The sensitivity improved two-test algorithm "SIT^2": a universal optimization strategy for SARS-CoV-2 serology

Thomas Perkmann MD*[1], Thomas Koller MD*[1], Nicole Perkmann-Nagele MD[1], Maria Oszvar-Kozma MSc[1], David W Eyre DPhil[2], Philippa Matthews DPhil[3], Abbie Bown MD[4], Nicole Stoesser DPhil[3], Marie-Kathrin Breyer MD PhD[5], Robab Breyer-Kohansal MD[5], Otto C Burghuber MD[6], Sylvia Hartl MD[5,6], Daniel Aletaha MD[7], Daniela Sieghart PhD[7], Peter Quehenberger MD[1], Rodrig Marculescu MD[1], Patrick Mucher MSc[1], Astrid Radakovics MSc[2], Miriam Klausberger PhD[8], Mark Duerkop PhD[9], Barbara Holzer PhD[10], Boris Hartmann PhD[10], Robert Strassl MD[1], Gerda Leitner MD[11], Florian Grebien PhD[12], Wilhelm Gerner PhD[13,14,15], Reingard Grabherr PhD[8], Oswald F Wagner MD[1], Christoph J Binder MD PhD[1], and Helmuth Haslacher MD PhD[1*]

[1]Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria
[2]Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, UK
[3]Nuffield Department of Medicine, University of Oxford, Oxford, UK
[4]Public Health England (PHE) Porton Down, Salisbury, UK
[5]Department of Respiratory and Critical Care Medicine and Ludwig Boltzmann Institute for Lung Health, Otto Wagner Hospital, Vienna, Austria
[6]Sigmund Freud University, Medical School and Ludwig Boltzmann Institute for Lung Health, Vienna, Austria
[7]Divison of Rheumatology, Department of Medicine III, Medical University of Vienna, Vienna, Austria
[8]Institute of Molecular Biotechnology, Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU) Vienna, Vienna, Austria
[9]Institute of Bioprocess Science and Engineering, Department of Biotechnology, University of Natural Resources and Life Sciences (BOKU) Vienna, Vienna, Austria
[10]Institute for Veterinary Disease Control, Austrian Agency for Health and Food Safety (AGES), Moedling, Austria
[11]Department of Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Vienna, Austria
[12]Institute for Medical Biochemistry, University of Veterinary Medicine, Vienna, Austria
[13]Institute of Immunology, University of Veterinary Medicine Vienna, Austria
[14]Christian Doppler Laboratory for an Optimized Prediction of Vaccination Success in Pigs, University of Veterinary Medicine, Vienna, Austria
[15]The Pirbright Institute, Pirbright, United Kingdom (current)

both authors contributed equally

Address all correspondence to:

Helmuth HASLACHER, LECT MD PhD MSc BSc BA

Medical University of Vienna

Department of Laboratory Medicine

Waehringer Guertel 18-20

1090 Vienna

Austria

Tel: +43 1 40400 53190

Fax: +43 1 40495 15547

Mail: helmuth.haslacher@meduniwien.ac.at

Background

Reliable antibody tests are an essential tool to identify individuals who have developed an adaptive immune response to SARS-CoV-2. However, attempts to maximize the specificity of SARS-CoV-2 antibody tests have come at the cost of sensitivity, exacerbating the total test error with increasing seroprevalence. Here, we present a novel method to maximize specificity while maintaining or even increasing sensitivity: the "Sensitivity Improved Two-Test" or "SIT²" algorithm.

Methods

SIT² involves confirmatory re-testing of samples with results falling in a predefined retesting-zone of an initial screening test, with adjusted cut-offs to increase sensitivity. We verified and compared the performance of SIT² to single tests and orthogonal testing (OTA) in an Austrian cohort (1,117 negative, 64 post-COVID positive samples) and validated the algorithm in an independent British cohort (976 negatives, 536 positives).

Results

The specificity of SIT² was superior to single tests and non-inferior to OTA. The sensitivity was maintained or even improved using SIT² when compared to single tests or OTA. SIT² allowed correct identification of infected individuals even when a live virus neutralization assay could not detect antibodies. Compared to single testing or OTA, SIT² significantly reduced total test errors to 0·46% (0·24-0·65) or 1·60% (0·94-2·38) at both 5% or 20% seroprevalence.

Conclusion

SIT² proved to be the best diagnostic choice at both 5% and 20% seroprevalence in all tested scenarios. It is an easy algorithm to apply to different available SARS-CoV-2

antibody testing systems and can potentially be helpful for the serology of other infectious diseases.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has led to a worldwide pandemic confronting us with unprecedented epidemiological, therapeutic, but not at last diagnostic challenges.

While case identification focuses primarily on polymerase chain reaction (PCR) or antigen testing, detecting antibodies to SARS-CoV-2 is essential to identify individuals who develop a SARS-CoV-2-specific adaptive immune response following infection or vaccination. So, reliable determination of SARS-CoV-2 antibodies can provide critical information for healthcare decision-making (1). Accurate measurement of antibody levels is paramount in estimating seroprevalences because subsequent attempts to compensate for antibody test errors can result in highly variable estimates of the actual seroprevalence rates(2). Moreover, antibody positivity has also been shown to correlate with protective immunity(3, 4) and several quantitative assays have been used to determine vaccine-induced antibody levels(5, 6).

Current test systems measure antibodies against either the SARS-CoV-2 nucleocapsid (NC) or spike (S) protein or fragments thereof(7). For correct serological results, both high specificity and sensitivity are crucial. The suboptimal specificities of some of the early SARS-CoV-2 antibody tests were problematic in the context of the low seroprevalence at the onset of the pandemic. This has changed rapidly, and currently available SARS-CoV-2 antibody tests are amongst the most specific serological tests available. In addition, despite marked regional differences, global seroprevalence has increased significantly(8). Unfortunately, the claim by many test manufacturers that assay sensitivities are close to 100% has proven to be inaccurate and too optimistic in

practice(7, 9). Two major possible reasons underlying this are that: i) lower antibody levels in non-hospitalized mild or asymptomatic cases sometimes fall below manufacturers' cut-offs, which are set to optimize specificity to limit false-positive results(10, 11), and ii) the decline in antibody titers over time towards the positivity threshold increases the demands on assay sensitivity(12, 13). The latter phenomenon is dependent on the test system used(14, 15) and can be a considerable problem, as shown by an epidemiological study from Brazil(2).

Consequently, maximizing both the sensitivity and specificity of SARS-CoV-2 serology is essential. However, there has been no attempt to optimize both parameters simultaneously through a systematic approach to our knowledge. Generally, there is a trade-off between sensitivity and specificity for a given test depending on the selected threshold for positivity; increasing specificity to near 100% comes at the expense of sensitivity, and vice versa. Attempts to increase sensitivity by lowering thresholds for positivity below the recommended values set by manufacturers lead to unacceptably low specificities. Orthogonal testing algorithms (OTA), as recommended by the Center for Disease Control (CDC) for low seroprevalence settings, increase specificity at the expense of sensitivity(16). Although this recommendation was justified at the beginning of the pandemic, it is no longer appropriate given the significant increase in seroprevalence in most areas of the world and the known sensitivity problem of SARS-CoV-2 antibody tests and contributes to the rise in the overall error rate.

In this paper, we present a new, simple, and widely applicable method, the Sensitivity-Improved Two-Test (SIT^2) algorithm, which, for the first time, allows the maximization of specificity while maintaining or even increasing sensitivity.

Methods

Study design and cohorts

Sera used in this non-blinded prospective cross-sectional study were either residual clinical specimens or samples stored in the MedUni Wien Biobank (n=1,181), a facility specialized in the preservation and storage of human biomaterial, which operates within a certified quality management system (ISO 9001:2015)(17).

For derivation of the SIT^2 algorithm, sample sets from individuals known to be negative and positive were established for testing. As previously described(11), samples collected before 01.01.2020 (i.e., assumed SARS-CoV-2 negative) were used as a specificity cohort (n=1117): a cross-section of the Viennese population (the LEAD study)(18), preselected for samples collected between November and April to enrich for seasonal infections (n=494); a collection of healthy voluntary donors (n= 265); a disease-specific collection of samples from patients with rheumatic diseases (n=358); (see also Tables S1 and S2).

