Evaluation of factors contributing to variability of qualitative and quantitative proficiency testing for SARS-CoV-2 nucleic acid detection

Yongzhuo Zhang [a,1], Xia Wang [a,1], Chunyan Niu [a], Di Wang [a], Qingfei Shen [a], Yunhua Gao [a], Haiwei Zhou [b], Yujing Zhang [a], Yan Zhang [a], Lianhua Dong [a,*]

[a] National Institute of Metrology, Beijing 100029, China
[b] Division II of In Vitro Diagnostics for Infectious Diseases, Institute for In Vitro Diagnostics Control, National Institutes for Food and Drug Control, Beijing 100050, China

ARTICLE INFO

ABSTRACT

The pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to unprecedented social and economic disruption. Many nucleic acid testing (NAT) laboratories in China have been established to control the epidemic better. This proficiency testing (PT) aims to evaluate the participants' performance in qualitative and quantitative SARS-CoV-2 NAT and to explore the factors that contribute to differences in detection capabilities. Two different concentrations of RNA samples (A, B) were used for quantitative PT. Pseudovirus samples D, E (different concentrations) and negative sample (F) were used for qualitative PT. 50 data sets were reported for qualitative PT, of which 74.00% were entirely correct for all samples. Forty-two laboratories participated in the quantitative PT. 37 submitted all gene results, of which only 56.76% were satisfactory. For qualitative detection, it is suggested that laboratories should strengthen personnel training, select qualified detection kits, and reduce cross-contamination to improve detection accuracy. For quantitative detection, the results of the reverse transcription digital PCR (RT-dPCR) method were more comparable and reliable than those of reverse transcription quantitative PCR (RT-qPCR). The copy number concentration of ORF1ab and N in samples A and B scattered in 85, 223, 50, and 106 folds, respectively. The differences in the quantitative result of RT-qPCR was mainly caused by the non-standard use of reference materials and the lack of personnel operating skills. Comparing the satisfaction of participants participating in both quantitative and qualitative proficiency testing, 95.65% of the laboratories with satisfactory quantitative results also judged the qualitative results correctly, while 85.71% of the laboratories with unsatisfactory quantitative results were also unsatisfied with their qualitative judgments. Therefore, the quantitative ability is the basis of qualitative judgment. Overall, participants from hospitals reported more satisfactory results than those from enterprises and universities. Therefore, surveillance, daily qualitiy control and standardized operating procedures should be strengthened to improve the capability of SARS-CoV-2 NAT.

1. Introduction

As of July 2022, over 568.77 million confirmed cases and over 6.38 million deaths caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been reported by WHO [1–3]. Now, SARS-CoV-2 is constantly mutating, and the rapid propagation and concealment of mutants such as Omicron [4], Alpha, and Beta [5] have made epidemic prevention and control extremely difficult [6,7]. In addition, the rapid spread of the virus has led to unprecedented social and economic disruption, with the government ordering schools and businesses to close.

Apart from epidemiologic clustering and virus-specific diagnostic testing, there is little clinically that distinguishes SARS-CoV-2 infection from other severe viral respiratory infections such as severe acute respiratory syndrome or influenza [8]. Nucleic acid testing (NAT) is the preferred method for detecting SARS-CoV-2 [9]. Suitable specimens for this testing include nasopharyngeal swabs, oropharyngeal swabs, nasal washes, and nasal aspirates [10,11]. This test methodology, also considered the gold standard of testing, has been the mainstay of COVID-19 diagnosis in the world [12]. NAT is a complex procedure, and its accuracy depends on well-defined laboratory practices and quality systems [13]. Under the continuous spread of the epidemic, the urgent and massive demand for testing led to rapid development and validation of commercial and laboratory-developed assays to

HIGHLIGHTS

Scientific question
The COVID-19 epidemic continues to ravage the world. Therefore, various influencing factors in SARS-CoV-2 NAT and the significance of quantitative and qualitative detection capabilities deserve to be discussed, and several factors for improving testing capacity are proposed to provide guidance for containing the spread of the outbreak.

Evidence before this study
Some institutions have carried out SARS-CoV-2 qualitative proficiency testing for domestic testing laboratories, and have raised issues such as insufficient overall testing capabilities and lack of personnel training. But there has been little discussion of quantitative capabilities and regulatory regimes.

New findings
For quantitative capabilities, dPCR methods are more comparable and reliable, while qPCR results are easily affected by the use of standard materials and personnel operation. Quantitative competence is the basis for qualitative judgment. Participants from hospitals reported more satisfactory outcomes, while participants from businesses were less satisfied. The systems such as "CE" certification based on "self-declaration" are temporarily not suitable for China, and national supervision and the development of various verification projects play an important role in improving nucleic acid detection capabilities.

Significance of the study
Not all laboratories have satisfactory results with SARS-CoV-2 testing, especially those from ICL. It is necessary to further improve its quality system, including selection of qualified testing kits, personnel training, standardized operations, daily quality control, and regular external assessments. Improving the overall testing capacity of the laboratory can greatly avoid the occurrence of testing errors and provide help for epidemic prevention and control.

detect SARS-CoV-2. Testing laboratories should accurately and promptly identify SARS-CoV-2 by their established methods. However, given the massive number of laboratories doing SARS-CoV-2 testing using different methods and kits, reliability and accuracy of results have always been a concern. Accurate results of SARS-CoV-2 testing are used for patient management, infection control in health care settings, and surveillance data that drive decisions on community-wide sheltering orders. Therefore, the accuracy of the testing results is the cornerstone of epidemic prevention and control.

Globally, it is almost impossible to deploy such a wide range of testing capabilities through administrative means and ensure the accuracy of testing results. Therefore, an objective evaluation of system-wide product and laboratory measurement capability is needed. Proficiency testing (PT) is an essential tool for monitoring the diagnostic proficiency of laboratories and providing results that allow the implementation of improved testing [14]. It is the procedure in which a government agency or professional institution sends identical samples to participating laboratories for blind analyses, and the results are then compiled and used as the basis for judging the testing ability of individual laboratories [15,16]. In this study, we designed a SARS-CoV-2 quantitative PT in parallel with a qualitative PT to evaluate the factors

affecting the performance of laboratory testing to provide information and guidance for the current clinical application of SARS-CoV-2 testing. Challenges and experience with implementing inter-laboratory comparison across a vast network of COVID-19 testing laboratories are also discussed.

