Increasing test specificity without impairing sensitivity – lessons learned from 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

Abstract

Background

Serological tests are widely used in various medical disciplines for diagnostic and monitoring purposes. Unfortunately, the sensitivity and specificity of test systems is often poor, leaving room for false positive and false negative results. However, conventional methods used to increase specificity decrease sensitivity and vice versa. Using SARS-CoV-2 serology as an example, we propose here a novel testing strategy: 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

For SARS-CoV-2 serology, SIT² proved to be the best diagnostic choice at both 5% and

20% seroprevalence in all tested scenarios. It is an easy to apply algorithm and can potentially be helpful for the serology of other infectious diseases.

Running Head

Sensitivity-improved two-test serology

Keywords

serology; allergy and immunology; medical laboratory science

Key messages

What is already known on this topic

Serological tests are widely used throughout medical disciplines. When a serological assay is to be established, usually a threshold is defined above or below which a result is considered indicative of a certain medical condition. This cut-off comes with a distinct sensitivity and specificity. Sensitivity and specificity are communicating vessels – increasing one comes at the cost of the other. Common orthogonal testing algorithms concentrate on confirming positive cases, thereby increasing specificity, but decreasing sensitivity.

What this study adds

Here, we propose a novel orthogonal test algorithm applying serological assays with adjusted cut-offs. The reduction of the thresholds for positivity in both the screening and confirmation tests, as well as the additional introduction of a high cut-off in the screening test, above which no further confirmation is required, allows to increase the specificity

without compromising the sensitivity. This alorithm, which we termed „Sensitivity-improved Two-Test, SIT²“, was derived in an Austrian cohort using 5 different SARS-CoV-2 antibody tests and validated in an independent UK cohort.

How this study might affect research, practice or policy

This paper clearly shows that, in the case of SARS-CoV-2 serology, the use of two randomly chosen test systems allowed for increasing test specificity without impairing its sensitivity. This is of specific interest, when an ongoing pandemic leads to waning antibody levels – in this case, sensitivities should not be further impaired. Furthermore, we are confident that the principle of SIT² is universally applicable and that this algorithm could also be used with serological assays other than those for SARS-CoV-2.

1. Introduction

Serological tests are commonly used diagnostic tools in a broad medical field, spanning from infectiology [1, 2] to autoimmunity [3, 4], oncology [5] and transplantation medicine [6]. They also play a critical role in animal disease surveillance [7]. However, many serological tests come with acceptable but imperfect sensitivities and specificities. Tests with specificities slightly above 90% are considered good [8] or even highly specific [5]. However, at low seroprevalence rates, every single percent counts: if the frequency of a given disease in the tested population is only 5%, a specificity of 90% would mean that - even at a sensitivity of 100% - 5 true positives would be matched by ten false positives. Thus, the probability of an individual with a positive test (positive predictive value, PPV) to be a true positive would be only 33%.

During the early phase of the SARS-CoV-2 pandemic, seroprevalences were far below 1% [9]. Therefore, highly specific test systems were necessary (>99,5%) to provide good positive predictive values [10]. Sensitivity and specificity can be seen as communicating vessels – the improvement of one is usually at the expense of the other [11]. Consequently, test systems adjusted by the manufacturers to very high specificities (>99%) showed moderate sensitivity. This problem was particularly evident when non-hospitalized patients were included in the cohorts studied [12-14]. To further increase specificity at very low seroprevalence levels, various methods have been proposed, e.g., raising the thresholds for positivity or confirming a positive result with a second test (orthogonal testing) [11, 15, 16]. Unfortunately, these specificity improvement strategies inevitably lead to a further reduction of the previously moderate sensitivities.

As the pandemic progressed, the problem became more pronounced as antibodies declined, and sophisticated statistical models were required to compensate for the waning sensitivity [17]. In the case of SARS-CoV-2, as with any evolving pandemic, increasing seroprevalence rates worldwide have attenuated the need for higher specificity.