Of the SARS-CoV-2 positive cohort (n=64 samples from n=64 individuals), five individuals were asymptomatic, 42 had mild-moderate symptoms, four reported severe symptoms, and 13 were admitted to the Intensive Care Unit (ICU). The timing of symptom onset was determined by a questionnaire for convalescent donors and by reviewing individual health records for patients. For asymptomatic donors (n=5), SARS-CoV-2 RT-PCR confirmation time was used instead (for more details, see Tables S1 and S3). All included participants gave written informed consent to donate their samples for research purposes. The overall evaluation plan conformed with the Declaration of

Helsinki as well as relevant regulatory requirements. It was reviewed and approved by the ethics committee of the Medical University of Vienna (1424/2020).

For validation of the SIT² algorithm, we used data from an independent United Kingdom cohort(19), including 1,512 serum/plasma samples (536 PCR confirmed SARS-CoV-2 positive cases; 976 negative cases, collected earlier than 2017).

Antibody testing

For the derivation analyses, SARS-CoV-2 antibodies were either measured according to the manufacturers' instructions on three different commercially available automated platforms (Roche Elecsys® SARS-CoV-2 [nuclecapsid total antibody assay, further referred to as Roche NC], Abbott SARS-CoV-2-IgG assay [nucleocapsid IgG assay, Abbott NC], DiaSorin LIASION® SARS-CoV-2 S1/S2 assay [S1/S2 combination antigen IgG assay, DiaSorin S1/S2]) or using 96-well enzyme-linked immunosorbent assays (ELISAs) (Technoclone Technozym® RBD and Technozym® NP) yielding quantitative results(20) (for details see Supplement, Supplemental Methods). The antibody assays used in the validation cohort were Abbott NC, DiaSorin S1/S2, Roche NC, Siemens RBD total antibody, and a novel 384-well trimeric spike protein ELISA (Oxford Immunoassay)(19), resulting in 20 evaluable combinations. All samples from the Austrian SARS-CoV-2-positive cohort also underwent live virus neutralization testing (VNT), and neutralization titers (NT) were calculated, as is described in detail in the Supplemental Methods.

Sensitivity improved two-test method (SIT²)

Our newly developed sensitivity improved two-test (SIT²) method consists of the following key components: i) sensitivity improvement by cut-off modification and ii) specificity rescue by a second, confirmatory test (Fig. 1A).

For the first component of the SIT² algorithm, positivity thresholds were optimized for sensitivity according to the first published alternative thresholds for the respective assays(21-23). Additionally, a high cut-off, above which a result can be reliably regarded as true positive without the need for further confirmation, was defined. These levels were based on in-house observations(11). The lowering of positivity thresholds improves sensitivity; the high cut-off prevents unnecessary re-testing of clearly positive samples. Moreover, the high cut-off avoids possible erroneous exclusion by the confirmatory test. The newly defined interval between the reduced threshold for positivity and the high cut-off is the re-testing zone (Fig. 1A). The initial antibody test (screening test) is then followed by a confirmatory test, whereby positive samples from the re-testing zone of the screening test are re-tested. Also, for the confirmatory test, sensitivity-adapted assay thresholds are needed (Figs.1A, 1B). As false-positive samples are usually only positive in one test system (Fig. S1), false positives can be identified, and specificity markedly restored with minimal additional testing as most samples do not fall within the re-testing zone(11, 24). A flowchart of the testing strategy and the applied cut-off levels and their associated quality criteria are presented in Figs. 1B, 1C.

Test strategy evaluation

On the derivation cohort, we compared the overall performance of the following SARS-CoV-2 antibody testing strategies: i) testing using single assays; ii) simple lowering of

thresholds; iii) classical orthogonal testing (OTA), and iv) our newly developed SIT^2 algorithm at assumed seroprevalences of 5% and 20%. As part of the derivation, we then compared the performance of OTAs and SIT^2 against the results of a virus neutralization assay. On the validation cohort, we then compared the performance of OTAs and SIT^2. Finally, we used data from this cohort to evaluate the performance of SIT^2 versus single tests at seroprevalences of 5%, 10%, 20%, and 50% if the Abbott and DiaSorin assays (i.e., assays with varying degrees of discrepancies in sensitivity and specificity) were used.

Statistical analysis

Unless otherwise indicated, categorical data are given as counts (percentages), and continuous data are presented as median (interquartile range). Total test errors were compared by Mann-Whitney tests or, in case they were paired, by Wilcoxon tests. 95% confidence intervals (CI) for sensitivities and specificities were calculated according to Wilson, 95% CI for predictive values were computed according to Mercaldo-Wald unless otherwise indicated. Sensitivities and specificities were compared using z-scores. P values <0.05 were considered statistically significant. All calculations were performed using Analyse-it 5.66 (Analyse-it Software, Leeds, UK) and MedCalc 19.6 (MedCalc bvba, Ostend, Belgium). Graphs were drawn using Microsoft Visio (Armonk, USA) and GraphPad Prism 7.0 (La Jolla, USA).

Results

Based on the derivation cohort of 1,117 pre-pandemic sera and 64 sera from convalescent COVID-19 patients (80% non-hospitalized, 20% hospitalized), the Roche NC, Abbott NC, and DiaSorin S1/S2 antibody assays gave rise to 545, 780, and 860 false-negative results per 100,000 tests, and 285, 760 and 1,710 false-positive results per 100,000 tests respectively, assuming a seroprevalence of 5% (Fig. 2A, left panel).

Effects of threshold lowering on Sensitivity and Specificity

Lowering the positivity thresholds for the Roche NC, Abbott NC, and Diasorin S1/S2 to 17%, 38% and 75% of the manufacturers cut-offs increased the sensitivity significantly and reduced false-negative results to 80, 155, and 545 per 100,000 tests, respectively, but substantially increased false-positive results to 1,520, 2,280 and 2,660 per 100,000 tests respectively (all at an assumed seroprevalence of 5%; Table S4, Fig. 2A, left panel).

Classical OTA compared to SIT^2

Subsequently, we evaluated 12 OTA combinations using the fully automated SARS-CoV-2 antibody tests from Roche NC, Abbott NC, and DiaSorin S1/S2 as screening tests, each combined with one of the other fully automated assays or a commercially available NC or RBD-specific ELISA as a confirmation test. Combining these tests as classical OTAs significantly increased specificity and reduced false positives to 0 (0-95) per 100,000 tests. However, the rate of false negatives was 1,095 (955-1,230) per 100,000 tests, and therefore considerably higher than for single testing strategies. In contrast, the SIT^2 algorithm minimized false positives to 0 (0-166·25) per 100,000 tests

while also reducing false negatives to 390 (235-607·5) per 100,000 tests (Fig. 2A left panel; Table S5).

At an assumed seroprevalence of 20%, the consequences of reduced test sensitivities led to an even more substantial rise in false-negative results for all testing strategies (Fig. 2A, right panel). Both SIT^2 and OTA reduced false-positive results substantially compared to single testing strategies (SIT^2: 0 [0-140] and OTA: 0 [0-80] false-positives per 100,000 tests). However, SIT^2 achieved a substantially lower rate of false-negative results when compared to OTA (1,560 [940-2,420] vs 4,380 [3,820-4,920] per 100,000 tests). The reduction in false-positive results using SIT^2 compared to single tests with the manufacturers' recommended thresholds (Table S6) was accompanied by a lower average false-negative rate for SIT2 than for single tests (Fig. 2A, right panel). Even at 20% seroprevalence, single tests resulted in a non-negligible number of false positives (240, 640, and 1,440 per 100,000 tests).

