Performance of SARS-CoV-2 nucleic acid amplification testing in Austria as measured by external quality assessment schemes during 3 years of the COVID-19 pandemic: an observational retrospective study

Christoph Buchta, Stephan W Aberle, Franz Allerberger, Bernhard Benka, Irene Görzer, Andrea Griesmacher, Wolfgang Hübl, Wolfgang Huf, Stylianos Kapiotis, Mathias M Müller, Erich Neuwirth, Elisabeth Puchhammer-Stöckl, Lukas Weseslindtner, Jeremy V Camp

Summary

Background The aim of external quality assessment (EQA) schemes is to evaluate the analytical performance of laboratories and test systems in a near-to-real-life setting. This monitoring service provides feedback to participant laboratories and serves as a control measure for the epidemiological assessment of the regional incidence of a pathogen, particularly during epidemics. Using data from EQA schemes implemented as a result of the intensive effort to monitor SARS-CoV-2 infections in Austria, we aimed to identify factors that explained the variation in laboratory performance for SARS-CoV-2 detection over the course of the COVID-19 pandemic.

Methods For this observational study, we retrospectively analysed 6308 reverse transcriptase quantitative PCR (RT-qPCR) test results reported by 191 laboratories on 71 samples during 14 rounds of three SARS-CoV-2 pathogen detection EQA schemes in Austria between May 18, 2020, and Feb 20, 2023. We calculated the overall rates of false and true-negative, false and true-positive, and inconclusive results. We then assessed laboratory performance by estimating the sensitivity by testing whether significant variation in the odds of obtaining a true-positive result could be explained by virus concentration, laboratory type, or assay format. We also assessed whether laboratory performance changed over time.

Findings 4371 (93·7%) of 4663 qPCR test results were true-positive, 241 (5·2%) were false-negative, and 51 (1·1%) were inconclusive. The mean per-sample sensitivity was 99·7% in samples with high virus concentrations (1383 [99·4%] true-positive, three [0·2%] false-negative, and five [0·4%] inconclusive results for 1391 tests in which the sample cycle threshold was ≤32), whereas detection rates were lower in samples with low virus concentrations (mean per-sample sensitivity 92·5%; 2988 [91·3%] true-positive, 238 [7·3%] false-negative, and 46 [1·4%] inconclusive results for 3272 tests in which the cycle threshold was >32). Of the 1645 results expected to be negative, 1561 (94·9%) were correctly reported as negative, 10 (0·6%) were incorrectly reported as positive, and 74 (4·5%) were reported as inconclusive. Notably, the overall performance of the tests did not change significantly over time. The odds of reporting a correct result were 2·94 (95% CI 1·75–4·96) times higher for a medical laboratory than for a non-medical laboratory, and 4·60 (2·91–7·41) times greater for automated test systems than for manual test systems. Automated test systems within medical laboratories had the highest sensitivity when compared with systems requiring manual intervention in both medical and non-medical laboratories.

Interpretation High rates of false-negativity in all PCR analyses evaluated in comprehensive, multiple, and repeated EQA schemes outline a clear path for improvement in the future. The performance of some laboratories (eg, non-medical laboratories or those using non-automated test systems) should receive additional scrutiny—for example, by requiring additional EQA schemes for certification or accreditation—if the aggregated data from EQA rounds suggest lower sensitivity than that recorded by others. This strategy will provide assurances that epidemiological data as a whole are reliable when testing on such a large scale. Although performance did not improve over time, we cannot exclude extenuating circumstances—such as shortages and weakened supply chains—that could have prevented laboratories from seeking alternative methods to improve performance.

Funding None.

Introduction

External quality assessment (EQA) schemes consist of recurring rounds in which panels of samples with identical undisclosed content and properties are sent to participating laboratories, who are invited to analyse them in the same way as they analyse patient samples

Lancet Microbe 2023

Published Online November 15, 2023 https://doi.org/10.1016/S2666-5247(23)00286-0

Austrian Association for Quality Assurance and Standardization of Medical and Diagnostic Tests (ÖQUASTA), Vienna, Austria (C Buchta MD, Prof A Griesmacher MD, Prof W Hübl MD, Prof S Kapiotis MD, Prof M M Müller MD); **Center for Virology, Medical University of Vienna, Vienna, Austria** (Prof S W Aberle MD, Prof I Görzer PhD, Prof E Puchhammer-Stöckl MD, Prof L Weseslindtner MD, J V Camp PhD); **Austrian Agency for Health and Food Safety (AGES), Vienna, Austria** (Prof F Allerberger MD, B Benka MD); **Central Institute for Medical and Chemical Laboratory Diagnosis, Innsbruck University Hospital, Innsbruck, Austria** (Prof A Griesmacher); **Department of Laboratory Medicine, Klinik Ottakring, Vienna, Austria** (Prof W Hübl); **Karl Landsteiner Institute for Clinical Risk Management, Vienna, Austria** (W Huf MD); **LABCON Medizinische Laboratorien, Vienna, Austria** (Prof S Kapiotis); **Faculty of Computer Science, University of Vienna, Vienna, Austria** (Prof E Neuwirth PhD)

Correspondence to:
Dr Christoph Buchta, Austrian Association for Quality Assurance and Standardization of Medical and Diagnostic Tests (ÖQUASTA), 1090 Vienna, Austria
christoph.buchta@oequasta.at

or

Dr Jeremy V Camp, Center for Virology, Medical University of Vienna, 1090 Vienna, Austria
jeremy.camp@meduniwien.ac.at

Research in context

Evidence before this study

We searched PubMed for articles published between Jan 1, 2020, and May 31, 2023, without language restrictions, using the search terms (((External Quality Assessment) OR (EQA) OR (Proficiency Testing) OR (PT) OR (interlaboratory comparison)) AND ((SARS-CoV-2) AND ((PCR) or (NAT) or (nucleic amplification))). We found 24 manuscripts reporting on single or a short series of large-scale external quality assessments (EQAs), typically with hundreds of medical laboratories and other test facilities participating at metropolitan, city-wide, country-wide, and worldwide scales. All of these studies clearly show that discordant results were obtained for samples with low virus concentrations, but overall performance was high. Although some studies compared assay formats (automated *vs* manual or commercial *vs* in-house) or laboratory type (public *vs* private), longitudinal data from EQA schemes measuring SARS-CoV-2 detection performance are absent. Such data would offer the possibility to test for improvement over time and for associations with laboratory type or assay format, and could be useful in the management of future epidemics, when regulations might be relaxed to enable an increase in testing while ensuring accurate epidemiological surveillance.