However, the problem persists in non-epidemic diseases where seroprevalence remains low. Moreover, each new pandemic begins with extremely low seroprevalence rates as well, and in the future, we should have better diagnostic strategies in infectious serology ready here.

In the present work, we propose for the first time an orthogonal test algorithm based on real-life data for the SARS-CoV-2 antibody tests of Roche, Abbott, DiaSorin, and two commercial SARS-CoV-2 ELISAs [18] intending to maximize both specificity and

sensitivity at the same time. Although our algorithm follows a general principle, it was developed based on SARS-CoV-2 antibody tests. The SARS-CoV-2 pandemic provided a unique opportunity in this regard, as historical samples from before the pandemic are negative by definition (thus allowing accurate specificity testing). In addition, sufficient PCR-confirmed positive cases were available quickly, ensuring a sophisticated and reliable sensitivity verification. Thus, for SARS-CoV-2 - in contrast to most other circulating microorganisms - a realistic and unusually accurate estimation of specificities and sensitivities of serological tests was possible. We benefited from this advantage to develop our sophisticated diagnostic algorithm.

2. Methods

2.1. 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) [19].

For derivation of the SIT^2 algorithm, sample sets from individuals known to be negative and positive were established for testing. As previously described [20], 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)[21], 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).

The SARS-CoV-2 positive cohort (n=64 samples from n=64 individuals) included patients testing positive with RT-PCR during the first wave and their close, symptomatic contacts. Of this cohort 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 and was in median 41 [26,25-49] days. 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 [22], including 1,512 serum/plasma samples (536 PCR confirmed SARS-CoV-2 positive cases; 976 negative cases, collected earlier than 2017).

2.2. 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 [total antibody assay detecting IgG, IgM and IgA antibodies against the viral nucleocapsid, 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[18] (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) [22], 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.

2.3. 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, calculated e.g. by ROC-analysis [23-25]. 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[20] and represent those values (including a safety margin) above which no more false positives were found. The highest results seen in false-positives were 1.800 COI, 2.86 Index, and 104.0 AU/mL, respectively. Hence, we defined the high cut-off for Roche and Abbott as 3.00 COI/Index and for DiaSorin as 150.0 AU/mL. 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 [16, 20]. A flowchart of the testing strategy

and the applied cut-off levels and their associated quality criteria are presented in Figs. 1B, 1C.

2.4. 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.

2.5. 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).

3. Results

In 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 7/64, 10/64, and 11/64 false-negative, as well as to 3/1,117, 9/1,117, and 20/1,117 false-positive results. Assuming a seroprevalence of 20%, this led to 2180, 3120, and 3440 false-negative results per 100,000 tests, and 240, 650 and 1,440 false-positive results per 100,000 tests respectively (Fig. 2A, right panel).

3.1. Effects of threshold lowering on Sensitivity and Specificity

Lowering the positivity thresholds for the Roche NC, Abbott NC, and Diasorin S1/S2 to 0.165 COI, 0.55 Index and 9 AU/mL increased the sensitivity significantly and reduced false-negative results to 63/64, 62/64, and 57/64 (320, 620, and 2,180 per 100,000 tests at a seroprevalence of 20%), but substantially increased false-positive results to 18/1,117, 27/1,117, and 31/1,117, respectively (1,280, 1,920 and 2,240 per 100,000 tests, an assumed seroprevalence of 20%; Table S4, Fig. 2A, right panel).

3.2. 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 (0-1)/1,117. However, the rate of false negatives was 14 (12-16)/64 (1,095 [955-1,230] per 100,000 tests at 20% seroprevalence), and therefore considerably higher than for single

testing strategies. In contrast, the SIT^2 algorithm minimized false positives to 0 (0-2)/1,117 (0 [0-140) per 100,000 tests at 20% seroprevalence) while also reducing false negatives to 5 (3-8)/64 (1,560 [940-2420] per 100,000 tests at 20% seroprevalence, Fig. 2A right panel; Table S5).

3.3. 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.

