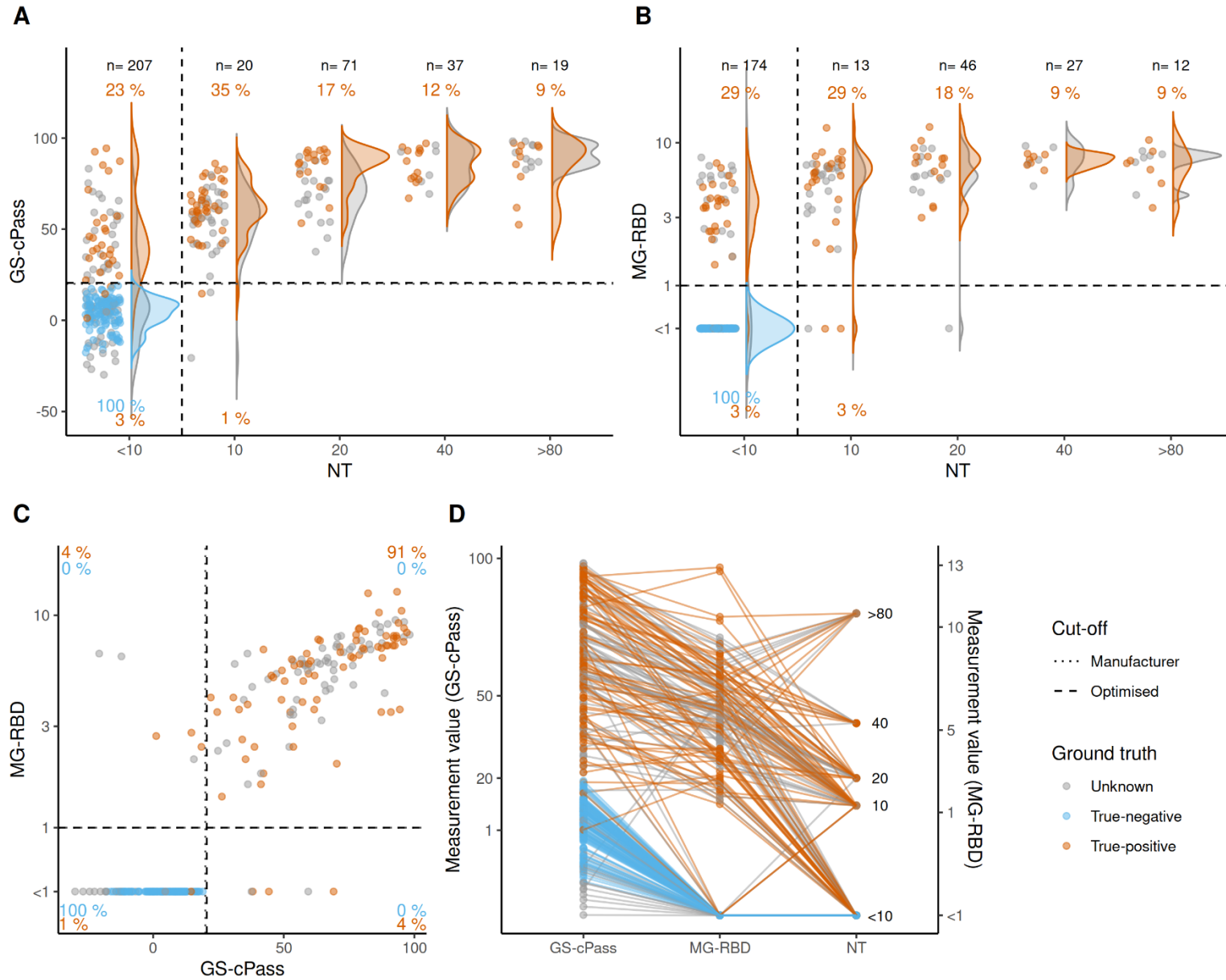


Figure 6: Comparison of confirmatory tests

Comparison of confirmatory tests for true-negatives (*blue*), true-positives (*orange*), and individuals with unknown SARS-CoV-2 status (*grey*).

(A) Association between the categorical endpoint of NT and the continuous results of GS-cPass. The test results were positively related; in cases of identical classification agreement with ground truth was frequent. However, more than 20% of true-positives were missed by NT (1:10 dilution; n=354).

(B) Association between the categorical endpoint of NT and the continuous results of MG-RBD. In true-positives with low direct neutralization capacity (NT <10), distinction between negative and positive populations was observed with GS-cPass (n=272), highlighting the limited sensitivity of NT (1:10 dilution).

(C) Association between GS-cPass and MG-RBD presented with discordant results in 8% (of true-positives) at intermediate test read-outs. The distribution in higher titre ranges were narrow (n=272).

(D) Parallel coordinate plot of the two neutralization surrogate tests and NT. Samples with high/low raw values demonstrated similarly in all three tests, as lines present horizontally. However, a subset of individuals with relatively high GS-cPass values and positive outcome in NT were below the cut-off in the MG-RBD assay.