Reduction of total error rates by the Sensitivity-Improved Two-Test

Of all the methods assessed, SIT^2 reached the lowest total error rates per 100,000 tests under both 5% and 20% assumed seroprevalence (455 [235-685] and 1,600 [940-2,490] per 100,000 tests) (Fig. 2B). At a seroprevalence of 5 %, OTA on average performed better than individual tests, and the total error rates of the single tests were higher for the Abbott NC and DiaSorin S1/S2 assay (OTA 1,095 [955-1,325] vs. 830 [Roche NC], 1,540 [Abbott NC] and 2,570 [DiaSorin S1/S2] per 100,000 tests). But with a seroprevalence of 20 %, performance of OTAs, worsened compared to single tests (OTA 4,380 [3,820-5,000] vs 1,600 [Roche], 2,540 [Abbott] and 4,420 [DiaSorin] per 100,000 tests) (Fig. 2B). Therefore, at both 5% and 20% seroprevalence, SIT^2 resulted

in the lowest overall errors. Compared to OTAs, SIT^2 yielded a similar improvement in specificity while not suffering from the significant sensitivity reduction (Fig. S2). Since the better overall performance of SIT^2 compared to OTAs was not due to increased specificity but improved sensitivity, we subsequently set out to examine these differences in more detail.

Sensitivities of single tests, OTA and SIT2 in relation to Neutralization Testing

Next, we compared the sensitivities of the three screening tests as single tests and in both two-test methods (OTA and SIT^2), benchmarking them using the Austrian sensitivity cohort (n=64) simultaneously evaluated with an authentic SARS-CoV-2 virus neutralization test (VNT). Regardless of the screening test used (Roche NC, Abbott NC, or DiaSorin S1/S2), OTAs had lower sensitivities than single tests (80·5% [78·5-83·6], 78·1% [75·8-82·8], or 75·8% [71·5-78·9] vs. 89·1%, 84·4%, or 82·8% respectively), and SIT^2 showed the best sensitivities of all methods (95·3% [93·0-96·5], 93·8% [92·2-96·5], or 87·5% [85·1-88·7]) (Fig. 3). SIT^2 algorithms, including the Roche NC and Abbott NC assays, achieved similar or even higher sensitivities than VNT (Fig. 3, VNT reference line), made possible by the unique re-testing zone of SIT^2 (Fig. S3).

Validation of the Sensitivity-Improved Two-Test using an independent cohort

To confirm the improved sensitivity of SIT^2 compared to OTA, we analyzed the sensitivities of OTAs and SIT^2 in an independent validation cohort of 976 pre-pandemic samples and 536 post-COVID samples. Out of 20 combinations using the assays Roche NC (total antibody), Abbott NC (IgG), DiaSorin S1/S2 (IgG), Siemens RBD (total antibody), and Oxford trimeric-S (IgG), a statistically significant improvement in sensitivities over OTAs was shown for SIT^2 in 18 combinations (Fig. 4). The

performance was comparable for the remaining two combinations (Siemens RBD with Oxford trimeric-S and vice versa). Still, no statistically significant improvement could be achieved due to the high pre-existing sensitivities of these assays on this particular sample cohort.

To further illustrate the effect of SIT^2 on the outcome of SARS-CoV-2 antibody testing, we compared single testing versus SIT^2 with the Abbott and DiaSorin assays at varying assumed seroprevalences (5, 10, 20, and 50%), given that the Abbott NC assay is a highly specific (99·9%), but moderately sensitive test (92·7%), and the DiaSorin S1/S2 assay has the most limited specificity (98·7%) of all evaluated assays but an acceptable sensitivity (96·3%). Regardless of whether a lack of specificity (DiaSorin S1/S2) or sensitivity (Abbott NC) had to be compensated for, SIT^2 improved the overall error rate compared to the individual tests in all four combinations and at all four assumed seroprevalence levels (Fig. 5).

Discussion

The simplest and most practical way to characterize the immune response to prior SARS-CoV-2 infection and vaccination is to measure SARS-CoV-2-specific antibodies. As long as the COVID-19 pandemic continues, serology-based assays' ongoing use and development will play a role in controlling and surveillance of this disease. The importance and value of SARS-CoV-2 antibody assays are well established(1, 3, 25-27), although some performance limitations may affect their applicability. To date, variable seroprevalence and very low antibody levels, especially in asymptomatic or mild disease(10), have challenged the performance of available tests(9, 11).

In the present work, we first sought to overcome these limitations by lowering positivity thresholds to improve test sensitivity(21-23, 28) or by using conventional orthogonal testing to maximize specificity(16, 29-32). As recommended by the CDC and used in several seroprevalence studies(31, 33), OTAs indeed maximized specificity and thus minimized false-positive results in the two sample cohorts evaluated, regardless of whether the assumed seroprevalence was 5% or 20%. However, this was accompanied by a further reduction of the a priori sub-optimal sensitivities of individual tests and increased false negatives, as shown previously(16). In seroprevalence studies, insufficient sensitivities can lead to apparent discrepancies. This is exemplified in a large study in Iceland, where only slightly more than 90% of individuals testing positive for SARS-CoV-2 by RT-PCR were seropositive(33), resulting from a priori low sensitivities of the single tests used or the OTA applied, or a combination of both.

Reducing the cut-offs for positivity leads to a significant increase in sensitivity, also shown previously(2, 21, 22), but is associated with a dramatic increase in false-positive

results at low seroprevalence. This effect only gradually decreases with increasing seroprevalence (Fig. 5).

The rates of false positives, false negatives, and overall errors of test systems depend on their specificity and sensitivity and the respective seroprevalence in the population. At a population level, at low seroprevalences, a test system with suboptimal specificity combined with insufficient sensitivity can add up to a correctly determined seroprevalence rate since false-positive, and false-negative results can compensate for each other (Fig. S2, left panel). However, at an individual level, the determination of serostatus in such a context is highly flawed, making individual-level results unreliable and unsuitable for management and subgroup analyses. For single tests, specificity and sensitivity always behave in opposite ways since an increase in the positivity threshold increases the specificity but lowers the sensitivity and vice versa. Low test specificity leads to a significant increase in the overall error rate at low seroprevalence. Low test sensitivity leads to a substantial increase in the overall error rate at high seroprevalence. However, since seroprevalence is often neither known and varies widely from region to region, it is difficult to judge whether a less specific or less sensitive test is the lesser of two evils.

To overcome these limitations, we propose a new, universally adaptable two-test system that can perform better than any other known approach regardless of the actual seroprevalence: the sensitivity-improved Two-Test or SIT^2. Its generalizability can be inferred from the following features: i) the adapted cut-offs used to optimize sensitivity were determined in various independent studies and were not explicitly calculated for our cohort(21-23); ii) SIT^2 was effective, albeit with different efficiencies, in a total of 32

different test combinations; and iii) SIT^2 was successfully validated in an independent cohort which was profoundly different from the derivation cohort. Therefore, SIT^2 does not require a particular infrastructure or the availability of high-performance individual test systems but achieves the best performance out of an available test.

Our SIT^2 strategy can rescue the specificity with minimal repeat testing required (see Table S6). For example, when applying the Roche NC as a screening test to our cohort, only 27 out of 1,181 samples needed confirmation testing with the Abbott NC test to correctly identify 62/64 true positives. Simultaneously, all false-positive results were eliminated, including those added by lowering the cut-offs (Table S4 and Fig. S1). Additionally, it was more sensitive than virus neutralization testing, which identified only 60/64 clinical positives (Fig. 3). This result is not completely surprising as it is known that not all patients who recovered from COVID-19 show detectable levels of neutralizing antibodies(34). Nevertheless, it should be noted that although antibody binding assays may have a higher sensitivity than neutralization assays, they only partially reflect the functional activity of SARS-CoV-2-reactive antibodies(35, 36).

In addition to the low or absent antibody formation shortly after COVID-19 infection(10, 34), the antibody levels also decline over time(14), and this is dependent on the assay used. For example, an antibody-positivity half-life of ~85 days has been specifically described for the Abbott NC test(15). The decrease in antibody levels or antibody reactivity measured in specific assays can lead to a status of "seroreversion" in post-COVID-19 individuals. While some of this disappearance may reflect a genuine loss of antibody (i.e., true seroreversion), some of it reflects an artifact caused by the decline of antibody levels below assay positivity thresholds, as remarkably illustrated in a recent

Brazilian study. The seroprevalence (measured by Abbott NC) apparently decreased from 46% to just under 21% between June 2020 and October 2020, when the manufacturer's recommended cut-off of 1·4 S/C was applied. In contrast, the same test results with a threshold reduced to 0·4 S/C yielded ~54% seroprevalence in June 2020 and around 45% in October 2020(2).