3.4. Sensitivities of single tests, OTA and SIT^2 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).

3.5. 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

297 highly specific (99.9%), but moderately sensitive test (92.7%), and the DiaSorin S1/S2
298 assay has the most limited specificity (98.7%) of all evaluated assays but an acceptable
299 sensitivity (96.3%). Regardless of whether a lack of specificity (DiaSorin S1/S2) or
300 sensitivity (Abbott NC) had to be compensated for, SIT^2 improved the overall error rate
301 compared to the individual tests in all four combinations and at all four assumed
302 seroprevalence levels (Fig. 5).

4. Discussion

Serology is a commonly used, multi-purpose analytical method [1-6]. However, not all serologic assays have appropriate sensitivities and specificities, especially in low-prevalence settings. The SARS-CoV-2 pandemic prompted the simultaneous development of several antibody tests and, which is rare otherwise, allowed to evaluate these tests with both confirmed positive and negative cases, the latter derived from biobank collections established before the virus emerged. In the case of SARS-CoV-2, false-positive samples are usually not simultaneously reactive in different test systems [16, 20]. This led to the hypothesis that reducing the threshold for positivity in screening and confirmation tests would increase the specificity without impairing the sensitivity. A further improvement in sensitivity was possible by defining a high cut-off for the screening test, above which, due to the excellent reliability of high test results, no further confirmation (and, thereby, a possible false-negative result in the confirmation test) was necessary.

In the early waves of the SARS-CoV-2 pandemic, many commercially available SARS-CoV-2 antibody tests did not provide sufficient specificity to achieve acceptable positive predictive values (PPVs), for example, at a seroprevalence rate of 1-5% [15, 20]. Lowering positivity thresholds might improve test sensitivity [23-25] and conventional orthogonal testing can maximize specificity [11, 26, 27]. The latter might increase the positive predictive value, but PPV will only be relevantly increased at low seroprevalences. 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.

Based on actual data related to SARS-CoV-2, we propose a new, universally adaptable two-test system that could, in the case of SARS-CoV-2, perform better than any other known approach regardless of the actual seroprevalence: the sensitivity-improved Two-Test or SIT^2. For this, we established the algorithm in our COVID-19 cohort (including 1181 samples, 1117 pre-pandemic negative, and 64 confirmed post-COVID positive samples) and validated it in a completely independent UK cohort (including 1512 samples, 976 negatives, and 536 positives). So, the associations found were neither exclusively related to a particular cohort nor the analyzing institutions. All Austrian cohort samples were tested with the following assays: Roche, Abbott, DiaSorin S1/S2, Technozym RBD, and Technozym NP. The UK cohort we used for validation included a complete data set of all samples analyzed with the Roche, Abbott, DiaSorin S1/S2, Siemens, Oxford assays. Hence, the Austrian and the UK cohorts shared three test systems (Roche, Abbott, DiaSorin S1/S2) but differed regarding specific characteristics of the included negative and positive samples. Besides these three overlapping test systems, each cohort included data of two more exclusive SARS-CoV-2 antibody assays in the analysis. The use of these different combinations should underscore the universality of SIT^2.

Its generalizability can be inferred further in detail 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 [23-25]; 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. The robustness of a diagnostic algorithm regarding analytical variability (lot-to-lot variability, instrument-dependent variability, or method-specific

confounders) is essential. Based on our study design with three overlapping assays (Roche, Abbott, DiaSorin) tested at two sites with two different cohorts but without lot matching, we did not find any adverse effects on the robustness of our algorithm by these potential confounders. Moreover, we estimated the SIT²-robustness to between-lot variability simulating how the algorithm's performance would change if results would vary according to their respective reference change values (RCVs). For this, we used a SIT²-algorithm consisting of Roche and Abbott as an example and could conclude that expectable between-assay variability might only marginally affect the algorithm (data not shown). Therefore, SIT² does not require a particular infrastructure, the availability of high-performance individual test systems or specific reagent lots to work, but can optimize the performance of any available test system.