2. Methods

2.1. Participants

The PT for SARS-CoV-2 detection was initiated and implemented by The National Institute of Metrology of China (NIM). Forty-eight laboratories, mainly those engaged in SARS-CoV-2 nucleic acid testing, registered for qualitative PT, and 42 of them registered for quantitative PT. These participants include laboratories from hospitals, academics, and enterprises. For the confidentiality of the PT, the laboratories were coded in sequence from NA01-001 to NA01-048. All participants were distributed in 14 provinces and cities in China (Supplementary Fig. S1).

2.2. Samples used in the PT

This PT consists of two parts: qualitative detection and quantitative measurement. The samples used are divided into two types to satisfy different application directions: purified RNA extracted from inactivated SARS-CoV-2, including samples A and B, and pseudovirus obtained by transfection cells by segmental construction of novel coronavirus sequences onto lentivirus vectors including sample D and E [17] (ORF1ab was longer in total length, it consisted of pseudoviruses of different segments). Pseudovirus samples were designed to evaluate the detection, including the extraction step.

Samples A and B were used for quantitative measurement, while samples D, E, and F were used for qualitative detection. Samples A and B were SARS-CoV-2 RNA samples with different concentrations. Samples D and E were diluted from a pseudovirus RM (NIM-RM5207), and sample F was a negative control. Two vials of each sample were distributed to all the participants. The reference values are listed in Table 1.

2.3. PT procedure

Since the primary purpose of the proficiency testing project is to examine the comprehensive testing ability of the laboratory. Participating laboratories can choose different methods, such as digital PCR (RT-dPCR) [18] or real-time quantitative PCR (RT-qPCR) [19], and use a kit commonly used in the laboratory or laboratory-developed tests. Samples D and E should be extracted before detection, and the extraction method should also be selected independently. The operating protocol referred to the manufacturer's instructions.

When the RT-dPCR method was used, the PCR reaction was prepared according to the manufacturer's instructions. Briefly, 5 μL of the sample was added to 15 μL of PCR reaction solution, and the PCR reaction mixture was transferred to generate droplets and amplified. Three replicated RT-dPCR measurements were required for each sample. The copy number concentration of ORF1ab and N were reported for samples A and B. Positive, or adverse results were reported for samples D, E, and F.

When using RT-qPCR, 5 μL of the sample was added to 20 μL of PCR reaction solution for amplification. The qualitative results of samples D, E, and F were directly judged by CT value according to the instruction of the testing kits. The standard curve should first be established for quantitative results using a serial dilution of certified SARS-CoV-2 reference material. Then test according to the instructions. Three replicated measurements were conducted for each vial. The copy number concentration of ORF1ab and N in samples A and B

Table 1
The information of panel samples.

Category	Sample name	Type	Copy number concentration ± expanded uncertainty (Copies/mL, $k = 2$)	
			ORF1ab Gene[##]	N Gene
Quantitative PT	A	SARS-CoV-2 Genomic RNA[*]	$(2.60 \pm 0.28) \times 10^5$	$(6.67 \pm 0.80) \times 10^5$
	B	SARS-CoV-2 Genomic RNA[**]	$(7.00 \pm 2.00) \times 10^3$	$(2.50 \pm 1.00) \times 10^4$
Qualitative PT	D	Pseudovirus[#]	(266–2,719) $(7.55 \pm 2.50) \times 10^2$	$(5.91 \pm 2.81) \times 10^2$
			(2,720–8,554) $(7.45 \pm 2.30) \times 10^2$	
			(8,555–13,024) $(9.09 \pm 0.28) \times 10^2$	
			(13,025–18,039) $(1.15 \pm 0.34) \times 10^3$	
			(18,040–21,552) $(4.91 \pm 1.50) \times 10^2$	
	E	Pseudovirus[#]	(266–2719) $(3.77 \pm 2.50) \times 10^2$	$(2.95 \pm 1.41) \times 10^2$
			(2,720–8,554) $(3.73 \pm 2.30) \times 10^2$	
			(8,555–13,024) $(4.55 \pm 0.28) \times 10^2$	
			(13,025–18,039) $(5.73 \pm 0.34) \times 10^2$	
			(18,040–21,552) $(2.45 \pm 1.50) \times 10^2$	
	F	TE buffer	/	/

[*] Sample A was certified reference materials GBW (E) 091,099 (This reference material was the purified RNA genome of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), including the entire length of ORF1ab, N, and E genes. For details, please refer to supplemental File S2, or visit https://www.ncrm.org.cn/Web/Ordering/MaterialDetail?autoID=20123).

[**] Sample B was diluted 30 times from GBW (E) 091099.

[#] Sample D and E were diluted from NIM-RM5207 in different batches, and D was diluted 110 times, E diluted 220 times.

[##] ORF1ab gene in samples D and E includes different segments with different concentrations. The final result is judged according to different fragments; refer to the certificate of NIM-RM5207 in supplemental File S1, or visit https://www.ncrm.org.cn/Web/Ordering/MaterialDetail?autoID=23949.

Table 2
z′-score evaluation criteria* .

z′-score	Judgment
$\| z' \| \leq 2$	Satisfactory result
$2 < \| z' \| < 3$	Questionable result (Warning signal)
$\| z' \| \geq 3$	Unsatisfactory result (Action signal)

* When the result of the participant is not used for the determination of the assigned value, the *z′-score* can be calculated.

can be calculated according to its corresponding target's standard curve.

2.4. Data analysis

For qualitative PT, each participant was assessed according to the consistency of sample attributes. The quantitative results of PT often need to be transformed into capability statistics for interpretation and comparison with other determined objectives. Its purpose is to measure deviations from the assigned values based on competency assessment criteria. As RMs were used in the PT, the reference values of RM can be used directly for the assigned values of the PT. Therefore; the *z′-score* was used to estimate the performance of each laboratory when the participant's result did not affect the assigned value [20]. The *z′-score* fraction formula is as follows:

x is the proficiency test result from the participant;

X is the assigned value;

σ is the standard deviation for proficiency assessment (determined by a normalized interquartile range of the participant);

u_X is the uncertainty of the assigned value.

$$z' = (x - X)/\sqrt{\sigma^2 + u^2{}_x}$$

Each participant was assessed according to the *z′-score* evaluation criteria shown in Table 2 [21,22,23].

Based on the fact that the logarithmic function is a monotonic increasing function in its definition domain, the relative relationship of data will not be changed after logarithmic processing. Therefore, the copy number of quantitative PT was processed by log10 and analyzed by violin and box diagram.