Our study has both strengths and limitations. One strength is the size of the cohorts examined, both in deriving the SIT^2 algorithm (N=1,181) and validating it (N=1,512). The composition of our specificity cohort is also unique: it consists of three sub-cohorts with selection criteria to further challenge analytical specificity. The lower cut-offs used to increase sensitivity were not modeled within our datasets but were derived from ROC-analyses data of independent studies(21-23). Furthermore, we were able to test the performance of the two-test systems in a total of 32 combinations, 12 in the derivation cohort and another 20 combinations in the validation cohort. As a limitation, in the Austrian cohort, only samples ≥14 days after symptom onset were included. Therefore, no conclusions on the sensitivity of the early seroconversion phase can be made from these data. Furthermore, mild and asymptomatic cases were underrepresented in the British cohort, perhaps leading to an observed higher sensitivity of the test systems.

In conclusion, we describe the novel two-test algorithm SIT^2 that, in contrast to classical orthogonal testing, can simultaneously maximize the specificity and the sensitivity of SARS-CoV-2 antibody tests. Poor sensitivity remains a problem for SARS-CoV-2 antibody tests. Low a priori antibody levels in many post-COVID-19 individuals combined with the steady antibody decline over time reduce effective sensitivity and lead to many errors in the assignment of SARS-CoV-2 serostatus. In the early post-

vaccine era of the pandemic, reliable assignation of SARS-CoV-2 serostatus will become even more relevant to public health and infection control policy. With SIT^2, we present a platform-independent and straightforward solution to minimize the overall error rate in determining SARS-CoV-2 antibody presence at the individual and population level, thereby increasing the accuracy of subsequent decisions.

Acknowledgments

We sincerely thank Marika Gerdov, Susanne Keim, Karin Mildner, Elisabeth Ponweiser, Manuela Repl, Ilse Steiner, Christine Thun, and Martina Trella for excellent technical assistance. Finally, we want to thank all the donors of the various study cohorts. The MedUni Wien Biobank is funded to participate in the biobank consortium BBMRI.at (www.bbmri.at) by the Austrian Federal Ministry of Science, Research and Technology. There was no external funding received for the work presented. However, test kits for the Technoclone ELISAs were kindly provided by the manufacturer.

Declaration of interests

NP received a travel grant from DiaSorin. DWE reports lecture fees from Gilead outside the submitted work. OCB reports grants from GSK, grants from Menarini, grants from Boehringer Ingelheim, grants from Astra, grants from MSD, grants from Pfizer, and grants from Chiesi, outside the submitted work. SH does receive unrestricted research grants (GSK, Boehringer, Menarini, Chiesi, Astra Zeneca, MSD, Novartis, Air Liquide, Vivisol, Pfizer, TEVA) for the Ludwig Boltzmann Institute of COPD and Respiratory Epidemiology, and is on advisory boards for G. SK, Boehringer Ingelheim, Novartis, Menarini, Chiesi, Astra Zeneca, MSD, Roche, Abbvie, Takeda, and TEVA for respiratory oncology and COPD. PQ is an advisory board member for Roche Austria and reports personal fees from Takeda outside the submitted work. The Dept. of Laboratory Medicine (Head: OWF) received compensations for advertisement on scientific symposia from Roche, DiaSorin, and Abbott and holds a grant for evaluating an in-vitro diagnostic device from Roche. CJB is a Board Member of Technoclone. HH receives

compensations for biobank services from Glock Health Science and Research and BlueSky immunotherapies.

References

1. Alter G, Seder R. The Power of Antibody-Based Surveillance. N Engl J Med 2020 Oct 29;383 18:1782-4. Epub 2020/09/02 as doi: 10.1056/NEJMe2028079.
2. Buss LF, Prete CA, Jr., Abrahim CMM, Mendrone A, Jr., Salomon T, de Almeida-Neto C, Franca RFO, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. Science 2021 Jan 15;371 6526:288-92. Epub 2020/12/10 as doi: 10.1126/science.abe9728.
3. Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, Marsden BD, et al. Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers. N Engl J Med 2020 Dec 23. Epub 2020/12/29 as doi: 10.1056/NEJMoa2034545.
4. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. Lancet 2021;397 10277:881—91 as doi: 10.1016/s0140-6736(21)00432-3 PMID - 33617777.
5. Eyre DW, Lumley SF, Wei J, Cox S, James T, Justice A, Jesuthasan G, et al. Quantitative SARS-CoV-2 anti-spike responses to Pfizer-BioNTech and Oxford-AstraZeneca vaccines by previous infection status. Clin Microbiol Infect 2021 Jun 7. Epub 2021/06/11 as doi: 10.1016/j.cmi.2021.05.041.

6. Perkmann T, Perkmann-Nagele N, Koller T, Mucher P, Radakovics A, Marculescu R, Wolzt M, et al. Anti-Spike Protein Assays to Determine SARS-CoV-2 Antibody Levels: a Head-to-Head Comparison of Five Quantitative Assays. Microbiol Spectr 2021 Jun 30:e0024721. Epub 2021/07/01 as doi: 10.1128/Spectrum.00247-21.

7. Wang H, Ai J, Loeffelholz MJ, Tang YW, Zhang W. Meta-analysis of diagnostic performance of serology tests for COVID-19: impact of assay design and post-symptom-onset intervals. Emerg Microbes Infect 2020 Dec;9 1:2200-11. Epub 2020/09/24 as doi: 10.1080/22221751.2020.1826362.

8. Bobrovitz N, Arora RK, Cao C, Boucher E, Liu M, Donnici C, Yanes-Lane M, et al. Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. Plos One 2021;16 6:e0252617. Epub 2021/06/24 as doi: 10.1371/journal.pone.0252617.

9. Oved K, Olmer L, Shemer-Avni Y, Wolf T, Supino-Rosin L, Prajgrod G, Shenhar Y, et al. Multi-center nationwide comparison of seven serology assays reveals a SARS-CoV-2 non-responding seronegative subpopulation. EClinicalMedicine 2020 Dec;29:100651. Epub 2020/11/26 as doi: 10.1016/j.eclinm.2020.100651.

10. Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, Hu JL, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020 Aug;26 8:1200-4. Epub 2020/06/20 as doi: 10.1038/s41591-020-0965-6.

11. Perkmann T, Perkmann-Nagele N, Breyer MK, Breyer-Kohansal R, Burghuber OC, Hartl S, Aletaha D, et al. Side-by-Side Comparison of Three Fully Automated SARS-CoV-2 Antibody Assays with a Focus on Specificity. Clin Chem 2020 Nov 1;66 11:1405-13. Epub 2020/08/11 as doi: 10.1093/clinchem/hvaa198.

12. Choe PG, Kim KH, Kang CK, Suh HJ, Kang E, Lee SY, Kim NJ, et al. Antibody Responses 8 Months after Asymptomatic or Mild SARS-CoV-2 Infection. Emerg Infect Dis;27 3:928-31 as doi: 10.3201/eid2703.204543 PMID - 33350923.

13. Seow J, Graham C, Merrick B, Acors S, Pickering S, Steel KJA, Hemmings O, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. Nat Microbiol;5 12:1598-607 as doi: 10.1038/s41564-020-00813-8 PMID - 33106674.

14. Muecksch F, Wise H, Batchelor B, Squires M, Semple E, Richardson C, McGuire J, et al. Longitudinal analysis of serology and neutralizing antibody levels in COVID19 convalescents. J Infect Dis 2020 Nov 3. Epub 2020/11/04 as doi: 10.1093/infdis/jiaa659.

15. Lumley SF, Wei J, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, et al. The duration, dynamics and determinants of SARS-CoV-2 antibody responses in individual healthcare workers. Clin Infect Dis 2021 Jan 6. Epub 2021/01/06 as doi: 10.1093/cid/ciab004.