Our SIT² 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 [28]. 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 [29, 30]. The sensitivity of SARS-CoV-2 tests may change over time, as prominently shown in a Brazilian study, where pronounced antibody waning led to an apparent decrease in seroprevalence already a

few months after a SARS-CoV-2 corona wave [17]. However, this was mainly caused by the strongly decreasing sensitivity of the test system used. The measured seroprevalence decreased from 46.3% in June 2020 to only 20.7% in October 2020, when the standard manufacturer cut-off of 1.4 was used for the Abbott NC test. When the same data were analyzed with a reduced cut-off of 0.4, the values changed from 54.3% in June to 44.6% in October, so the apparent decrease in seroprevalence was much less pronounced. Lowering the cut-off to increase the sensitivity of a test system (and therefore also to compensate for such time-dependent sensitivity losses) is the first step of our SIT^2 algorithm. As this cut-off lowering reduces the specificity of a test (so with the 0.4 cut-off, the seroprevalence rate in June was 8% higher than with the 1.4 cut-off, including more false-positives), it is necessary to rescue this loss of specificity by a second test (also highly sensitive by cut-off lowering). This should illustrate that while there are test systems whose sensitivity changes more rapidly over time and there is physiologically a time-dependent decrease in antibody levels, SIT^2 offers a strategy to counteract this development with an increase in sensitivity by cut-off lowering and subsequent correction of specificity. Thus, these time-dependent sensitivity changes are not a significant problem for SIT^2. Accordingly, there are far-reaching potential applications. Regarding SARS-CoV-2, on the one hand, the use of an algorithm of this kind could increase the reliability of seroprevalence analyses, especially in low-prevalence areas. On the other hand, its use in routine clinical diagnostics is also conceivable. In the case of SARS-CoV-2, the emergence of new viral variants particularly affects test sensitivity [31].This could be counteracted by increasing sensitivity through modified cut-offs, and specificity would subsequently be restored by a second test. For SARS-CoV-2 testing, it must be further emphasized that different

mechanisms of immunization induce different humoral responses. Whereas an infection usually leads to both anti-nucleocapsid and anti-spike antibodies, the immune response to an mRNA-, vector-, or protein-based vaccine that introduces only the spike-protein lacks the anti-nucleocapsid antibody [32]. Accordingly, amongst the vaccinated, tests assessing anti-spike antibodies might not be useful in detecting individuals after SARS-CoV-2 infection, as the measured amount would have at least partly been induced by the vaccine. However, add-on infection could boost anti-spike levels [33]. These conditions must be considered when searching for the optimal combination of tests for a SIT^2 approach.

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 [23-25]. 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. Moreover, the analysis did only include samples collected during the first wave, therefore, positive individuals were most likely infected by the wildtype virus. However, as stated above, the emergence of new variants challenges a test system’s sensitivity

even more, which only reinforces the need to increase sensitivity without harming specificity, as we propose here by using SIT^2.

In conclusion, we describe the novel two-test algorithm SIT^2, which makes it possible to maintain or even significantly improve sensitivity while approaching 100% specificity.

5. Funding

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.

6. 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.

7. Competing interests

All authors have read the journal's policy on disclosure of potential conflicts of interest and the following has to be disclosed: NP-N 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.

8. Ethics approval statement

The overall evaluation plan was reviewed and approved by the ethics committee of the Medical University of Vienna (1424/2020).

9. Contributorship Statement

TP und TK contributed equally. Conceptualization: TP, TK, OFW, HH. Methodology: TP, TK, NP-N, HH; Investigation: TP, TK, NP-N, MO-K, DWE, PM, AB, NS, M-LB, RB-K, OCB, SH, DA, DS, PQ, RM, PM, AR, MK, MD, BHo, BHa, RS, GL, FG, WG, RG, HH; Data curation: TP, TK, DW, PM, AB, NS, M-LB, RB-K, OCB, SH, DA, DS; Project administration: PM, AR; Formal analysis: HH; Validation: DWE, PM, AB, NS; Writing – original draft: TP, TK, NP-N, HH; Visualization: HH; Supervision: OFW, CJB, HH; Resources: DWE, PM, AB, NS, M-LB, RB-K, OCB, SH, DA, DS, PQ, RM, MK, MD, BHo, BHa, RS, GL, FG, WG, RG, OFW, CJB; Writing – review & editing: all authors.