Added value of this study

This study retrospectively analysed comprehensive results from 14 rounds of three EQA schemes in the context of their relevance for epidemiological assessments and public health considerations. The EQA scheme in Austria is, to our knowledge, among the most intensive and responsive programmes in the world, and Austria was one of the world leaders in terms of the number of pathogen detection tests performed per 1000 inhabitants during the COVID-19 pandemic. As far as we are aware, this study is the first to analyse results from multiple EQA rounds with regard to false-negativity rates in the regional, widespread application of PCR analyses during the COVID-19 pandemic, accounting for virus concentration and investigating differences in performance due to laboratory type and assay format. Notably, these analyses show that false-negativity rates were 5·2%, increasing to 7·0% in some samples with low viral concentration. Furthermore, detection rates did not improve over time during 3 years of the pandemic. Because the EQA schemes evaluate analytical performance in near-to-real-life settings, the rate of undetected infections in the general population can be assumed to be similarly high.

Implications of all the available evidence

These data call for the development of appropriate quality control strategies for pathogen detection in preparation for future pandemics. Furthermore, the data highlight the benefit of comprehensive, regular EQA schemes that clearly define which parameters of analytical performance should be monitored considering the effect of the diagnostic setting (medical *vs* non-medical laboratory) on the quality of regional epidemiological data. The design of accompanying EQA schemes should include regular and timely reporting of the overall analytical performance data to public health authorities.

and report the results back to the EQA provider. EQA samples are required to be highly homogeneous and stable, such that all participating laboratories use identical samples and therefore have the same conditions for measurement or determination. The International Standard ISO 15189:2022, entitled Medical laboratories—requirements for quality and competence, requires EQA samples to mimic patient samples for clinically relevant challenges, and the EQA provider must select or prepare samples accordingly.[1]

The EQA provider assesses reported quantitative and qualitative results by comparison with the assigned target for each sample, and confidential feedback is provided to participating laboratories in an individual report. A summary report compares the results of each peer group (consisting of laboratories grouped by the test system used), describes the specifics of each round, and highlights overall areas for improvement where identified. Participation in EQA schemes enables laboratories to have the analytical performance of their methods confidentially evaluated by an independent third party, but also to compare their methods and procedures with those of other laboratories and, if necessary, to identify the need for improvement or other appropriate measures. Although participation in EQA schemes is of considerable self-interest to laboratories, there is also a legal obligation to do so, which, in Austria, is anchored in the Hospitals Act, the Physicians Act, the Quality Assurance Ordinance of the Medical Association, the Medical Devices Act, and the Epidemic Diseases Act.

In addition to the benefits for individual participating laboratories, EQA schemes also provide valuable data that, as a summary of aggregated results, allow an evaluation of the general analytical performance of individual assays, devices, and reagents. The results of several consecutive rounds can show performance over time. During a pandemic, when legal procedures for the approval of test procedures and for the qualification of test facilities and their staff can be suspended, EQA data should be relevant for public health authorities because of the unique ability of these data to provide an overview of the general analytical performance, and therefore the reliability, of the test results.[2] Aggregated results are therefore indicators of the quality of the data used in the epidemiological management of pandemics (eg, number of cases and incidence).

	SARS-CoV-2 RNA	Mean Ct (E gene)	Sample type and characteristics
1 (first round of SARS CoV-2 virus genome detection; May 18, 2020, 66 [99%] of 67 participants reporting results)			
1	Positive	27·8	Clinical sample, wild-type*
2	Positive	33·6	Clinical sample, wild-type*
3†	Positive	38·3	Clinical sample, weakly positive, wild-type*
4	Negative	..	..
2 (second round of SARS CoV-2 virus genome detection; Sept 30, 2020, 101 [96%] of 105 participants reporting results)			
1	Positive	28·9	Clinical sample, wild-type*
2	Negative	..	..
3	Positive	24·8	Clinical sample, wild-type*
4	Negative	..	..
5	Positive	36·3	Clinical sample, wild-type*
3 (third round of SARS-CoV-2 virus genome detection; Feb 1, 2021, 133 [90%] of 147 participants reporting results)			
1	Negative	..	Human DNA
2	Positive	32·2	Clinical sample, wild-type B.1.1.170
3	Positive	33·6	Clinical sample, 1:10 dilution of sample 6
4	Positive	35·7	Clinical sample, alpha variant B.1.1.7
5	Negative	..	..
6†	Positive	35·9	Clinical sample, wild-type B.1.1
7†	Positive	38·5	Clinical sample, 1:100 dilution of sample 6
4 (fourth round of SARS-CoV-2 virus genome detection; Aug 16, 2021, 139 [91%] of 153 participants reporting results)			
1	Negative	..	HCoV-OC43
2	Positive	34·5	Clinical sample, delta variant B.1.617.2 (same material as sample 6)
3	Positive	32·6	Clinical sample, alpha variant B.1.1.7
4	Positive	35·9	Clinical sample, delta variant B.1.617.2
5	Negative	..	Human DNA
6	Positive	33·8	Clinical sample, delta variant B.1.617.2 (same material as sample 2)
5 (fifth round of SARS-CoV-2 virus genome detection; Nov 16, 2021, 121 [91%] of 133 participants reporting results)			
1	Positive	23·9	Clinical sample§
2	Negative	..	Human DNA
3	Positive	32·5	Standard (~5000 copies per mL)‡
4	Positive	34·8	Standard (~1000 copies per mL)‡§
6 (first round of SARS-CoV-2 virus genome detection POCT; Jan 18, 2022, 16 [43%] of 37 participants reporting results)			
1	Positive	36·8	Standard (~1000 copies per mL)‡§
2	Negative	..	Human DNA
3	Positive	25·2	Clinical sample, omicron variant BA.1§
4	Positive	36·5	Standard (~1000 copies per mL)‡§

(Table 1 continues in next column)

(Continued from previous column)