16. Turbett SE, Anahtar M, Dighe AS, Garcia Beltran W, Miller T, Scott H, Durbin SM, et al. Evaluation of Three Commercial SARS-CoV-2 Serologic Assays and Their Performance in Two-Test Algorithms. J Clin Microbiol 2020 Dec 17;59 1. Epub 2020/10/07 as doi: 10.1128/JCM.01892-20.

17. Eskandary F, Regele H, Baumann L, Bond G, Kozakowski N, Wahrmann M, Hidalgo LG, et al. A Randomized Trial of Bortezomib in Late Antibody-Mediated Kidney Transplant Rejection [eng]. J Am Soc Nephrol 2018 Feb;29 2:591-605. Epub 2017/12/16 as doi: 10.1681/asn.2017070818.

18. Breyer-Kohansal R, Hartl S, Burghuber OC, Urban M, Schrott A, Agusti A, Sigsgaard T, et al. The LEAD (Lung, Heart, Social, Body) Study: Objectives, Methodology, and External Validity of the Population-Based Cohort Study. J Epidemiol 2019 Aug 5;29 8:315-24. Epub 2018/10/23 as doi: 10.2188/jea.JE20180039.

19. National S-C-SAEG. Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison. Lancet Infect Dis 2020 Dec;20 12:1390-400. Epub 2020/09/27 as doi: 10.1016/S1473-3099(20)30634-4.

20. Klausberger M, Duerkop M, Haslacher H, Wozniak-Knopp G, Cserjan-Puschmann M, Perkmann T, Lingg N, et al. A comprehensive antigen production and characterisation study for easy-to-implement, specific and quantitative SARS-CoV-2 serotests. Ebiomedicine 2021 Apr 24;67:103348. Epub 2021/04/28 as doi: 10.1016/j.ebiom.2021.103348.

21. Favresse J, Eucher C, Elsen M, Tre-Hardy M, Dogne JM, Douxfils J. Clinical Performance of the Elecsys Electrochemiluminescent Immunoassay for the Detection of SARS-CoV-2 Total Antibodies. Clin Chem 2020 Aug 1;66 8:1104-6. Epub 2020/06/03 as doi: 10.1093/clinchem/hvaa131.

22. Lau CS, Oh HML, Hoo SP, Liang YL, Phua SK, Aw TC. Performance of an automated chemiluminescence SARS-CoV-2 IG-G assay. Clin Chim Acta 2020 Nov;510:760-6. Epub 2020/09/11 as doi: 10.1016/j.cca.2020.09.005.

23. Bonelli F, Sarasini A, Zierold C, Calleri M, Bonetti A, Vismara C, Blocki FA, et al. Clinical and Analytical Performance of an Automated Serological Test That Identifies S1/S2-Neutralizing IgG in COVID-19 Patients Semiquantitatively. J Clin Microbiol 2020 Aug 24;58 9. Epub 2020/06/26 as doi: 10.1128/JCM.01224-20.

24. Pfluger LS, Bannasch JH, Brehm TT, Pfefferle S, Hoffmann A, Norz D, van der Meirschen M, et al. Clinical evaluation of five different automated SARS-CoV-2 serology assays in a cohort of hospitalized COVID-19 patients. J Clin Virol 2020 Sep;130:104549. Epub 2020/08/09 as doi: 10.1016/j.jcv.2020.104549.

25. Anand S, Montez-Rath M, Han J, Bozeman J, Kerschmann R, Beyer P, Parsonnet J, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. Lancet 2020 Sep 25. Epub 2020/09/29 as doi: 10.1016/S0140-6736(20)32009-2.

26. Pollan M, Perez-Gomez B, Pastor-Barriuso R, Oteo J, Hernan MA, Perez-Olmeda M, Sanmartin JL, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Lancet 2020 Aug 22;396 10250:535-44. Epub 2020/07/10 as doi: 10.1016/S0140-6736(20)31483-5.

27. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, De Ridder D, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. Lancet 2020 Aug 1;396 10247:313-9. Epub 2020/06/15 as doi: 10.1016/S0140-6736(20)31304-0.

28. Horber S, Soldo J, Relker L, Jurgens S, Guther J, Peter S, Lehmann R, et al. Evaluation of three fully-automated SARS-CoV-2 antibody assays. Clin Chem Lab Med 2020 Aug 3;58 12:2113-20. Epub 2020/08/04 as doi: 10.1515/cclm-2020-0975.

29. Xu G, Emanuel AJ, Nadig S, Mehrotra S, Caddell BA, Curry SR, Nolte FS, et al. Evaluation of Orthogonal Testing Algorithm for Detection of SARS-CoV-2 IgG

Antibodies. Clin Chem 2020 Sep 7. Epub 2020/09/08 as doi: 10.1093/clinchem/hvaa210.

30. Ripperger TJ, Uhrlaub JL, Watanabe M, Wong R, Castaneda Y, Pizzato HA, Thompson MR, et al. Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity. Immunity 2020 Nov 17;53 5:925-33 e4. Epub 2020/11/02 as doi: 10.1016/j.immuni.2020.10.004.

31. Bolotin S, Tran V, Osman S, Brown KA, Buchan SA, Joh E, Deeks SL, et al. SARS-CoV-2 seroprevalence survey estimates are affected by anti-nucleocapsid antibody decline. J Infect Dis 2021 Jan 5. Epub 2021/01/06 as doi: 10.1093/infdis/jiaa796.

32. Irsara C, Egger AE, Prokop W, Nairz M, Loacker L, Sahanic S, Pizzini A, et al. Evaluation of four commercial, fully automated SARS-CoV-2 antibody tests suggests a revision of the Siemens SARS-CoV-2 IgG assay. Clin Chem Lab Med 2021 Jan 14;0 0. Epub 2021/02/09 as doi: 10.1515/cclm-2020-1758.

33. Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, Arnthorsson AO, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. N Engl J Med 2020 Oct 29;383 18:1724-34. Epub 2020/09/02 as doi: 10.1056/NEJMoa2026116.

34. Kalkan Yazici M, Koc MM, Cetin NS, Karaaslan E, Okay G, Durdu B, Sumbul B, et al. Discordance between Serum Neutralizing Antibody Titers and the Recovery from COVID-19. J Immunol 2020 Nov 15;205 10:2719-25. Epub 2020/09/27 as doi: 10.4049/jimmunol.2000840.

35. Bal A, Pozzetto B, Trabaud MA, Escuret V, Rabilloud M, Langlois-Jacques C, Paul A, et al. Evaluation of high-throughput SARS-CoV-2 serological assays in a longitudinal cohort of patients with mild COVID-19: clinical sensitivity, specificity and association with virus neutralization test. Clin Chem 2021 Jan 5. Epub 2021/01/06 as doi: 10.1093/clinchem/hvaa336.

36. Padoan A, Bonfante F, Pagliari M, Bortolami A, Negrini D, Zuin S, Bozzato D, et al. Analytical and clinical performances of five immunoassays for the detection of SARS-CoV-2 antibodies in comparison with neutralization activity. EBioMedicine 2020 Dec;62:103101. Epub 2020/11/08 as doi: 10.1016/j.ebiom.2020.103101.