10. References

[1] C. Lass-Flörl, E. Samardzic, M. Knoll, Serology anno 2021-fungal infections: from invasive to chronic, Clin Microbiol Infect 27(9) (2021) 1230-1241.

[2] S.A. Davison, S.I. Strasser, Ordering and interpreting hepatitis B serology, Bmj 348 (2014) g2522.

[3] A. Popp, L. Kivelä, V. Fuchs, K. Kurppa, Diagnosing Celiac Disease: Towards Wide-Scale Screening and Serology-Based Criteria?, Gastroenterol Res Pract 2019 (2019) 2916024.

[4] B. Terziroli Beretta-Piccoli, G. Mieli-Vergani, D. Vergani, Serology in autoimmune hepatitis: A clinical-practice approach, Eur J Intern Med 48 (2018) 35-43.

[5] J. Hibbert, G. Halec, D. Baaken, T. Waterboer, N. Brenner, Sensitivity and Specificity of Human Papillomavirus (HPV) 16 Early Antigen Serology for HPV-Driven Oropharyngeal Cancer: A Systematic Literature Review and Meta-Analysis, Cancers (Basel) 13(12) (2021) 3010.

[6] M.K. Smatti, D.W. Al-Sadeq, N.H. Ali, G. Pintus, H. Abou-Saleh, G.K. Nasrallah, Epstein-Barr Virus Epidemiology, Serology, and Genetic Variability of LMP-1 Oncogene Among Healthy Population: An Update, Front Oncol 8 (2018) 211.

[7] M. Gaudino, A. Moreno, C.J. Snoeck, S. Zohari, C. Saegerman, T. O'Donovan, E. Ryan, I. Zanni, E. Foni, A. Sausy, J.M. Hübschen, G. Meyer, C. Chiapponi, M.F. Ducatez, Emerging Influenza D virus infection in European livestock as determined in serology studies: Are we underestimating its spread over the continent?, Transbound Emerg Dis 68(3) (2021) 1125-1135.

[8] Y. Shibata, N. Horita, M. Yamamoto, T. Tsukahara, H. Nagakura, K. Tashiro, H. Watanabe, K. Nagai, K. Nakashima, R. Ushio, M. Ikeda, A. Narita, A. Kanai, T. Sato, T.

Kaneko, Diagnostic test accuracy of anti-glycopeptidolipid-core IgA antibodies for Mycobacterium avium complex pulmonary disease: systematic review and meta-analysis, Sci Rep 6 (2016) 29325.

[9] R.K. Arora, A. Joseph, J. Van Wyk, S. Rocco, A. Atmaja, E. May, T. Yan, N. Bobrovitz, J. Chevrier, M.P. Cheng, T. Williamson, D.L. Buckeridge, SeroTracker: a global SARS-CoV-2 seroprevalence dashboard, Lancet Infect Dis 21(4) (2020) e75-e76.

[10] D. Bailey, D. Konforte, V.E. Barakauskas, P.M. Yip, V. Kulasingam, M. Abou El Hassan, L.A. Beach, I.M. Blasutig, P. Catomeris, K.C. Dooley, Y. Gong, P. Kavsak, E.W. Randell, J.L. Robinson, J. Shaw, J. Taher, N. White-Al Habeeb, Canadian society of clinical chemists (CSCC) interim consensus guidance for testing and reporting of SARS-CoV-2 serology, Clin Biochem 86 (2020) 1-7.