2.5. Homogeneity and stability of the PT samples

The previously established digital PCR method [18] was used to analyze the stability and homogeneity [18]. The homogeneity and stability of the six samples were tested at NIM. For samples A and B, the homogeneity was evaluated by *F*-test. For pseudovirus samples D and E, the whole vial would be extracted at once but not necessary to evaluate the homogeneity within the bottle. Therefore, the standard deviation among different vials was calculated as S_{bb}, and the S_{bb}^2 was compared with the method's repeatability (R). The result showed that all the samples were homogeneous (supplemental Table S1).

Short-term stability was assessed to ensure stability during the transportation. Copy number concentration under 4℃ was compared with the reference temperature of -70℃ for five days. Three vials were randomly selected at each time point under 4℃. Trend analysis was used to check if a significant change occurred. The results showed that the samples were stable at 4℃ for five days (supplemental Table S2). Therefore, the sample was shipped with dry ice to ensure the temperature was below 4℃, and all participants received samples within four days.

3. Results

3.1. Overall evaluation of qualitative PT

Fifty sets of valid data (including ORF1ab/N gene detection for sample D/E/F) were submitted from 48 participants (two of which provide two data sets with two methods). The overall accuracy was 88.00 %, 84.00 %, and 92.00 % for samples D, E, and F, respectively (Table 3). Among them, eleven participants presented false negatives (either for sample D or E), indicating that problems existed in the operation or testing process of the laboratory personnel involved. Sample F was a negative control, but 4 participants reported false positives indicating contamination occurred in these laboratories. Only 74.00 % of the participants reported correctly for all samples, which is relatively lower compared with 83.10 % of SARS-CoV-2 nucleic acid detection inter-laboratory quality evaluation conducted by the National Health Commission Clinical Laboratory Center in 2020. We speculated that the reason was that the copy number concentrations in our study were relatively lower, especially the copy number concentration in sample E even lower than the detection limit of testing kits used by some partic-

Table 3
Overall performance in SARS-CoV-2 qualitative proficiency testing.

Sample	Total number	False negative/positive number	Correct rate	False negative/positive number
Sample D	50	6	88.00%	12.00%
Sample E	50	4 + 4*	84.00%	16.00%
Sample F	50	4	92.00 %	8.00 %
All correct	50	13	74.00 %	26.00 %

* The detection limit of the kits used by some participating laboratories is higher than the gene concentration of sample E.
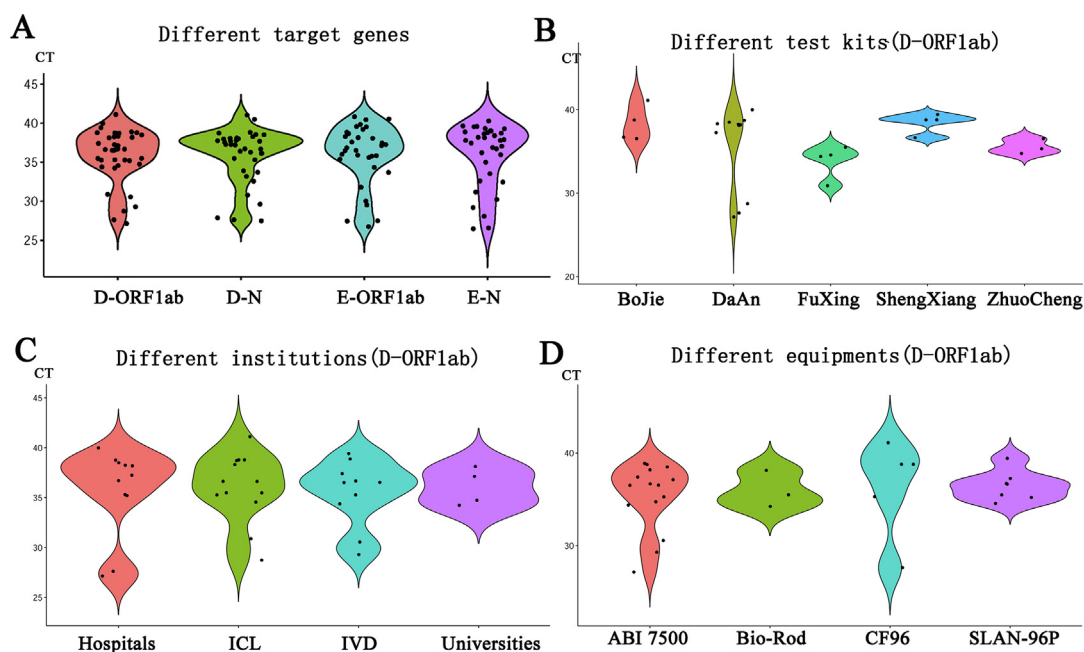
Fig. 1. Violin diagram of CT value distribution for qualitative proficiency testing (PT). CT value of ORF1ab and N in samples D and E (A). CT value of ORF1ab in Sample D with different nucleic acid detection kits (B), different participants (C) and on different equipment (D). Less than or equal to 3 categories were not included in the statistics.

Table 4
Classification of the participants in SARS-CoV-2 qualitative proficiency testing.

Classification	Kit (50 sets) *		Method (50 sets) *		Institution (48)			
	Approved	RUO**	RT-qPCR	RT-dPCR	Enterprises		Hospitals	Universities
					ICL	IVD		
Kind	12	11	/	/	ICL	IVD	/	/
Number of tests (Set)	37	13	42	8	17	14	11	6
Number of errors (Set)	5	6	8	3	7	4	0	1
Correct rate (%)	86.49	53.85	80.95	62.50	58.82	71.43	100.00	83.33
Error rate (%)	13.51	46.15	19.05	37.50	41.18	28.57	0.00	16.67

* Exclude errors caused by sample F.
** RUO, research use only.

ipants. Interestingly, the false negative rate increased from 12.00 % to 16.00 %, decreasing the concentration from about 600 copies/mL (Sample D) to 300 copies/mL (Sample E), as sample E was more challenging regarding its low copy number concentration.

3.2. Factors impacting testing performance in qualitative PT

3.2.1. CT distributions had no preference

Among the 50 sets of data, eight were detected by RT-dPCR and 42 by RT-qPCR. RT-qPCR was the primary method used for the detection of SARS-CoV-2 nucleic acid. The distribution of CT values and detection kits, equipment, and extraction kits used by all participants in the qualitative PT were plotted (Fig. 1, Fig. S2), and no CT values pref-

erence was observed. The CT values were distributed similarly and concentrated between 35 and 40 for two targets in two samples (Fig. 1A). The majority of CT values of sample E were slightly higher than that of sample D, which is consistent with the twofold difference in concentration of the two samples. Significant fluctuation of CT values was observed for those participants using DaAn test kits and the CF96 RT-qPCR platform. No significant factors directly affected the accuracy were found in the qualitative PT.