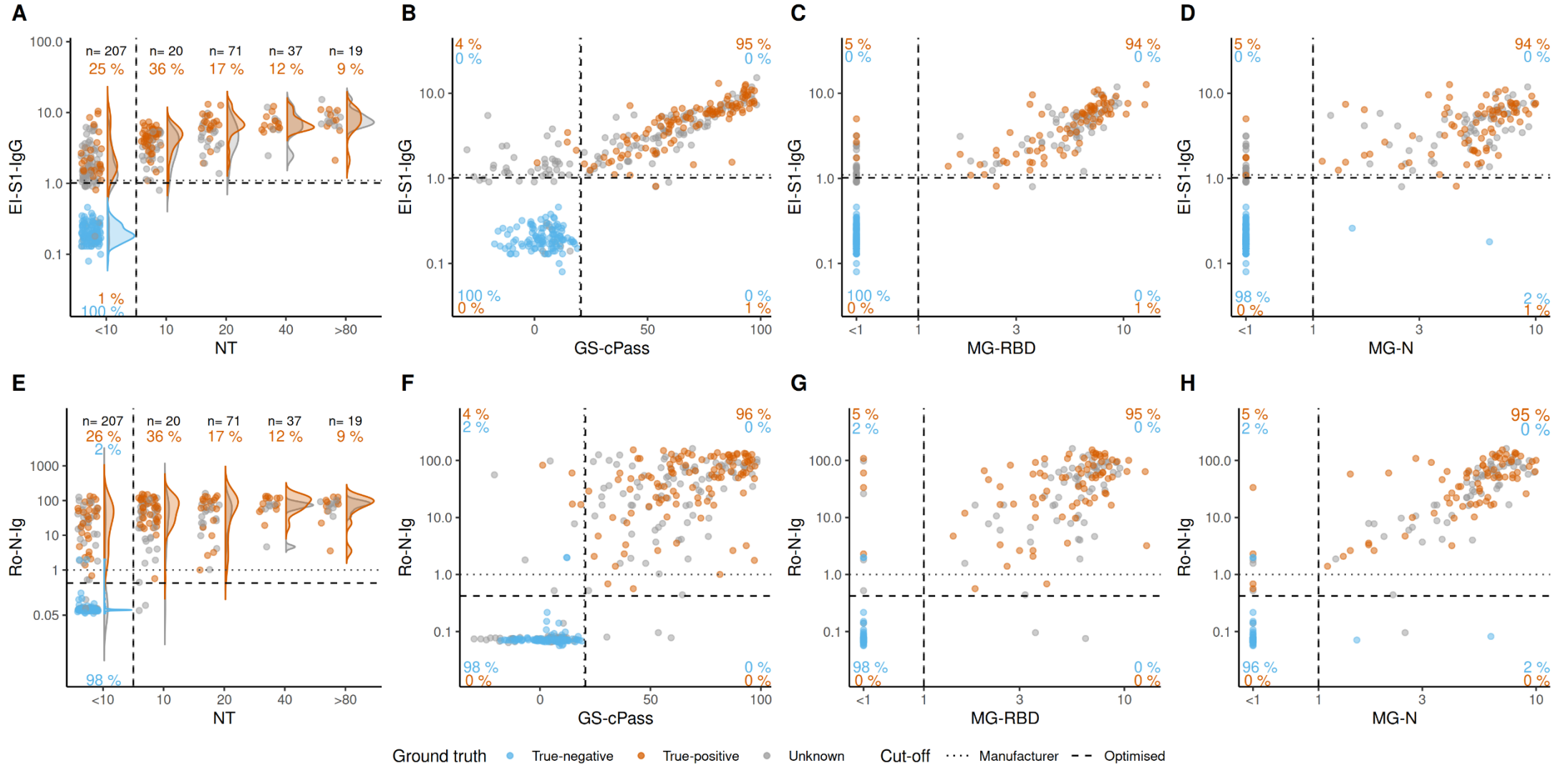


Figure 7: Comparison of primary tests (EI-S1-IgG, Ro-N-Ig) with confirmatory tests (NT, GS-cPass MG-RBD, MG-N).

Comparison of primary with confirmatory tests for true-negatives (*blue*), true-positives (*orange*), and individuals with unknown SARS-CoV-2 status (*grey*).

(A) EI-S1-IgG and NT show positive correlation (n=354), although 1% of the true-positives were missed by NT (1:10 dilution).

(B, C) Correlation of EI-S1-IgG with GS-cPass (n=361) and MG-RBS (n=272) presented as dose dependent for the double-positive values. Concordance for the true-positive subset was >94%.

(D) Association of EI-S1-IgG to MG-N (n=355) presented dose-dependent, but not as markedly.

(E-H) Association between Ro-N-Ig and the confirmatory tests (n=362, n=273, n= 354, n=354) did not present as dose-dependent, compared to EI-S1-IgG. Concordance between the tests was >95%; except NT, which presented similar patterns to EI-S1-IgG.

We observed a population in the upper left quadrant, (B, C, D), clearly negative in the confirmatory tests.