[11] S.E. Turbett, M. Anahtar, A.S. Dighe, W. Garcia Beltran, T. Miller, H. Scott, S.M. Durbin, M. Bharadwaj, J. Thomas, T.S. Gogakos, M. Astudillo, J. Lennerz, E.S. Rosenberg, J.A. Branda, Evaluation of Three Commercial SARS-CoV-2 Serologic Assays and Their Performance in Two-Test Algorithms, J Clin Microbiol 59(1) (2020) e01892-20.

[12] P. Gattinger, K. Borochova, Y. Dorofeeva, R. Henning, R. Kiss, B. Kratzer, B. Muhl, T. Perkmann, D. Trapin, M. Trella, P. Ettel, I. Tulaeva, W.F. Pickl, R. Valenta, Antibodies in serum of convalescent patients following mild COVID-19 do not always prevent virus-receptor binding, Allergy 76(3) (2020) 878-883.

[13] Q.X. Long, X.J. Tang, Q.L. Shi, Q. Li, H.J. Deng, J. Yuan, J.L. Hu, W. Xu, Y. Zhang, F.J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X.M. Liu, J.J. Li, J.F. Qiu, J. Chen, A.L. Huang, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections, Nat Med 26(8) (2020) 1200-1204.

512 [14] R.C. Baron, L. Risch, M. Weber, S. Thiel, K. Grossmann, N. Wohlwend, T. Lung, D.
513 Hillmann, M. Ritzler, S. Bigler, K. Egli, F. Ferrara, T. Bodmer, M. Imperiali, S. Heer, H.
514 Renz, L. Flatz, P. Kohler, P. Vernazza, C.R. Kahlert, M. Paprotny, M. Risch, Frequency
515 of serological non-responders and false-negative RT-PCR results in SARS-CoV-2
516 testing: a population-based study, Clin Chem Lab Med 58(12) (2020) 2131-2140.
517 [15] D.M. Manthei, J.F. Whalen, L.F. Schroeder, A.M. Sinay, S.H. Li, R. Valdez, D.A.
518 Giacherio, C. Gherasim, Differences in Performance Characteristics Among Four High-
519 Throughput Assays for the Detection of Antibodies Against SARS-CoV-2 Using a
520 Common Set of Patient Samples, Am J Clin Pathol 155(2) (2020) 267-279.
521 [16] L.S. Pfluger, J.H. Bannasch, T.T. Brehm, S. Pfefferle, A. Hoffmann, D. Norz, M. van
522 der Meirschen, S. Kluge, M. Haddad, S. Pischke, J. Hiller, M.M. Addo, A.W. Lohse,
523 J.S.Z. Wiesch, S. Peine, M. Aepfelbacher, M. Lutgehetmann, Clinical evaluation of five
524 different automated SARS-CoV-2 serology assays in a cohort of hospitalized COVID-19
525 patients, J Clin Virol 130 (2020) 104549.
526 [17] L.F. Buss, C.A. Prete, Jr., C.M.M. Abrahim, A. Mendrone, Jr., T. Salomon, C. de
527 Almeida-Neto, R.F.O. França, M.C. Belotti, M. Carvalho, A.G. Costa, M.A.E. Crispim,
528 S.C. Ferreira, N.A. Fraiji, S. Gurzenda, C. Whittaker, L.T. Kamaura, P.L. Takecian, P. da
529 Silva Peixoto, M.K. Oikawa, A.S. Nishiya, V. Rocha, N.A. Salles, A.A. de Souza Santos,
530 M.A. da Silva, B. Custer, K.V. Parag, M. Barral-Netto, M.U.G. Kraemer, R.H.M. Pereira,
531 O.G. Pybus, M.P. Busch, M.C. Castro, C. Dye, V.H. Nascimento, N.R. Faria, E.C.
532 Sabino, Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a
533 largely unmitigated epidemic, Science 371(6526) (2021) 288-292.
534 [18] M. Klausberger, M. Duerkop, H. Haslacher, G. Wozniak-Knopp, M. Cserjan-
535 Puschmann, T. Perkmann, N. Lingg, P.P. Aguilar, E. Laurent, J. De Vos, M. Hofner, B.