Figure Legends

Fig. 1. A) *The Sensitivity Improved Two-Test (SIT^2) algorithm includes sensitivity improvement by adapted cut-offs and a subsequent specificity rescue by re-testing all samples within the re-testing zone of the screening test by a confirmatory test.* ***B)*** *Testing algorithm for SIT^2 utilizing a screening test on an automated platform (ECLIA/Roche, CMIA/Abbott, CLIA/DiaSorin) and a confirmation test, either on one of the remaining platforms or tested by means of ELISA (Technozym RBD, NP).* ***C)*** *All test results between a reduced cut-off suggested by the literature, and a higher cut-off, above which no more false-positives were observed, were subject to confirmation testing. **… results between 12·0 and 15·0, which are according to the manufacturer considered borderline, were treated as positives; ***… suggested as a cut-off for seroprevalence testing; ****… determined by in-house modeling; [1]… see (21); [2]… see (22); [3]… see (23).*

Fig. 2. *False-positives (FP)/false-negatives (FN) (**A**) and total error (**B**) of single tests, tests with reduced thresholds according to (21-23), orthogonal testing algorithms (OTAs) and the Sensitivity Improved Two-Test (SIT^2) algorithm at 5 and 20% estimated seroprevalence. Data in (B) were compared by Mann-Whitney tests (unpaired) or Wicoxon tests (paired). *… P<0·05; **...P<0·01; ***…P<0·001.*

Fig. 3. *Sensitivities of single tests, orthogonal testing algorithms (OTAs) and the Sensitivity Improved Two-Test (SIT^2) algorithm. The dotted line indicates the sensitivity of virus neutralization test (VNT).*

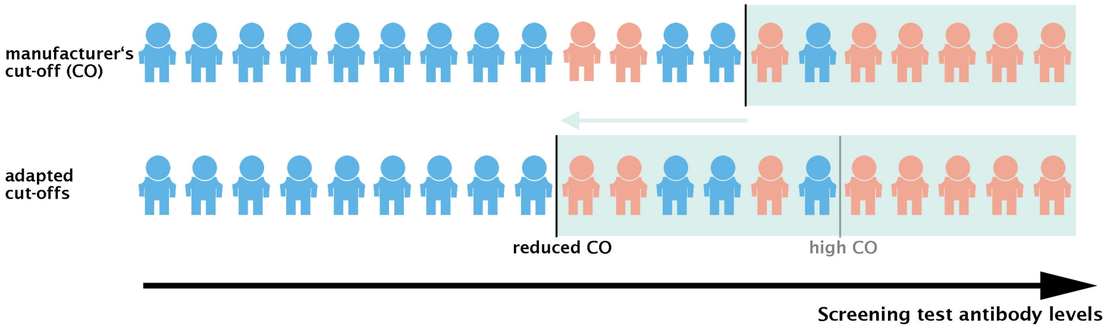
Fig. 4. *Differences in sensitivity and specificity (mean±95% confidence interval) between the Sensitivity Improved Two-Test (SIT^2) algorithm and standard orthogonal testing*

*algorithms (OTAs) within the UK validation cohort. *… P<0·05; **...P<0·01; ***…P<0·001; ****...P<0·0001*

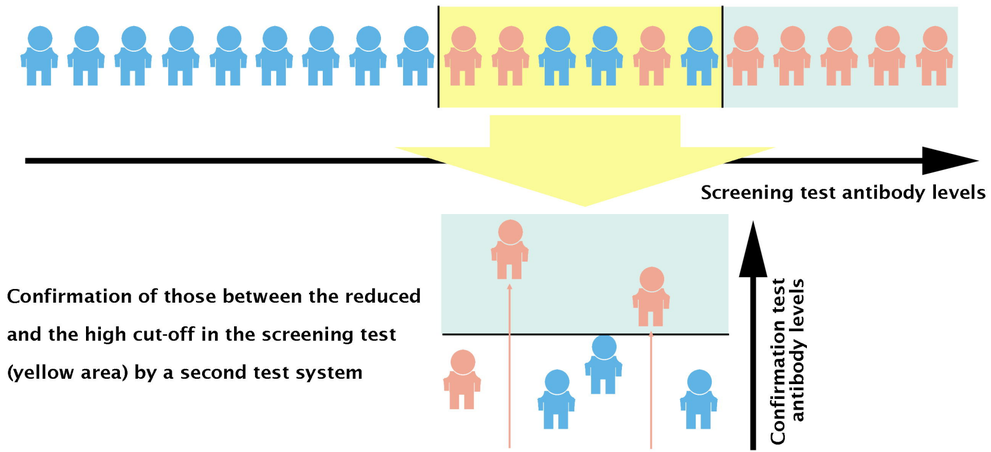
Fig. 5. *Comparing false-positives (FP), false-negatives (FN), and total error (TE) for two selected test systems,* ***A)*** *Abbott,* ***B)*** *DiaSorin, between different Sensitivity Improved Two-Test (SIT^2) combinations and the respective single test within the UK validation cohort for different estimated seroprevalences.*