	SARS-CoV-2 RNA	Mean Ct (E gene)	Sample type and characteristics
7 (sixth round of SARS-CoV-2 virus genome detection; Feb 21, 2022, 160 [94%] of 171 participants reporting results)			
1†	Positive	36·2	Standard (~100 copies per mL)‡
2	Positive	24·3	Clinical sample, omicron variant BA.2
3	Negative	..	Human DNA
4	Positive	34·6	Standard (~500 copies per mL)‡
5	Positive	33·5	Standard (~1000 copies per mL)‡
6†	Positive	37·9	Standard, weakly positive (50 copies per mL)‡
7	Positive	33·5	Standard (~1000 copies per mL)‡
8 (first round of SARS-CoV-2 virus genome detection II; May 16, 2022, 32 [94%] of 34 participants reporting results)			
1	Positive	35·4	Standard (~1000 copies per mL)‡
2	Positive	26·8	Clinical sample, omicron variant BA.2
3	Positive	32·8	Standard (~5000 copies per mL)‡
4	Negative	..	..
5	Positive	34·8	Standard (~1000 copies per mL)‡
9 (second round of SARS-CoV-2 virus genome detection POCT; May 16, 2022, 15 [79%] of 19 participants reporting results)			
1	Positive	36·7	Standard (~1000 copies per mL)‡
2	Positive	28·0	Clinical sample, omicron variant BA.5
3	Positive	34·4	Standard (~5000 copies per mL)‡
4	Negative	..	..
5	Positive	35·5	Standard (~1000 copies per mL)‡
10 (seventh round of SARS-CoV-2 virus genome detection; Aug 22, 2022, 161 [90%] of 178 participants reporting results)			
1	Positive	33·9	Standard (~1000 copies per mL)‡§
2	Positive	33·8	Standard (~1000 copies per mL)‡§
3	Negative	..	Human DNA
4	Positive	24·9	Clinical sample, omicron variant BA.4§
5†	Positive	37·5	Standard (~100 copies per mL)‡§
11 (third round of SARS-CoV-2 virus genome detection POCT; Aug 22, 2022, 11 [73%] of 15 participants reporting results)			
1	Positive	36·5	Standard (~1000 copies per mL)‡
2	Positive	36·6	Standard (~1000 copies per mL)‡
3	Negative	..	Human DNA
4	Positive	28·0	Clinical sample, omicron variant BA.4
5†	Positive	39·4	Standard (~100 copies per mL)‡
12 (second round of SARS CoV-2 virus genome detection II; Nov 8, 2022, 34 [92%] of 37 participants reporting results)			
1	Positive	33·8	Standard (~1000 copies per mL)‡
2	Positive	28·1	Clinical sample, omicron variant BA.5
3	Positive	27·7	Clinical sample, omicron variant BA.2.75
4	Negative	..	Human DNA
13 (fourth round of SARS-CoV-2 virus genome detection POCT; Jan 16, 2023, 28 [76%] of 37 participants reporting results)			
1¶	Positive	32·6	Virus culture, omicron variant BA.5§
2¶	Positive	32·7	Virus culture, omicron variant BA.5§
3¶	Positive	30·7	Virus culture, omicron variant BA.2§
4¶	Negative	..	Human DNA

(Table 1 continues on next page)

In Austria, the number of SARS-CoV-2 detection tests per 1000 inhabitants during the COVID-19 pandemic was one of the highest in the world.[3] The aim of this study was to retrospectively evaluate results from SARS-CoV-2 virus genome detection EQA schemes over 3 years of the pandemic, to infer the accuracy of epidemiological data and to identify areas for improvement.

	SARS-CoV-2 RNA	Mean Ct (E gene)	Sample type and characteristics
(Continued from previous page)			
14 (eighth round of SARS-CoV-2 virus genome detection; Feb 20, 2023, 128 [90%] of 142 participants reporting results)			
1¶	Positive	24·5	Virus culture, omicron variant BA.2§
2	Negative	..	Influenza A§
3¶	Negative	..	Human DNA
4¶	Positive	29·4	Virus culture, omicron variant BA.5§
5	Positive	28·9	RSV-B and SARS-CoV-2 omicron variant XBB§
6¶	Positive	26·0	Virus culture, omicron variant BA.5§

Rounds are numbered sequentially and are part of one of three EQA schemes: SARS-CoV-2 virus genome detection, SARS CoV-2 virus genome detection II, or SARS-CoV-2 virus genome detection POCT. Participants reporting results are defined as laboratories that were registered, received samples, and submitted a result for at least one sample. Ct=mean reported cycle threshold value. EQA=external quality assessment. POCT=point-of-care testing. *The clinical samples in rounds 1 and 2 are all designated as wild-type; more specific information is not available, although some samples might be of pre-alpha-variant B-like lineage (ie, with an Asp614Gly mutation in the spike protein). †An educational sample, for which a correct result was not required to pass the EQA round. ‡Standard was obtained from the AccuPlex SARS-CoV-2 Molecular Controls Kit (SeraCare; Milford, MA, USA) and is the ancestral (Wuhan-Hu-1) strain. §Human cell culture material was added to provide human housekeeping genes for assays that test and require their detection for a valid result. ¶Virus cultivation and preparation of EQA samples from cell-culture supernatants were conducted as previously described[5] with exceptions as outlined in the Methods.

Table 1: **Rounds of SARS-CoV-2 virus genome detection EQA schemes and samples used**

Methods

Study design

The EQA schemes in Austria were jointly operated by the Austrian Association for Quality Assurance and Standardization of Medical and Diagnostic Tests (ÖQUASTA) and the Center for Virology of the Medical University of Vienna (Vienna, Austria), which is the national reference laboratory for respiratory viruses. ÖQUASTA provides the technical infrastructure, communicates with the participants, distributes the sample materials, collects results, and conducts preparatory activities for the assessment, and the Center for Virology is responsible for the selection, procurement, production, and provision of sample materials and provides the expertise for assessing reported results.