3.2.2. Hospitals have better testing capabilities and quality control systems

Among the participants, 31 were from enterprises, 11 from hospitals, and six from academics (research institutes and universities). The accuracy rates were 100.00 %, 83.33 %, and 64.52 % for partici-

pants from hospitals, academics, and enterprises, respectively (Table 4). For SARS-CoV-2 nucleic acid testing, hospitals have better testing capabilities and quality control systems with rigorous and accurate results. Some enterprises (14/31) were the in vitro diagnostic (IVD) manufacturers, who focus on testing kit development, and some enterprises (17/31) were the independent clinical laboratories (ICL, authorized and unauthorized). The error rate of IVD manufacturers was 28.60 %, while the error rate of ICL was as high as 41.20 % (Table 4). Furthermore, this indicates that enterprise participants, mainly from ICL, must improve the quality system, including personnel training on qualifications, quality control, and supervision.

3.2.3. Approved kits for better fault tolerance and compatibility

All participants used twenty-three different testing kits, of which 12 were approved by the National Medical Products Administration (NMPA), and 11 were not approved yet. In the face of major infectious diseases worldwide, most enterprises and academics have actively participated in the development and production of kits, and the overall R&D capabilities of the nucleic acid detection industry have been improved. Most participants chose the approved testing kits for the PT (Table 4). Excluding the errors in the detection of sample F entirely caused by the contamination of laboratory operations, considering comprehensive factors such as the type of kits and the level of participants [24], the error rate was 13.51 % and 46.15 % for those using approved kits and RUO kits, respectively. The low error rate of approved SARS-CoV-2 nucleic acid detection kits indicates that approved detection kits are relatively stable, reliable, and fault-tolerant. National supervision contributes significantly to the improvement of the accuracy of SARS-CoV-2 detection.
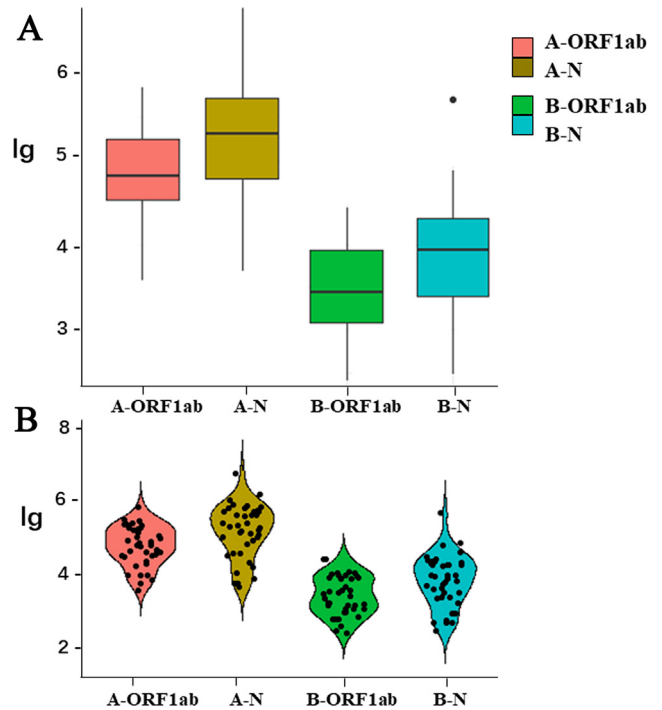

Fig. 3. Trend of quantitative proficiency testing (PT) results. A) Box diagram of samples A and B. B) Violin diagram of samples A and B. X-axis represented sample types; Y-axis was the result of log10 (lg) processing for different sample copy numbers (After logarithmic processing, the relative relationship of data will not change).
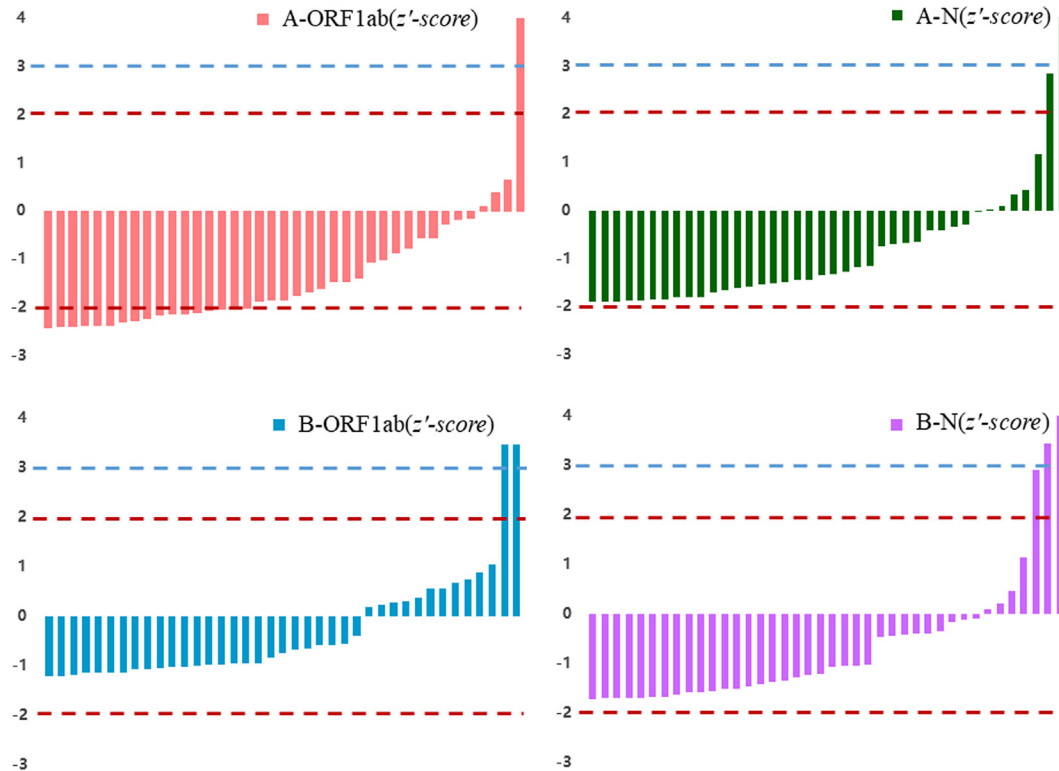

Fig. 2. z'-score diagram for quantitative proficiency testing (PT). The red dotted line is the absolute value of 2, and the blue dotted line is 3. Incomplete display when it is higher than 4.