A

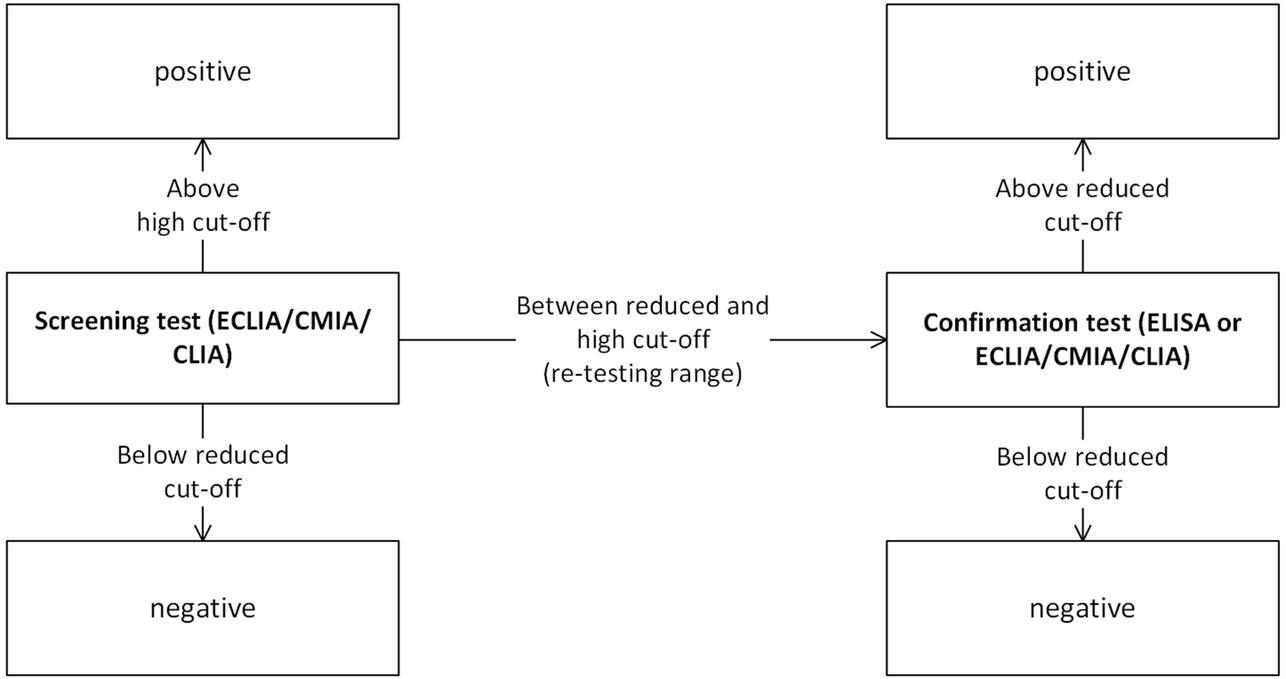
Step 1: Sensitivity Improvement



Step 2: Specificity Rescue

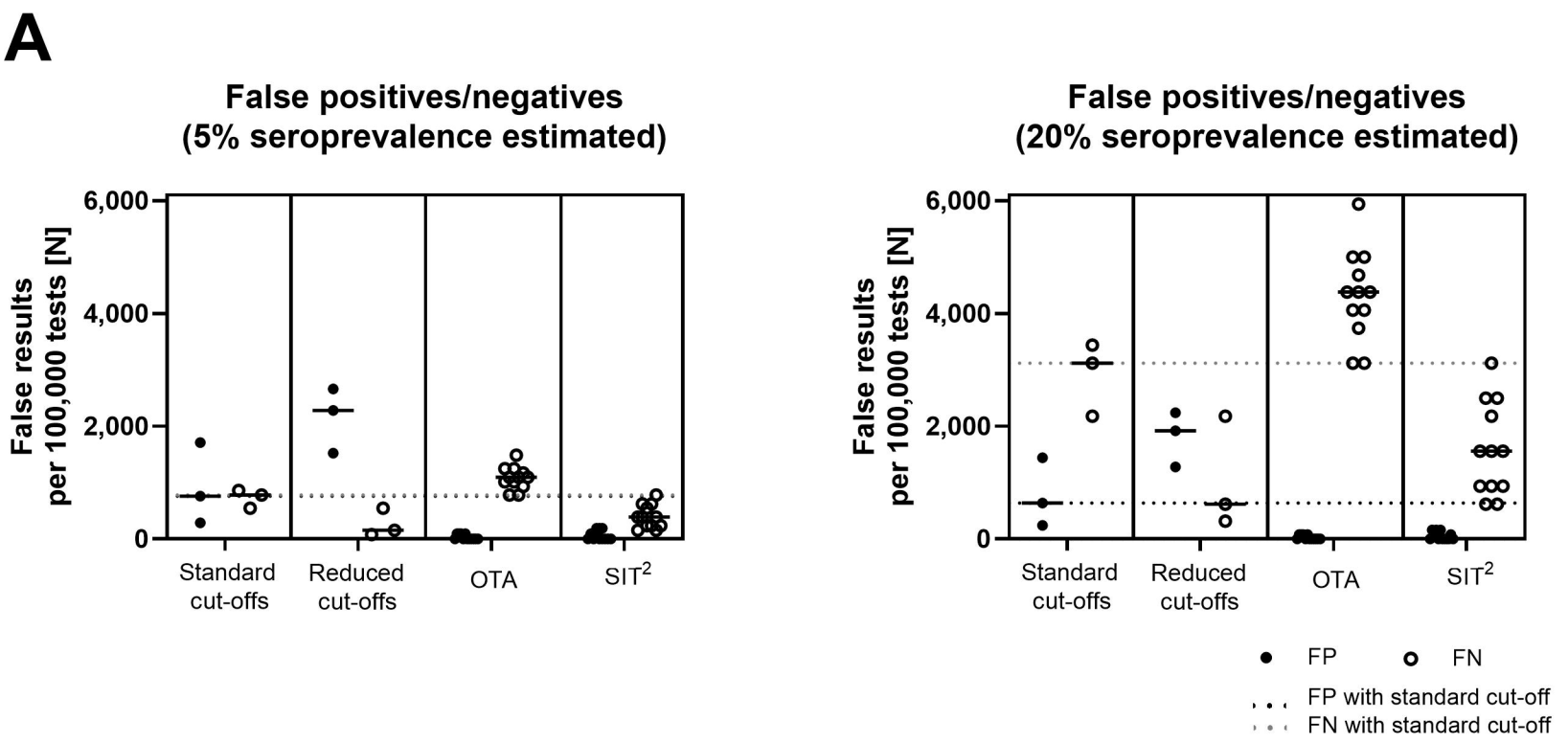


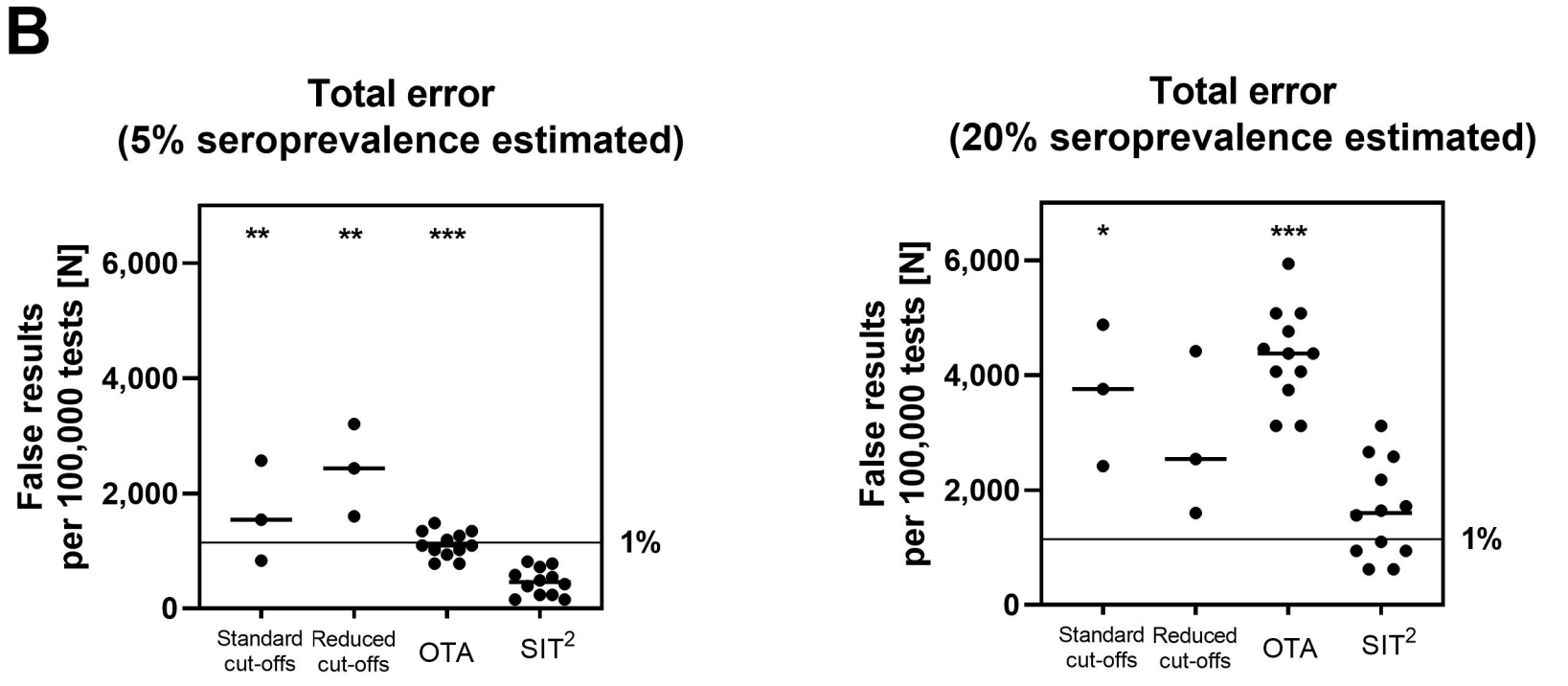
B



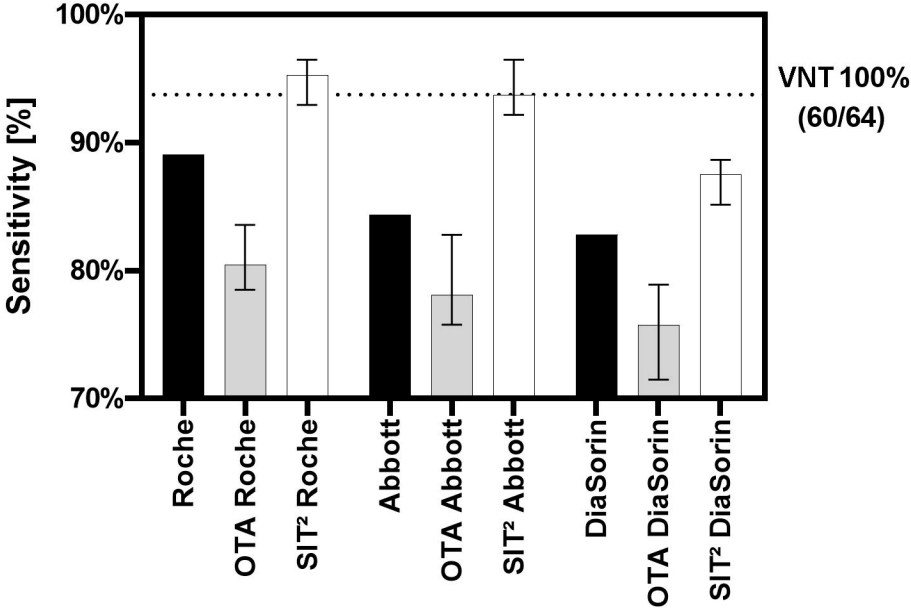
C

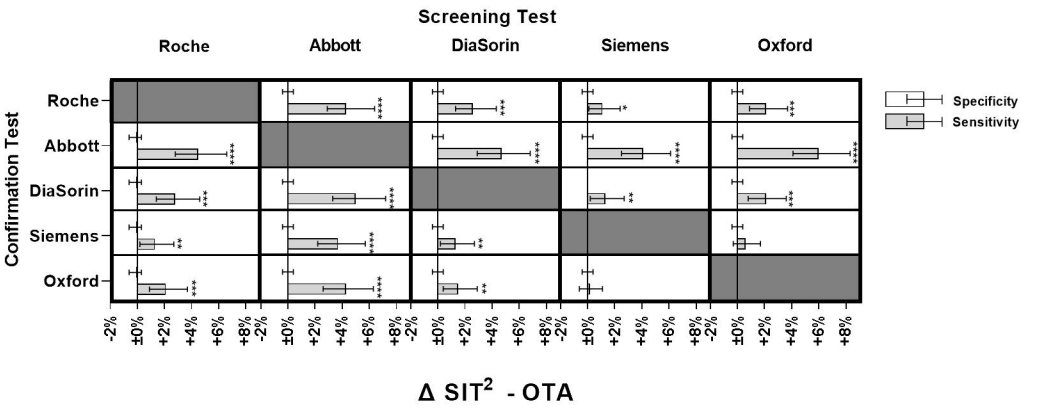
	Roche	Abbott	DiaSorin	Technozym RBD	Technozym NP
Sensitivity	89.1 (79.1-94.6)	84.4 (73.6-91.3)	82.8 (71.8-90.1)	85.9 (75.4-92.4)	78.1 (66.6-86.5)
Specificity	99.7 (99.2-99.9)	99.2 (98.5-99.6)	98.2 (97.3-98.8)	99.2 (98.5-99.6)	99.1 (98.4-99.5)
PPV (5%)	94.6 (84.9-98.2)	84.6 (74.0-91.4)	70.9 (60.2-79.2)	84.9 (74.4-91.6)	82.1 (71.0-89.6)
NPV (5%)	99.4 (98.9-99.7)	99.2 (98.6-99.5)	99.1 (98.4-99.5)	99.3 (98.7-99.6)	98.9 (98.2-99.3)
Cut-off (manufacturer)	1.000 COI	1.40 I	12.0 AU/mL**	8.000 U***	8.000 U***
Reduced Cut-off	>0.165[1]	≥0.55[2]	≥9.0[3]	>3.000 U****	>3.500 U****