Samples for the first round of SARS-CoV-2 virus genome detection were dispatched to 67 laboratories on May 18, 2020, and the deadline for reporting results was 1 week later.[4] Two rounds per year were planned and additional rounds were to be added according to the development of the pandemic and the evolution of the virus (table 1). In early 2022, several laboratories requested two additional SARS-CoV-2 detection EQA rounds per year and justified their need for closer monitoring for accreditation purposes, contractual obligations, or because of personal interest. These additional rounds, conducted in a scheme called SARS-CoV-2 virus genome detection II, started on May 16 and Nov 8, 2022. From Jan 18, 2022, participant pharmacies were gathered in an EQA scheme named SARS-CoV-2 virus genome detection point-of-care testing (POCT), as they needed higher scheduling flexibility and were allowed to exclusively use assays compliant with the requirements for test systems appropriate for POCT.[6]

Procedures

Samples were prepared from residual clinical samples, virus culture supernatant, or by dilution of a standard (AccuPlex SARS-CoV-2 Molecular Controls Kit; SeraCare, Milford, MA, USA), as described elsewhere.[4,5,7] Virus cultivation and preparation of EQA samples from cell-culture supernatants were conducted as previously described,[5] with the exception that the cell-culture medium used for virus dilution (Minimal Essential Medium) additionally contained 0·2 µg/mL human placental DNA (Sigma-Aldrich, D7011). The fetal calf serum concentration in the cell-culture medium was adjusted to 10% (v/v). The EQA samples were prepared in liquid form and were stored below –20°C (Zeichhardt H, Kammel M, GBD Gesellschaft fur Biotechnologische Diagnostik, Berlin, Germany, personal communication). Ethical approval was not required for the EQA rounds or for this study. The requirement of ISO 15189:2022 to mimic patient samples for clinically relevant challenges was met by using samples with relatively low viral load in addition to at least one negative and one clearly positive sample in each round and an overall coverage of mean cycle threshold (Ct) values of 23·9–39·4 over the 14 rounds (table 1).

Samples were intended to be either core or educational. Participant laboratories were required to submit correct results for core samples (all negative and 46 of a total of 53 samples positive for SARS-CoV-2) to pass the EQA round, and educational samples (seven positive samples) were used to provide more detailed information for participants and the EQA provider (table 1). From Jan 18, 2022, participants were required to submit scans or screenshots of raw data to verify that they carried out the analysis themselves with the specified test system. Feedback and results from all samples, including educational samples, were provided to participants in the general summary report and confidential individual reports for each round.

Participants were classified as medical laboratories (registered medical diagnostic laboratories, hospital diagnostic laboratories or special-care clinics, and microbiological or virological departments within university hospitals) or non-medical laboratories (blood banks, academic teaching or research laboratories, military and governmental laboratories, general practitioners and walk-in clinics, pharmacies, distributors or manufacturers of diagnostic tests, and laboratories dedicated solely to SARS-CoV-2 testing). In Austria, non-medical laboratories were authorised by the Epidemic Act to carry out SARS-CoV-2 detection tests, and we were interested to assess

whether they performed as well as laboratories that were experienced in routine clinical diagnostics.

Statistical analysis

Data from 14 rounds of three EQA schemes for SARS-CoV-2 virus genome detection were aggregated. The analysis is based on results reported by the participant laboratories: positive (detected); negative (not detected); or inconclusive, which included all other results reported by the test system—ie, not determinable, invalid, or error—and cases of participants submitting no result. Samples for which no result was submitted were considered inconclusive because the participant could resolve all other samples, could have requested new material, and would have been penalised for submitting an incorrect result but would not be penalised for submitting no result. The overall rates of false or true-negative, false or true-positive, and inconclusive results were calculated. Performance was assessed in two ways. First, the odds of getting a correct answer for virus-positive samples were modelled by mixed-effects logistic regression with laboratories as a random effect, reporting odds ratios (ORs) with 95% CIs to describe the association against four predictors: virus concentration, laboratory type, assay format or type, and time since the beginning of the COVID-19 pandemic. Because there were few false-positive results, we did not consider virus-negative samples. Therefore, the probability of a correct answer estimated by logistic regression is equal to the sensitivity (TP/[FN+TP]≈odds/[1+odds], where odds=TP/FN; TP=true-positive, FN=false-negative). The virus concentration per sample was the mean of the Ct values reported by the participants, and was treated as a continuous predictor. As we observed near-perfect performance (>99·9% sensitivity) for samples with Ct≤32, only samples with relatively low concentrations of viral RNA (set to Ct>32) were used for these statistical comparisons (appendix p 5). Second, inconclusive results were analysed with respect to these main comparisons using the same technique (mixed-effects logistic regression models) with the dummy variable inconclusive=1 and not-inconclusive=0. As we have previously established that within-laboratory variance explained a significant amount of variance in Ct values for SARS-CoV-2 genome detection EQAs,[8] we used mixed-effects logistic regression models, treating within-laboratory variance as a random effect.

The dataset was highly structured in several ways. A small subset of laboratories reported multiple test systems at irregular intervals across the many rounds (appendix p 2). We could not necessarily assume that participation within a round (and the missing values) could be treated as random, and we could not exclude the possibility of a reporting bias. Therefore, we conducted an analysis of sensitivity (hereafter termed robustness) on each mixed model by iteratively subsampling the data to include only one randomly selected assay per laboratory per round. As a further test of robustness, we analysed subsets of the data, using only core samples, including only rounds that contained more than 100 participants, or both. The statistical analyses were done in R (version 4.0.3) using software packages lme4 (version 1.1-29), MASS (version 7.3-53) and car (version 3.0-12) for mixed model analysis; data were handled with tidyverse (version 1.3.1); and graphs were created using ggplot2 (version 3.3.5), ggsci (version 2.9), and ggpubr (version 0.4.0).

Role of the funding source

There was no funding source for this study.

Results

191 individual Austrian laboratories participated in SARS-CoV-2 detection EQA schemes between May 18, 2020 and Feb 20, 2023 (appendix p 3), classified as 102 medical laboratories and 89 non-medical laboratories. 63 (32·9%) of these laboratories were newly registered with ÖQUASTA as they had not previously participated in any of their EQA schemes. In these 191 laboratories, 42 different devices and 60 different reagents were used, in addition to more than 20 unspecified in-house assays. In addition to the in-house methods, 23 reagents and 17 devices were each used by only one laboratory, and, of these, eight reagents and four devices were used in only one round (appendix pp 6–8). 4663 results were reported for the 53 positive samples that were used in the 14 rounds of the EQA schemes; among them, 4371 (93·7%) were true-positive and 241 (5·2%) were false-negative. Among the 1645 results reported for the 18 samples that were negative for SARS-CoV-2 RNA, 1561 (94·9%) were true-negative, and ten (0·6%) were falsely reported as positive. 51 (1·1%) of 4663 results for positive samples and 74 (4·5%) of 1645 results for negative samples were reported as inconclusive (table 2).