Additionally, RT-dPCR showed a higher error rate than RT-qPCR because the kits used for RT-dPCR were under development, as none of them were approved by NMPA by the time of running the PT.

3.3. Overall evaluation of quantitative PT

A total of 42 laboratories participated in the quantitative PT. A few laboratories submitted partial results (submit ORF1ab or N only). The initial copy number concentration is shown in Fig. S3. The result of the z'-score for samples A and B was shown in Fig. 2. The satisfactory results of ORF1ab and N in sample A accounted for 77.22 %, the questionable results for 20.25 %, and the unsatisfactory results accounted for 2.53 %. The satisfactory results of ORF1ab and N in sample B accounted for 93.67 %, the questionable results accounted for 1.27 %, and the unsatisfactory results accounted for 5.06 % (Fig. 2). The quantitative results of ORF1ab in sample A showed a higher percentage of unsatisfactory/questionable than others. The possible reason is that the entire length of the ORF1ab gene was about 21kbp; more diverse primers and probes were used compared with the N gene (1.3kbp). Different primer/probes may lead to inconsistent amplification efficiency. It was observed that most participants used RT-qPCR with a standard curve to report the copy number concentration. Therefore, errors in preparing standard curves for each target were introduced due to unreasonable dilution gradients and additional operations, leading to a significant deviation in the quantitative result. Furthermore, the majority of z'-score was negative for the two targets of the samples, indicating RNA degradation occurred during the quantitative process.

The data was shown as logarithm ten due to the scattered copy number concentration. Through boxplot and violin plot analysis (Fig. 3), it was found that the medians of the tested samples A-N and B-ORF1ab were located in the middle of the entire boxplot, and the distances between the upper and lower edges are similar, indicating that the overall data distribution level of the samples was relatively uniform. One outlier was checked out for sample B-N (Fig. 3A). It was found that the ORF1ab results of sample B were concentrated on two peaks according to the violin plot (Fig. 3B). This indicates the results were biased and showed different aggregation areas in different laboratories. The quantitative capacity of the participants needs further improvement.

3.4. Factors impacting testing performance in quantitative PT

3.4.1. Digital PCR method with better comparability for quantitative measurement

For quantitative PT, 30 participants reported results with RT-qPCR and 9 with RT-dPCR. Therefore the total number of RT-dPCR tests was 36 times, and the number of unsatisfactory/questionable tests accounted for 8.33 %. On the other hand, the total number of RT-qPCR tests was 122, of which 16.39 % were unsatisfactory/questionable tests. Generally, RT-qPCR has more strict requirements for operators and other factors due to its quantification relying on an external standard curve, which may cause an increase in the error rate. Therefore, results obtained from the RT-dPCR method were more comparable and reliable for quantitative measurement because RT-dPCR provides quantification results independent of an external standard.

Table 5
Summary result of SARS-CoV-2 quantitative proficiency testing.

Classification	Type					Method		Institution [**]			
	Sample A		Sample B		Both[*]	RT-qPCR	RT-dPCR	Enterprises		Hospitals	Universities
	ORF1ab	N	ORF1ab	N	/	/	/	ICL	IVD	/	/
Number	39	40	39	40	37	122	36	50	60	28	20
Satisfactory	23	38	37	37	21	102	33	44	50	26	15
Questionable	15	1	0	1	16	15	2	6	6	2	3
Unsatisfactory	1	1	2	2	5	1	0	4	0	2	
Percentage of satisfactory	58.94 %	95.00 %	94.87 %	92.50 %	56.76 %	83.61 %	91.67 %	88.00 %	83.33 %	92.86 %	75.00 %
Percentage of unsatisfactory/questionable	41.03 %	5.00 %	5.13 %	7.50 %	43.24 %	16.39 %	8.33 %	12.00 %	16.67 %	7.14 %	25.00 %

ORF1ab and N genes were calculated separately.

[*] Submit complete data for sample A and sample B. If a laboratory was dissatisfied with one data, it was dissatisfied.
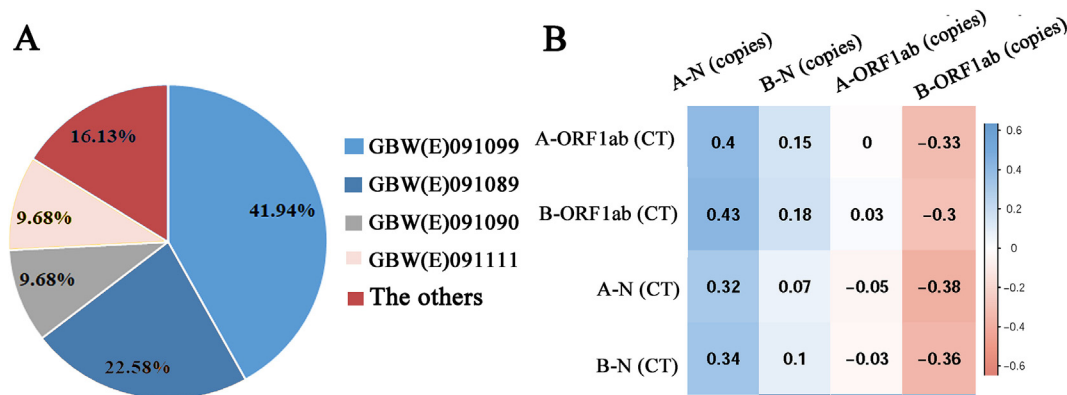
[**] A number of experimental results were submitted by different institutions.

Fig. 4. Proportion of reference materials used (A) and correlation analysis between CT and copies value (B). B) Analyzed the data in the same conditions of reference material, target gene, and detection kit. The correlation was reflected in the personnel operation and the establishment of the standard curve.

In the entire quantitative proficiency testing, the hospital's satisfactory rate was as high as 92.86 %, with only two questionable or not unsatisfactory (Table 5). Moreover, the satisfaction data of universities was only 75.00 % (Table 5), which indicates that colleges and universities were primarily scientific research and exploration and compared with hospitals, they lack quality system construction. Therefore, regardless of qualitative or quantitative research, a relatively complete quality system and special training of personnel can ensure the accuracy of different methods and capabilities.