Holzer, M. Stadler, G. Manhart, K. Vierlinger, M. Egger, L. Milchram, E. Gludovacz, N. Marx, C. Köppl, C. Tauer, J. Beck, D. Maresch, C. Grünwald-Gruber, F. Strobl, P. Satzer, G. Stadlmayr, U. Vavra, J. Huber, M. Wahrmann, F. Eskandary, M.K. Breyer, D. Sieghart, P. Quehenberger, G. Leitner, R. Strassl, A.E. Egger, C. Irsara, A. Griesmacher, G. Hoermann, G. Weiss, R. Bellmann-Weiler, J. Loeffler-Ragg, N. Borth, R. Strasser, A. Jungbauer, R. Hahn, J. Mairhofer, B. Hartmann, N.B. Binder, G. Striedner, L. Mach, A. Weinhäusel, B. Dieplinger, F. Grebien, W. Gerner, C.J. Binder, R. Grabherr, A comprehensive antigen production and characterisation study for easy-to-implement, specific and quantitative SARS-CoV-2 serotests, EBioMedicine 67 (2021) 103348.

[19] F. Eskandary, H. Regele, L. Baumann, G. Bond, N. Kozakowski, M. Wahrmann, L.G. Hidalgo, H. Haslacher, C.C. Kaltenecker, M.B. Aretin, R. Oberbauer, M. Posch, A. Staudenherz, A. Handisurya, J. Reeve, P.F. Halloran, G.A. Böhmig, A Randomized Trial of Bortezomib in Late Antibody-Mediated Kidney Transplant Rejection, J Am Soc Nephrol 29(2) (2018) 591-605.

[20] T. Perkmann, N. Perkmann-Nagele, M.K. Breyer, R. Breyer-Kohansal, O.C. Burghuber, S. Hartl, D. Aletaha, D. Sieghart, P. Quehenberger, R. Marculescu, P. Mucher, R. Strassl, O.F. Wagner, C.J. Binder, H. Haslacher, Side-by-Side Comparison of Three Fully Automated SARS-CoV-2 Antibody Assays with a Focus on Specificity, Clin Chem 66(11) (2020) 1405-1413.

[21] R. Breyer-Kohansal, S. Hartl, O.C. Burghuber, M. Urban, A. Schrott, A. Agusti, T. Sigsgaard, C. Vogelmeier, E. Wouters, M. Studnicka, M.K. Breyer, The LEAD (Lung, Heart, Social, Body) Study: Objectives, Methodology, and External Validity of the Population-Based Cohort Study, J Epidemiol 29(8) (2019) 315-324.

[22] S.-C.-S.A.E.G. National, Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison, Lancet Infect Dis 20(12) (2020) 1390-1400.

[23] J. Favresse, C. Eucher, M. Elsen, M. Tre-Hardy, J.M. Dogne, J. Douxfils, Clinical Performance of the Elecsys Electrochemiluminescent Immunoassay for the Detection of SARS-CoV-2 Total Antibodies, Clin Chem 66(8) (2020) 1104-1106.

[24] C.S. Lau, H.M.L. Oh, S.P. Hoo, Y.L. Liang, S.K. Phua, T.C. Aw, Performance of an automated chemiluminescence SARS-CoV-2 IG-G assay, Clin Chim Acta 510 (2020) 760-766.

[25] F. Bonelli, A. Sarasini, C. Zierold, M. Calleri, A. Bonetti, C. Vismara, F.A. Blocki, L. Pallavicini, A. Chinali, D. Campisi, E. Percivalle, A.P. DiNapoli, C.F. Perno, F. Baldanti, Clinical and Analytical Performance of an Automated Serological Test That Identifies S1/S2-Neutralizing IgG in COVID-19 Patients Semiquantitatively, J Clin Microbiol 58(9) (2020) e01224-20.

[26] S. Bolotin, V. Tran, S. Osman, K.A. Brown, S.A. Buchan, E. Joh, S.L. Deeks, V.G. Allen, SARS-CoV-2 seroprevalence survey estimates are affected by anti-nucleocapsid antibody decline, J Infect Dis 223(8) (2021) 1334-1338.