The probability of reporting a correct result decreased with the estimated concentration of the sample; the odds

See Online for appendix

	N	True-positive	False-negative	False-positive	True-negative	Inconclusive
Overall results						
Positive	4663	4371 (93·7%)	241 (5·2%)	..	..	51 (1·1%)
Positive (Ct ≤32)	1391	1383 (99·4%)	3 (0·2%)	..	..	5 (0·4%)
Positive (Ct >32)	3272	2988 (91·3%)	238 (7·3%)	..	..	46 (1·4%)
Positive educational	823	659 (80·1%)	144 (17·5%)	..	..	20 (2·4%)
Negative	1645	..	..	10 (0·6%)	1561 (94·9%)	74 (4·5%)
Early *vs* late rounds						
Early rounds (2020 and 2021)						
Positive	2085	1984 (95·2%)	74 (3·5%)	..	..	27 (1·3%)
Negative	933	..	..	6 (0·6%)	885 (94·9%)	42 (4·5%)
Late rounds (2022 and 2023)						
Positive	2578	2387 (92·6%)	167 (6·5%)	..	..	24 (0·9%)
Negative	712	..	..	4 (0·6%)	676 (94·9%)	32 (4·5%)

(Table 2 continues on next page)

	N	True-positive	False-negative	False-positive	True-negative	Inconclusive
(Continued from previous page)						
Laboratory type						
Medical						
Positive	3104	2966 (95·6%)	116 (3·7%)	..	..	22 (0·7%)
Positive (Ct ≤32)	939	937 (99·8%)	0 (0%)	..	..	2 (0·2%)
Positive (Ct >32)	2165	2029 (93·7%)	116 (5·4%)	..	..	20 (0·9%)
Positive educational	570	487 (85·4%)	73 (12·8%)	..	..	10 (1·8%)
Negative	1107	..	..	1 (0·1%)	1090 (98·5%)	16 (1·4%)
Non-medical						
Positive	1559	1405 (90·1%)	125 (8·0%)	..	..	29 (1·9%)
Positive (Ct ≤32)	452	446 (98·7%)	3 (0·7%)	..	..	3 (0·7%)
Positive (Ct >32)	1107	959 (86·7%)	122 (11·0%)	..	..	26 (2·3%)
Positive educational	253	172 (68·0%)	71 (28·1%)	..	..	10 (4·0%)
Negative	538	..	..	9 (1·7%)	471 (87·5%)	58 (10·8%)
Assay type						
Automated						
Positive	2653	2544 (95·9%)	77 (2·9%)	..	..	32 (1·2%)
Negative	935	..	..	4 (0·4%)	907 (97·0%)	24 (2·6%)
Manual commercial	..	..	..	..	..	..
Positive	1892	1717 (90·8%)	157 (8·3%)	..	..	18 (1·0%)
Negative	664	..	..	6 (0·9%)	612 (92·2%)	46 (6·9%)
Manual in-house	..	..	..	..	..	..
Positive	118	110 (93·2%)	7 (5·9%)	..	..	1 (0·8%)
Negative	46	..	..	0 (0%)	42 (91·3%)	4 (8·7%)
Medical laboratories						
Automated assays						
Positive	2265	2219 (98·0%)	36 (1·6%)	..	..	10 (0·4%)
Negative	801	..	..	0 (0%)	793 (99·0%)	8 (1·0%)
Manual commercial	..	..	..	..	..	..
Positive	769	683 (88·8%)	74 (9·6%)	..	..	12 (1·6%)
Negative	282	..	..	1 (0·4%)	274 (97·2%)	7 (2·5%)
Manual in-house	..	..	..	..	..	..
Positive	70	64 (91·4%)	6 (8·6%)	..	..	0 (0%)
Negative	24	..	..	0 (0%)	23 (95·8%)	1 (4·2%)
Non-medical laboratories						
Automated assays						
Positive	388	325 (83·3%)	41 (10·6%)	..	..	22 (5·7%)
Negative	134	..	..	4 (3·0%)	114 (85·1%)	16 (11·9%)
Manual commercial						
Positive	1123	1034 (92·1%)	83 (7·4%)	..	..	6 (0·5%)
Negative	382	..	..	5 (1·3%)	338 (88·5%)	39 (10·2%)
Manual in-house						
Positive	48	46 (95·8%)	1 (2·1%)	..	..	1 (2·1%)
Negative	22	..	..	0 (0%)	19 (86·4%)	3 (13·6%)

Ct=cycle threshold. EQA=external quality assessment.

Table 2: **Results reported in SARS-CoV-2 genome detection EQA schemes**

of a correct result decreased by 0·64 (95% CI 0·59–0·69) for every unit increase in Ct value (p<0·0001; figure 1A). Visually, the logistic curve did not appear to fit low-concentration samples, and performance decreased below a particular concentration. For samples with Ct>32 (mean 35·4, SD 1·8, range 32·2–39·4; appendix p 5), the mean sensitivity per sample was 92·5% (SD 9·5; range 59·1–100·0), whereas it was 99·7% (SD<0·9; range 99·4–100·0) for all other samples (Mann-Whitney U test, p<0·0001; table 2). We therefore focused on low-concentration samples (Ct>32) to assess factors that influenced performance (appendix p 5).

We compared performance in rounds that took place in 2020–21 versus those in 2022–23 (table 2). The mean sensitivity for earlier rounds (95·4%, SD 9·8, range 59·1–100·0) was similar to that of later rounds (94·5%, 7·9, 72·7–100·0). However, the odds of a correct response in earlier rounds was 2·76 times greater than the odds of a correct response in later rounds (95% CI 2·00–3·83; p<0·0001 for samples with Ct>32). This difference was similar when only testing rounds in which more than 100 results per sample were submitted (OR 3·29 [95% CI 2·37–4.61], p<0·0001) and when conducting robustness analysis (2·74 [1·64–4·49], p<0·0001; appendix p 10). However, when stratifying by year, performance was significantly better in 2021 and 2023 than in 2020 (p<0·0001), with an increase in the odds of correctly identifying a sample as positive or negative of 2·80 (1·75–4·50) times in 2021 and 4·53 (2·32–9·52) times in 2023, relative to 2020. No difference was observed between sensitivity in 2022 versus 2020 (OR 0·72 [0·48–1·07]). When focusing only on core samples, we found no difference in performance between any year: the mean sensitivities were 97·4% (SD 4·3) for 2020, 97·9% (3·5, p=0·87 *vs* 2020) for 2021, 97·3% (3·3, p=0·58) for 2022, and 95·8% (7·2, p=0·069) for 2023 (figure 1B).