3.4.2. The higher the concentration, the greater the fluctuation of the quantitative results

The fluctuation was compared between samples A and B to explore the relations between variability and different sample concentrations. The fluctuation of ORF1ab and N in sample A was much more extensive than that of sample B. By excluding the outliers (Grubbs' Test), the data of sample A dispersed in a range of 85 and 223 folds for ORF1ab and N, respectively. The results of sample B were scattered in a range of 50 and 106 folds for ORF1ab and N, respectively. Additionally, for sample A, 18 out of 79 were unsatisfactory/questionable, accounting for 22.78 %. For sample B, 5 out of 79 were dissatisfied, accounting for 6.33 %. The proportion of unsatisfactory was higher for sample A with high concentration, and most of them were used by the RT-qPCR method, indicating that the higher the sample concentration, the greater the fluctuation for quantitative determination.

3.4.3. The establishment of the standard curve of each laboratory is quite different

The participants selected nine kinds of reference materials using RT-qPCR [25]. One hundred sixteen standard curves need to be constructed, of which 74.20 % were reference materials using NIM (GBW(E)091099, GBW(E)091089, and GBW(E)091090) (Fig. 4A). Linear correlation was analyzed between CT value and copy number under the same conditions of reference material (GBW(E)091099), target gene and detection kit (DaAn). The correlation coefficients were 0, 0.32, −0.30, and 0.10 for A-ORF1ab, A-N, B-ORF1ab, and B-N, respectively (Fig. 4B). It was found that $|R^2|$ was close to 0, indicating that the correlation between the establishment or operation of standard curves in different laboratories was very poor. Preparing the standard curve is a critical factor that leads to a big difference in quantitative results. Thus the procedure should be standardized.

3.5. Accurate quantitative measurement ensuring reliable qualitative detection

To further explore the impact of quantitative measurement on qualitative detection, we compared the satisfactory and unsatisfactory participants in both quantitative and qualitative PT. There were 22 laboratories with both correct (satisfactory) qualitative and quantitative results, accounting for 95.65 % of the satisfactory quantitative PT. Most of the laboratories with correct quantitative results in quantitative PT can report correctly in qualitative PT (Fig. 5A). There were six laboratories with both incorrect (unsatisfactory) qualitative and quantitative results, accounting for 85.71 % of the unsatisfactory qualitative PT (Fig. 5B). Most laboratories with incorrect qualitative results were also unsatisfied with quantitative results. Thus, quantitative ability is the basis of qualitative judgment.

4. Discussion

Evaluation of PT results from a large group of laboratories helps assess the performance of test methods applied in various conditions [26]. The significance of this measure is compounded when a test is new and widely practiced, and the results have several applications. Through this PT, it was found that the overall results are not very satisfactory regarding qualitative or quantitative PT. Only 74.00 % of the laboratories were utterly correct in qualitative PT, and the number of satisfactory quantitative participants accounted for 56.76 % (all genes were satisfied). Especially the copy number concentration fluctuated in the range of tens to hundreds of times. Given the spread of such a large-scale worldwide epidemic, it is still tricky to promptly establish a complete and accurate detection system. Therefore, on the one hand, it is necessary to strengthen the training and management of personnel within the laboratory. On the other hand, it is necessary to deepen the supervision and construction of the quality system comprehensively to cope with the outbreak of a series of infectious diseases such as SARS and MERS (Fig. 6) [27].

4.1. National supervision and the development of various verification projects play an essential role in improving nucleic acid detection capabilities

By analyzing some laboratories with poor performance, incorrect results can be attributed to two main reasons: 1) Cross-contamination. Many laboratories reported false positive results for sample F, indicating sample contamination was caused by personnel operation and other reasons. 2) Low sensitivity of NAT (Fig. 6). Eleven laboratories report false negatives for samples D or E. On the one hand, the test kits used have not been registered and approved by NMPA, with the performance not sufficiently verified and evaluated. On the other hand, the stability and sensitivity of the detection kit are insufficient, and the detection limit is not enough. However, testing laboratories from hospitals showed a better performance than those from enterprises in this PT. Testing laboratories from hospitals have been engaged in medical and health testing for a long time and often participate in various verification evaluations, its supervision and manage-
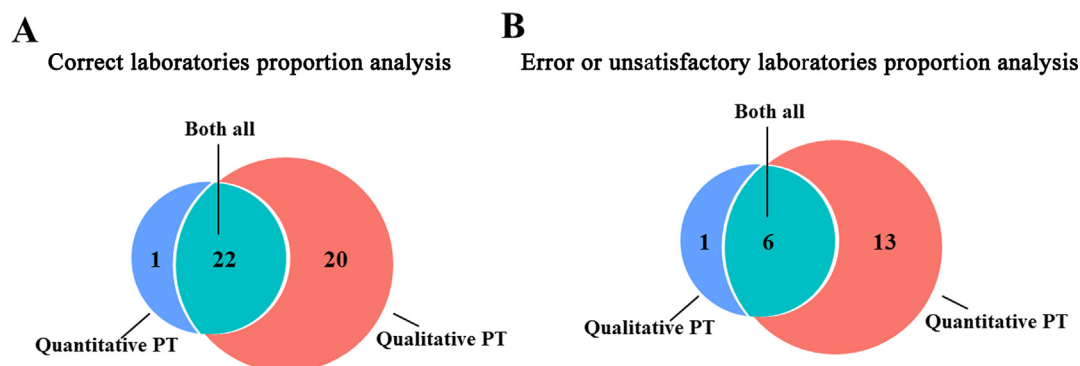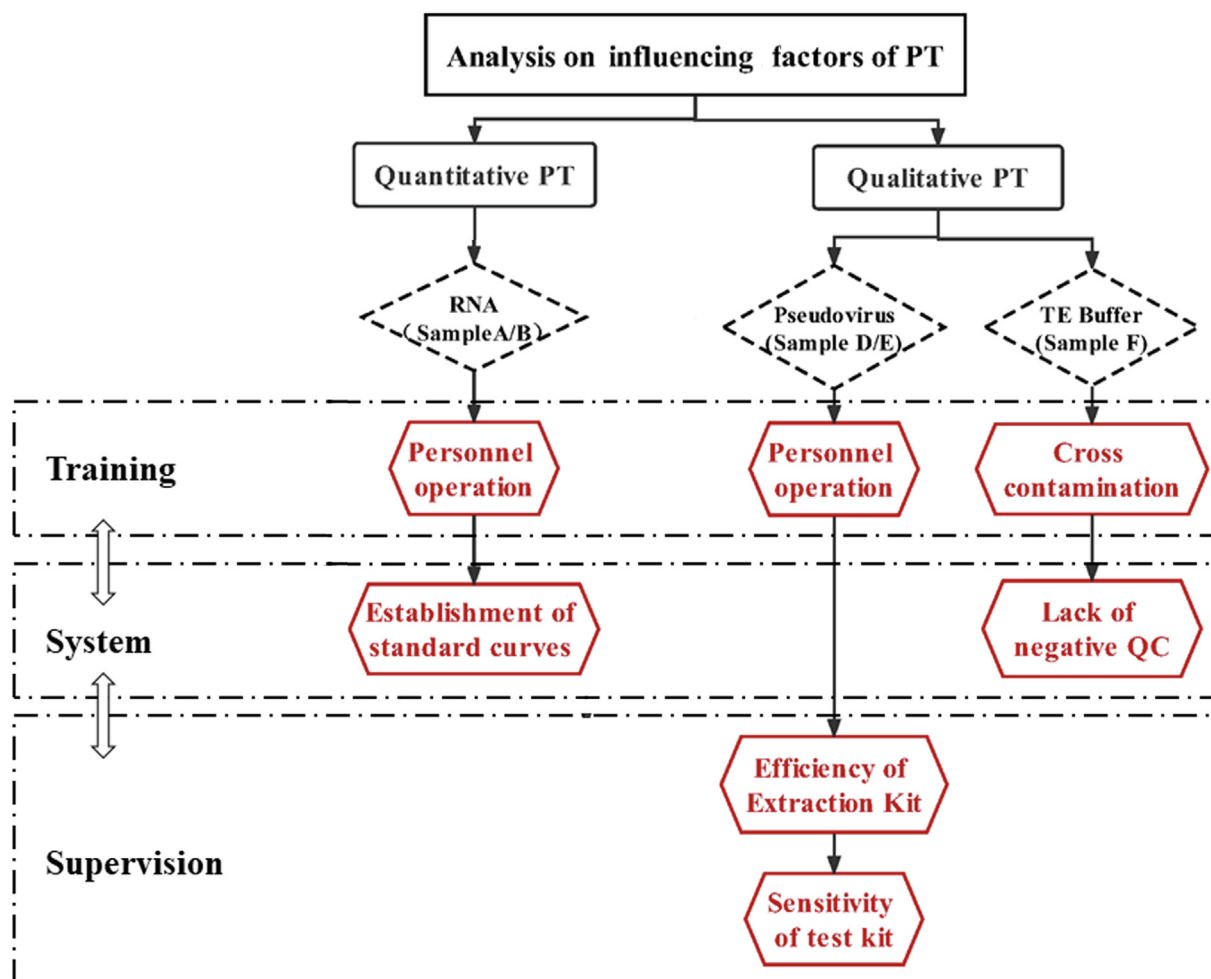