[27] C. Irsara, A.E. Egger, W. Prokop, M. Nairz, L. Loacker, S. Sahanic, A. Pizzini, T. Sonnweber, W. Mayer, H. Schennach, J. Loeffler-Ragg, R. Bellmann-Weiler, I. Tancevski, G. Weiss, M. Anliker, A. Griesmacher, G. Hoermann, 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 59(6) (2021) 1143-1154.

582 [28] M. Kalkan Yazici, M.M. Koc, N.S. Cetin, E. Karaaslan, G. Okay, B. Durdu, B.
583 Sumbul, M.Z. Doymaz, Discordance between Serum Neutralizing Antibody Titers and
584 the Recovery from COVID-19, J Immunol 205(10) (2020) 2719-2725.
585 [29] A. Bal, B. Pozzetto, M.A. Trabaud, V. Escuret, M. Rabilloud, C. Langlois-Jacques,
586 A. Paul, N. Guibert, C. D'Aubarede-Frieh, A. Massardier-Pilonchery, N. Fabien, D.
587 Goncalves, A. Boibieux, F. Morfin-Sherpa, V. Pitiot, F. Gueyffier, B. Lina, J.B. Fassier, S.
588 Trouillet-Assant, G. Covid Ser Study, Evaluation of high-throughput SARS-CoV-2
589 serological assays in a longitudinal cohort of patients with mild COVID-19: clinical
590 sensitivity, specificity and association with virus neutralization test, Clin Chem 67(5)
591 (2021) 742-752.
592 [30] A. Padoan, F. Bonfante, M. Pagliari, A. Bortolami, D. Negrini, S. Zuin, D. Bozzato,
593 C. Cosma, L. Sciacovelli, M. Plebani, Analytical and clinical performances of five
594 immunoassays for the detection of SARS-CoV-2 antibodies in comparison with
595 neutralization activity, EBioMedicine 62 (2020) 103101.
596 [31] T.P. David Niklas Springer, Claudia Jani, Patrick Mucher, Katja Prüger, Rodrig
597 Marculescu, Elisabeth Reuberger, Jeremy Vann Camp, Marianne Graninger, Christian
598 Borsodi, Josef Deutsch, Oliver Lammel, Stephan Walter Aberle, Elisabeth Puchhammer-
599 Stöckl, Helmuth Haslacher, Eva Höltl, Judith Helene Aberle, Karin Stiasny, Lukas
600 Weseslindtner, Reduced sensitivity of commercial Spike-specific antibody assays after
601 primary infection with the SARS-CoV-2 Omicron variant. PREPRINT (Version 3)
602 available at Research Square 2022. https://doi.org/10.21203/rs.3.rs-1669740/v3.
603 (Accessed 2022-07-15.

[32] R. Midorikawa, M. Nakama, H. Furukawa, S. Oka, T. Higuchi, H. Nagai, N. Nagai, S. Tohma, Detection of SARS-CoV-2 Nucleocapsid, Spike, and Neutralizing Antibodies in Vaccinated Japanese, Viruses 14(5) (2022).

[33] T. Perkmann, N. Perkmann-Nagele, T. Koller, P. Mucher, A. Radakovics, M. Wolzt, O.F. Wagner, C.J. Binder, H. Haslacher, Serum antibody response to BNT162b2 after natural SARS-CoV-2 infection, Eur J Clin Invest 51(11) (2021) e13632.

11. Figure Legends

Fig. 1. A) *The Sensitivity Improved Two-Test (SIT²) 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 SIT² 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 [23]; [2]…see [24]; [3]… see [25].*

Fig. 2. *False-positives (FP)/false-negatives (FN) (**A**) and total error (**B**) of single tests, tests with reduced thresholds according to [23-25], orthogonal testing algorithms (OTAs) and the Sensitivity Improved Two-Test (SIT²) 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 (SIT²) 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 (SIT²) 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.*