Medical laboratories (mean sensitivity 96·5%, SD 6·4, range 69·8–100·0) performed better than non-medical laboratories (92·9%, SD 12·6, 39·1–100·0). For samples with Ct>32, the odds of a medical laboratory correctly reporting a sample were 2·94 [95% CI 1·75–4·96] times higher than those of a non-medical laboratory (p<0·0001; table 2, figure 1C). An analysis of model robustness showed some variability around this value, but the effect was always significant: OR 3·84 (2·16–6·85) for the rounds with more than 100 participants, 5·66 (1·64–19·47) when considering only core samples, and between 3·01 and 3·70 (mean 3·35, range of 95% CIs 1·74–6·53) when iterating over randomly selected results (appendix p 10).

Only six laboratories reported laboratory-developed (also known as in-house) tests, in nine of the 14 rounds. One participating laboratory reported results from two different laboratory-developed tests in each of these nine rounds. Among the five laboratories reporting results from one laboratory-developed test each, one participated in five rounds, three in two rounds, and one in one round. With one exception, all laboratory-developed tests had 100% sensitivity. One laboratory that was using

two different laboratory-developed tests reported three of four positive samples as negative in one round (sample Cts 33·9, 33·8, and 37·5), but correctly reported the fourth positive sample (Ct 24·9) in that round.

Fully automated commercial test systems (mean sensitivity 97·4%, SD 3·9, range 84·7–100·0) performed better than test systems requiring one or more manual steps (91·1%, 13·9, 48·9–100·0), when considering rounds in which more than 17 results were reported per group. The odds of an automated test system correctly reporting a sample were 4·60 (95% CI 2·91–7·41) times greater than for a manual test system (p<0·0001), and even greater (6·30 [3·83–10·37]) when considering only rounds in which more than 100 test systems were reported per sample (p<0·0001; figure 1D). Testing model robustness revealed the OR to be between 3·87 and 6·50 (mean 4·99, range of 95% CIs 2·26–12·03; appendix p 10).

The two most frequently used test systems were the Roche cobas assays (either the cobas SARS-CoV-2 and Influenza A/B multiplex or the cobas SARS-CoV-2 using the Roche cobas 6800 device) and the Cepheid GeneXpert Xpress assays (SARS-CoV-2 or the SARS-CoV-2/Flu/RSV multiplex), both of which are fully automated test systems. These test systems were used by 82 of the 191 registered laboratories; 67 participating laboratories used one of these two test systems exclusively. These two test systems combined had a mean sensitivity of 99·5% per sample over all samples (SD 1·5, range 92·2–100·0)—this was significantly better than for all other test systems combined (93·8%, 11·3, 50·0–100·0), with 15·1 (95% CI 7·76–29·31) times greater odds of a correct response (p<0·0001).

3048 (73·0%) of 4173 results reported by medical laboratories were obtained by automated assays, 1032 (24·7%) by manual commercial assays, and 93 (2·2%) by manual in-house assays. Although medical laboratories (mean sensitivity 96·5% per sample) performed better than non-medical laboratories (92·9%), and automated assays (97·4%) performed better than manual assays (91·1%), the best performances overall were for medical laboratories using automated assays, with a mean sensitivity of 98·6% (SD 3·2). The odds of this combination reporting a correct answer were 4·93 (95% CI 3·29–7·37) times greater than for non-medical laboratories using manual assays, 6·88 (4·56–10·39) times greater than for medical laboratories using manual assays, and 7·51 (4·68–12·03) times greater than for non-medical laboratories using automated assays. In a statistical model that included laboratory type, assay format, and the interaction term, the coefficients for laboratory type and the interaction term were significantly associated with the probability of reporting a correct answer (p<0·0001), but the assay format coefficient was not (p=0·64; similar results were obtained from tests of model robustness; appendix p 11). This finding was explained by the fact that, although automated systems (mean sensitivity 98·6%, SD 3·2) performed better than manual systems (90·2%, 13·5) in medical laboratories, manual systems (94·2%, 18·5) outperformed automated systems (89·5%, 12·9) in non-medical laboratories (table 2, appendix p 9).