Fig. 5. Analysis of laboratories with correct (A) and incorrect (B) results in qualitative and quantitative proficiency testing (PT). Only analyze the data involved in both qualitative and quantitative PT and exclude the error results of sample F (cross-contamination) and sample E caused by detection limit issue.

Fig. 6. Main factors contributing to the proficiency testing (PT). The red box represents the problems in this PT, and the dotted line on the left represents the solution direction.

ment are rigorous. Meanwhile, these participants used approved testing kits with more stable performance, leading to better accuracy. A lack of enough evidence for the unapproved testing kits could easily lead to an inaccurate test. Therefore, systems such as "CE" certification based on "self-declaration" are temporarily unsuitable for China. National supervision and PT program can be essential in promoting and optimizing infection disease detection.

4.2. Accurate quantitative detection is the cornerstone of qualitative judgment, and the ability of quantitative detection needs to be improved

Through this PT, we found that there are still many contents worth discussing low-concentration samples and accurate quantitative detection. Several unsatisfactory results were reported for samples D and E. Participating laboratories used different extraction and elution volumes according to different kit instructions (supplemental Table S3). Due to the extraction steps, it is easy to cause a low amount of extracted samples by low extraction efficiency [28,29], which may result in false negatives. Furthermore, the CT values of low-concentration samples were near the gray area, and it was not easy to decide whether it was negative or positive without further checking. Thus, using the weak positive control material in the daily control is necessary. For quantitative measurement, the deviation of high-concentration samples was extensive, resulting in many unsatisfactory results. Although participating laboratories can give good results for

negative and positive judgments of high concentration samples, the capability for quantitative measurement is still poor. Especially for these IVD manufacturers, the limit of detection (LoD) of the testing kits cannot be determined accurately without qualified quantitative capability. All qualitative detection results were judged according to the testing kits with reliable LoD. Therefore, it is necessary further to improve the quantitative measurement capability of NAT for SARS-CoV-2.

4.3. The establishment of standardized processes and traceability systems is the basis for ensuring accurate nucleic acid detection

Overall, establishing a traceability system and a standardized process is also critical to comprehensively improve the accuracy of nucleic acid detection, to ensure the kit's stability and personnel qualification. Before starting the detection, the problems caused by the equipment should be eliminated first, so the calibration should be carried out in time. At the same time, focus on strengthening the operation training of testing personnel, especially implementing reference materials and standard procedures. Good traceability delivery and system establishments are essential to support and prerequisites for ensuring nucleic acid detection.

In conclusion, this study showed that not all laboratories could obtain a satisfactory result, especially those from ICL need further improve their quality system, including a selection of qualified testing

kits, personal training, standardizing operation, daily quality control, and regular external evaluation, etc. However, continuing to participate in various verification programs, improving the overall detection capability of the laboratory, paying attention to technical details, and staff training can significantly avoid testing errors and provide help for epidemic prevention and control.

Acknowledgements

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Yongzhuo Zhang: Conceptualization, Formal Analysis, Validation, Visualization, Writing. **Xia Wang**: Investigation, Methodology, Validation. **Chunyan Niu**: Investigation, Methodology, Visualization. **Di Wang**: Investigation, Methodology, Resources. **Qingfei Shen**: Investigation, Methodology, Resources. **Yunhua Gao**: Supervision, Resources, Conceptualization. **Haiwei Zhou**: Resources, Conceptualization, Investigation. **Yujing Zhang**: Resources, Conceptualization, Investigation. **Yan Zhang**: Resources, Conceptualization, Investigation. **Lianhua Dong**: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Supervision, Writing.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bsheal.2022.08.004.

References

[1] World Health Organization, Novel coronavirus (2019-nCoV) situation report. https://covid19.who.int, 2022 (accessed 5 May 2022).