High Cut-off	3.000	3.00	150.0	-	-


A
False positives/negatives
(5% seroprevalence estimated)
False positives/negatives
(20% seroprevalence estimated)
False results
per 100,000 tests [N]
6,000
4,000
2,000
0
Standard cut-offs
Reduced cut-offs
OTA
SIT2
FP
FN
FP with standard cut-off
FN with standard cut-off

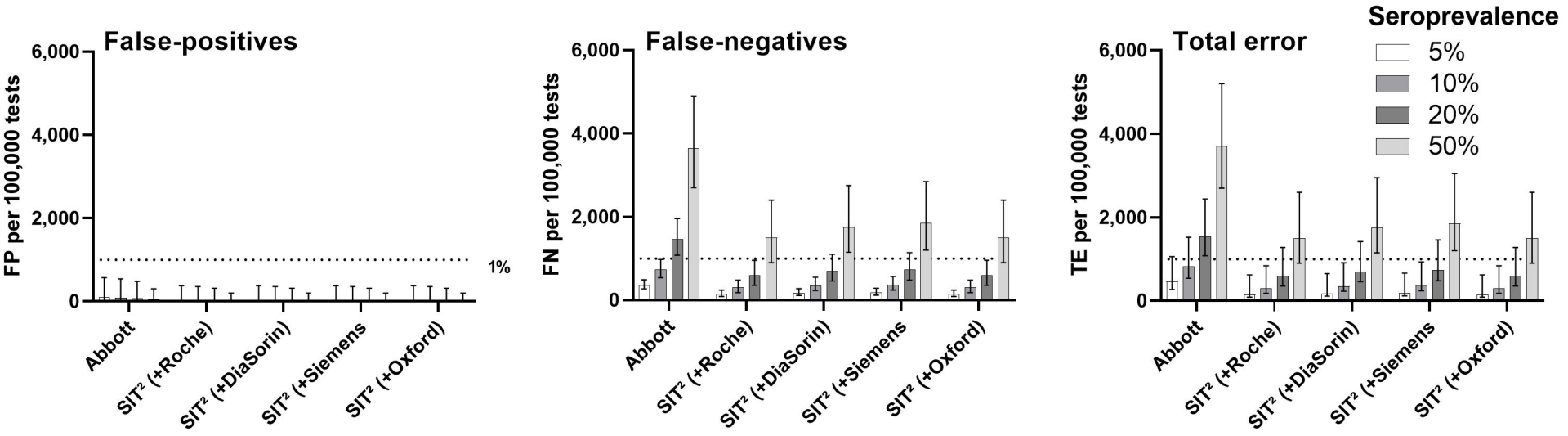
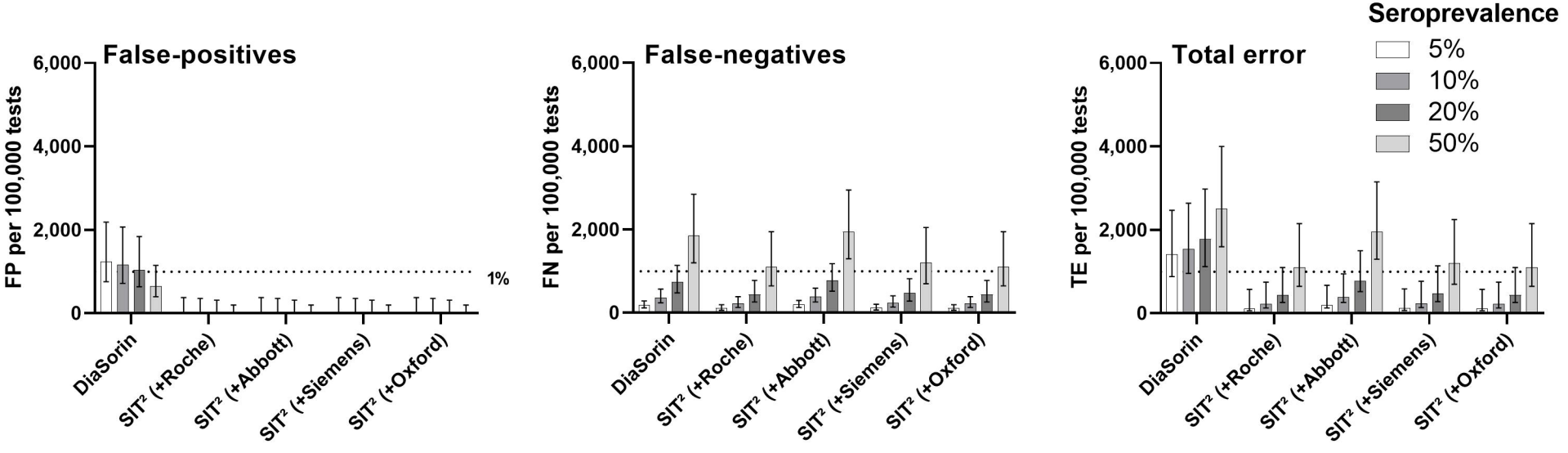

B
Total error
(5% seroprevalence estimated)
Total error
(20% seroprevalence estimated)
False results
per 100,000 tests [N]
6,000
4,000
2,000
0
**
**
***
*
***
1%
Standard cut-offs
Reduced cut-offs
OTA
SIT2

Sensitivity with standard testing, OTA and SIT^2




Screening Test
Roche
Abbott
DiaSorin
Siemens
Oxford
Confirmation Test
Roche
Abbott
DiaSorin
Siemens
Oxford
Specificity
Sensitivity
-2%
±0%
+2%
+4%
+6%
+8%
Δ SIT2 - OTA

A Abbott



B DiaSorin