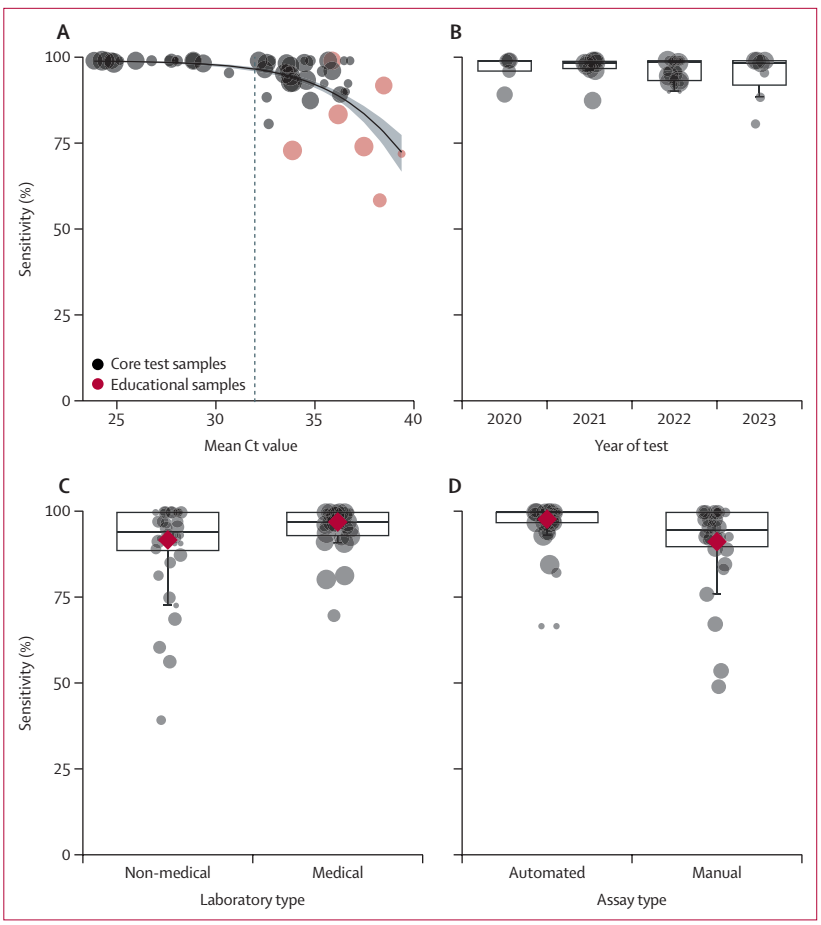

Figure 1: **Sensitivity per sample for positive samples in multiple rounds of SARS-CoV-2 genomic detection EQA schemes**
(A) The relationship between sensitivity (observed positive results/total expected positive results) and sample concentration measured by the reference laboratory as a Ct value. The predicted probability of detecting a sample is shown as a logistic curve with 95% CI. The vertical dashed line shows an arbitrary cutoff at Ct 32, to the right of which samples are considered to be low-concentration. (B) Box plots showing the sensitivity per sample over 4 years of EQA schemes for core samples. (C) Sensitivity per sample for low-concentration samples measured by medical laboratories versus non-medical laboratories. (D) Sensitivity per sample for low-concentration samples measured by fully automated test systems versus assay formats requiring manual intervention. In all panels, the size of the dots is relative to the number of reported results per sample. For boxplots, the thick line represents the median, and the upper and lower bounds of the box show the interquartile range. In panels (C) and (D), the red diamonds show the predicted probability of a correct response in each group based on a mixed effects logistic regression model. Ct=cycle threshold. EQA=external quality assessment.

Overall, inconclusive results were reported for 51 (1·1%) of 4663 positive samples and 74 (4·5%) of 1645 negative samples. A small but significant association was found between inconclusive results and sample composition, with the odds of reporting an inconclusive result increasing 1·23 times (95% CI 1·11–1·37) for each unit increase in Ct value. Notably, greater than 10% of results were submitted as inconclusive for three negative samples in which human DNA was not included (ie, the sample contained

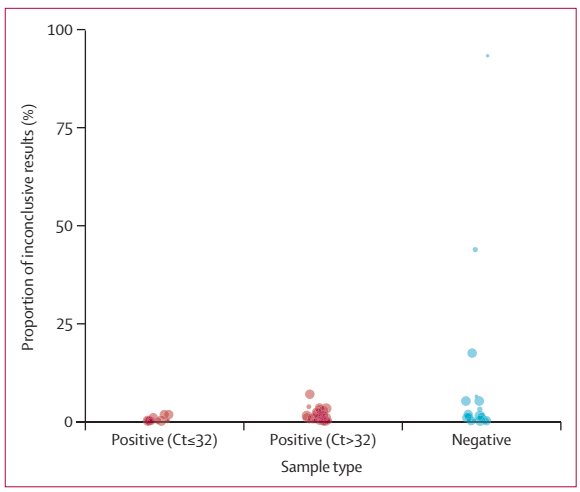

Figure 2: **Percentage of inconclusive results stratified by sample type**
Samples positive for SARS-CoV-2 are shown with red dots and negative samples with blue dots; the size of the dots is relative to the number of reported results per sample. Results were defined as inconclusive if they were reported as not determinable, invalid, or error, or if the participant laboratory submitted no result for the sample but did submit results for other samples in the round. Negative samples with greater than 10% inconclusive results are those that did not contain spiked human DNA.

only an NaCl solution). Excluding these two rounds (10 and 11; table 1), the odds of obtaining an inconclusive result was 4·73 (95% CI 3·21–6·96) times higher for negative samples than for positive samples (p<0·0001; figure 2); testing model robustness revealed the OR to be between 4·34 and 5·07 (range of 95% CIs 2·88–7·62; appendix p 10). Similarly, the odds of submitting an inconclusive result were higher for non-medical laboratories than for medical laboratories (OR 5·92 [95% CI 3·04–11·51]), for manual assay formats than for automated assays (1·87 [1·06–3·27]), and when a test system other than the two most common systems was used (10·20 [4·11–25·28]).

Discussion

The aim of this study was to evaluate laboratory performance for the detection of SARS-CoV-2 by qPCR over the course of the COVID-19 pandemic. The most notable findings were the high rate of false-negative results in low-concentration samples compared with high-concentration samples and the fact that this rate did not decrease significantly during the pandemic. Additionally, we identified that laboratory type and assay format were key factors that described performance, with medical laboratories performing better than non-medical laboratories, and automated assays performing better than assays that require one or more manual interventions.

As expected, the rates of false-negative results were highest in samples with relatively low virus concentration (Ct>32 in our analysis). This finding is potentially important given that approximately 25% of samples from individuals who are infected with SARS-CoV-2 but are asymptomatic have initial Ct values greater than 30. We can conclude that the under-reporting of cases in Austria was considerable.[9–11] The actual rate of false-negative results could have been even higher, due to limitations in analytical and pre-analytical procedures. Pre-analytical procedures that affect overall outcome could include incorrect sampling and inappropriate transport conditions, in addition to pooling samples for analysis to increase testing capacity, which potentially dilutes positive samples with excess negative samples.[12–15] In such cases, it is only certain when a sample (or pool) is virus-positive; negative test results are best interpreted as not detected. With EQAs, performances can be evaluated with a reasonable expectation that sampling errors—and, to an extent, transport conditions—are being controlled, and therefore incorrect results must arise from other pre-analytical or analytical procedures.

We note that the performance of non-medical laboratories was worse than that of medical laboratories, and that performance was also associated with assay format. The better performance of medical laboratories could be explained, at least in part, by the test systems used, as these laboratories mostly used automated methods. However, why non-medical laboratories performed better when using methods that required manual intervention is not clear. Our conclusions are limited principally by a lack of detailed information about the participant laboratories, their relative experience levels, and details about the specific analytical procedures they used (eg, were all manufacturers' recommendations strictly followed?). We could not account for laboratories using EQAs to test new assays; similarly, we could not account for the batch-to-batch variability of assays, which laboratories are required to validate outside of an EQA scheme.

Nonetheless, data from our EQA schemes suggest that the number of individuals falsely diagnosed as positive for SARS-CoV-2 infection was low and likely to be negligible. We also noted that inconclusive results were significantly higher in samples that were negative for SARS-CoV-2—particularly samples that did not contain internal control material required for assay validation. Inconclusive results do not indicate excellent performance of a test system, but are still preferable to false-negative or false-positive results.