[2] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, Nat. Microbiol. 5 (4) (2020) 536–544, https://doi.org/10.1038/s41564-020-0695-z.

[3] Centers for Disease Control and Prevention, COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/index.html, 2019 (accessed 5 May 2022).

[4] X.M. He, W.Q. Hong, X.Y. Pan, et al., SARS-CoV-2 Omicron variant: Characteristics and prevention, Med. Comm. 2 (4) (2020) 838–845, https://doi.org/10.1002/mco2.110.

[5] M.M. Rajah, M. Hubert, E. Bishop, et al., SARS-CoV-2 Alpha, Beta, and Delta variants display enhanced Spike-mediated syncytia formation, EMBO J. 40 (2021), e108944. https://doi.org/10.15252/embj.2021108944.

[6] N.L. Miller, T. Clark, R. Raman, et al., Insights on the mutational landscape of the SARS-CoV-2 Omicron variant [Preprint], bioRxiv. (2021), https://doi.org/10.1101/2021.12.06.471499.

[7] M.A. Lauxmann, N.E. Santucci, A.M. Autrán-Gómez, The SARS-CoV-2 coronavirus and the COVID-19 Outbreak, Int. Braz. J. Urol. 46 (2020) 6–18, https://doi.org/10.1590/s1677-5538.ibju.2020.s101.

[8] D.C. Edson, D.L. Casey, S.E. Harmer, et al., Identification of SARS-CoV-2 in a proficiency testing Program, Am. J. Clin. Pathol. 154 (4) (2020) 475–478, https://doi.org/10.1093/ajcp/aqaa128.

[9] P. Rai, B.K. Kumar, V.K. Deekshit, et al., Detection technologies and recent developments in the diagnosis of COVID-19 infection, Appl. Microbiol. Biotechnol. 105 (2) (2021) 441–455, https://doi.org/10.1007/s00253-020-11061-5.

[10] C.K.C. Lai, W. Lam, Laboratory testing for the diagnosis of COVID-19, Biochem. Biophys. Res. Commun. 538 (2021) 226–230, https://doi.org/10.1016/j.bbrc.2020.10.069.

[11] M. Pradhan, K. Shah, A. Alexander, et al., COVID-19: clinical presentation and detection methods, J. Immunoassay Immunochem. 43 (2022), 1951291. https://doi.org/10.1080/15321819.2021.1951291.

[12] H. Wang, X. Li, T. Li, et al, The genetic sequence, origin, and diagnosis of SARS-CoV-2, Eur. J. Clin. Microbiol. Infect. Dis. 39 (2020) 1629–1635, https://doi.org/10.1007/s10096-020-03899-4.

[13] R.J. Lu, X. Zhao, J. Li, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet. 395 (10224) (2020) 565-574, https://doi.org/10.1016/S0140-6736(20)30251-8.

[14] M. Różycki, W. Korpysa-Dzirba, A. Bełcik, et al, Results of proficiency testing for Trichinella in Poland, 2015–2019, J. Clin. Med. 10 (22) (2021), 5389. https://doi.org/10.3390/jcm10225389.

[15] S.S. Ehrmeyer, R.H. Laessig, Interlaboratory proficiency-testing programs: a computer model to assess their capability to correctly characterize intralaboratory performance, Clin. Chem. 6 (1987) 784–787.

[16] I. Görzer, Ch. Buchta, P. Chiba, et al, First results of a national external quality assessment scheme for the detection of SARS-CoV-2 genome sequences, J. Clin. Virol. 129 (2020), 104537. https://doi.org/10.1016/j.jcv.2020.104537.

[17] C. Ceraolo, F.M. Giorgi, Genomic variance of the 2019-nCoV coronavirus, J. Med. Virol. 92 (2020) 522–528, https://doi.org/10.1002/jmv.25700.

[18] L. Dong, J. Zhou, C. Niu, et al, Highly accurate and sensitive diagnostic detection of SARS-CoV-2 by digital PCR, Talanta. 224 (2021), 121726. https://doi.org/10.1016/j.talanta.2020.121726.

[19] V.M. Corman, O. Landt, M. Kaiser, et al, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro. Surveill. 25 (3) (2020), 2000045. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

[20] ISO/IEC 17043:2010, Conformity assessment—general requirements for proficiency testing, International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland, 2010.

[21] ISO 13528, Statistical methods for use in proficiency testing by interlaboratory comparison, International Organization for Standardization, Geneva, Switzerland, 2015.

[22] Proficiency testing of sampling, Anal. Methods 9 (28) (2017) 4110–4111, https://doi.org/10.1039/C7AY90092A.

[23] China National Accreditation Service for Conformity Assessment. Guidance on Statistic Treatment of Proficiency Testing Results and Performance Evaluation: CNASGL002:2018. Beijing: China National Accreditation Service for Conformity Assessment, (2018).

[24] R.T. Hayden, X. Yan, M.T. Wick, et al, Factors contributing to variability of quantitative viral PCR results in proficiency testing samples: a multivariate analysis, J. Clin. Microbiol. 50 (2012) 337–345, https://doi.org/10.1128/JCM.01287-11.

[25] L. Vierbaum, N. Wojtalewicz, H.P. Grunert, 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 17 (2022), e0262656. https://doi.org/10.1371/journal.pone.0262656.

[26] H. Kaur, L. Mukhopadhyay, N. Aggarwal, et al., Inter-laboratory testing as a strategy for external quality assessment for qualitative detection of SARS-CoV-2 by real-time RT-PCR testing in India, Indian J. Med. Res. 155 (1) (2022) 86–90, https://doi.org/10.4103/ijmr.ijmr_2433_21.

[27] T. Theis, K.A. Lau, J.L. Gray, et al, Proficiency testing for the detection of Middle East respiratory syndrome coronavirus demonstrates global capacity to detect Middle East respiratory syndrome coronavirus, J. Med. Virol. 90 (2018) 1827–1833, https://doi.org/10.1002/jmv.25266.

[28] C. Beltrán-Pavez, L.A. Alonso-Palomares, F. Valiente-Echeverría, et al, Accuracy of a RT-qPCR SARS-CoV-2 detection assay without prior RNA extraction, J. Virol. Methods. 287 (2021), 113969. https://doi.org/10.1016/j.jviromet.2020.113969.

[29] C. Ambrosi, C. Prezioso, P. Checconi, et al, SARS-CoV-2: Comparative analysis of different RNA extraction methods, J. Virol. Methods. 287 (2021), 114008. https://doi.org/10.1016/j.jviromet.2020.114008.