Various assays were used for routine SARS-CoV-2 testing and in EQA participation, and these assays varied in analytical performance. Not only established manufacturers and distributors of diagnostic tests but also non-medical laboratories that were not previously experienced in routine clinical diagnostics promoted and provided their test systems for SARS-CoV-2 detection. If newly developed assays for the detection of recently emerged infectious agents initially underperformed, it might be expected that these assays would be improved by manufacturers to enable them to compete in the

market. The better-performing assays would then prevail, and overall performance would therefore increase continuously. However, our analysis of aggregated EQA results showed no change in performance over time, even though assays for laboratory and point-of-care-testing applications that consistently demonstrated excellent performance have been available since the first round of the EQA scheme—two of which were the most frequently used assays among participating laboratories. Owing to unprecedented shortages of reagents and consumables, it is possible that participating laboratories could not switch to better-performing assays, or were reluctant to do so, even if feedback from participation in the EQAs indicated that their assay of choice had a low performance.

The high number of initial enrolments at ÖQUASTA through registrations for SARS-CoV-2-associated EQAs shows how widely SARS-CoV-2 testing was conducted in Austria and that testing was not limited to medical laboratories and institutions. Our findings provide a picture of the general performance of a wide variety of laboratories conducting large-scale diagnostics during a pandemic, and reveal patterns that suggest how performance (and quality assurance) could be improved in the future (appendix p 12). The responsibilities for improved implementation of diagnostic testing during outbreaks are shared by laboratories, public health authorities, regulatory authorities, legislators, manufacturers of in-vitro diagnostics, and the EQA provider. Furthermore, we refer to recommendations on the design of EQA schemes and the roles of their providers in future epidemics.[16] These recommendations offer best practices for monitoring the performance of pandemic-associated analytical procedures to provide relevant information to public health authorities. Most importantly, our analysis makes a strong case for the importance of EQAs during a pandemic, and lack of consideration of them should be added to the failures listed in the *Lancet* Commission's report on lessons for the future from the COVID-19 pandemic.[17]

Contributors

CB conceptualised the study. CB, SWA, IG, EN, LW, and JVC developed the methodology. CB curated the data and planned and organised the schemes. CB, WHu, EN, and JVC conducted the formal analysis. SWA, AG, and EP-S contributed resources. SWA, FA, BB, IG, AG, WHü, SK, MMM, EN, EP-S, LW, and JVC supervised the project. CB was responsible for project administration. CB wrote the initial draft of the manuscript, which was reviewed and edited by all other authors. All authors had full access to all the data in the study and accept responsibility for the decision to submit for publication. The underlying data were verified by CB and JVC.

Declaration of interests

We declare no competing interests.

Data sharing

The full dataset may be provided upon reasonable request to the corresponding authors.

References

1 ISO 15189. Medical laboratories—requirements for quality and competence. Geneva: International Organization for Standardization, 2022.

2 Buchta C, Zeichhardt H, Griesmacher A, Schellenberg I, Kammel M. Ignoring SARS-CoV-2 testing performance during COVID-19. *Lancet Microbe* 2023; **4:** e296.

3 TheGlobalEconomy.com. Covid total tests per thousand by country: the latest data. https://www.theglobaleconomy.com/rankings/covid_total_tests_per_thousand/ (accessed Feb 23, 2023).

4 Görzer I, Buchta C, Chiba P, et al. First results of a national external quality assessment scheme for the detection of SARS-CoV-2 genome sequences. *J Clin Virol* 2020; **129:** 104537.

5 Vierbaum L, Wojtalewicz N, Grunert HP, et al. RNA reference materials with defined viral RNA loads of SARS-CoV-2—a useful tool towards a better PCR assay harmonization. *PLoS One* 2022; **17:** e0262656.

6 Buchta C, Zeichhardt H, Badrick T, et al. Classification of "near-patient" and "point-of-care" SARS-CoV-2 nucleic acid amplification test systems and a first approach to evaluate their analytical independence of operator activities. *J Clin Virol* 2023; **165:** 105521.

7 Buchta C, Camp JV, Jovanovic J, et al. The versatility of external quality assessment for the surveillance of laboratory and in vitro diagnostic performance: SARS-CoV-2 viral genome detection in Austria. *Clin Chem Lab Med* 2021; **59:** 1735–44.

8 Buchta C, Kollros D, Jovanovic J, et al. Converting to an international unit system improves harmonization of results for SARS-CoV-2 quantification: results from multiple external quality assessments. *J Clin Virol* 2023; **158:** 105352.

9 Glenet M, Lebreil AL, Heng L, N'Guyen Y, Meyer I, Andreoletti L. Asymptomatic COVID-19 adult outpatients identified as significant viable SARS-CoV-2 shedders. *Sci Rep* 2021; **11:** 20615.

10 Platten M, Hoffmann D, Grosser R, et al. SARS-CoV-2, Ct-values, and infectivity—conclusions to be drawn from side observations. *Viruses* 2021; **13:** 1459.

11 Wu Q, Shi L, Li H, et al. Viral RNA load in symptomatic and asymptomatic COVID-19 omicron variant-positive patients. *Can Respir J* 2022; **2022:** 5460400.

12 Richard-Greenblatt M, Ziegler MJ, Bromberg V, et al. Quantifying the impact of nasopharyngeal specimen quality on severe acute respiratory syndrome coronavirus 2 test performance. *Open Forum Infect Dis* 2021; **8:** ofab235.

13 Buchta C, Camp JV, Jovanovic J, et al. Results of a SARS-CoV-2 virus genome detection external quality assessment round focusing on sensitivity of assays and pooling of samples. *Clin Chem Lab Med* 2022; **60:** 1308–12.

14 Mulu A, Alemayehu DH, Alemu F, et al. Evaluation of sample pooling for screening of SARS CoV-2. *PLoS One* 2021; **16:** e0247767.

15 Torres I, Albert E, Navarro D. Pooling of nasopharyngeal swab specimens for SARS-CoV-2 detection by RT-PCR. *J Med Virol* 2020; **92:** 2306–07.

16 Buchta C, Zeichhardt H, Aberle SW, et al. Design of external quality assessment schemes and definition of the roles of their providers in future epidemics. *Lancet Microbe* 2023; **4:** e552–62.

17 Sachs JD, Karim SSA, Aknin L, et al. The *Lancet* Commission on lessons for the future from the COVID-19 pandemic. *Lancet* 2022; **400:** 1224–80.