Analytical and Clinical Comparison of Three Nucleic Acid Amplification Tests for SARS-CoV-2 Detection

Elizabeth Smith, Wei Zhen, Ryhana Manji, Deborah Schron, Scott Duong, Gregory J. Berry

Infectious Disease Diagnostics, Northwell Health Laboratories, Lake Success, New York, USA
Department of Pathology and Laboratory Medicine, The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, New York, USA

ABSTRACT
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in December 2019 and has quickly become a worldwide pandemic. In response, many diagnostic manufacturers have developed molecular assays for SARS-CoV-2 under the Food and Drug Administration (FDA) Emergency Use Authorization (EUA) pathway. This study compared three of these assays, the Hologic Panther Fusion SARS-CoV-2 assay (Fusion), the Hologic Aptima SARS-CoV-2 assay (Aptima), and the BioFire Defense COVID-19 test (BioFire), to determine analytical and clinical performance as well as workflow. All three assays showed similar limits of detection (LODs) using inactivated virus, with 100% detection, ranging from 500 to 1,000 genome equivalents/ml, whereas use of a quantified RNA transcript standard showed the same trend but had values ranging from 62.5 to 125 copies/ml, confirming variability in absolute quantification of reference standards. The clinical correlation found that the Fusion and BioFire assays had a positive percent agreement (PPA) of 98.7%, followed by the Aptima assay at 94.7%, compared to the consensus result. All three assays exhibited 100% negative percent agreement (NPA). Analysis of discordant results revealed that all four samples missed by the Aptima assay had cycle threshold (Ct) values of >37 by the Fusion assay. In conclusion, while all three assays showed similar relative LODs, we showed differences in absolute LODs depending on which standard was employed. In addition, the Fusion and BioFire assays showed better clinical performance, while the Aptima assay showed a modest decrease in overall PPA. These findings should be kept in mind when making platform testing decisions.

KEYWORDS
SARS-CoV-2, COVID-19, EUA, molecular diagnostics, TMA, PCR, quantified virus

The novel coronavirus was first detected in the United States in Washington State on 20 January 2020 (1), and by 13 March 2020, the World Health Organization (WHO) characterized the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak as a worldwide pandemic (2). Since then, SARS-CoV-2 has rapidly spread across the country. At the time of this report in June 2020, the number of cases in the United States has reached nearly 2.2 million, and the number of deaths has exceeded 100,000, according to the Johns Hopkins University database (3). Certain regions and cities have much higher prevalences than others, especially the New York metropolitan area that our laboratories serve. New York City alone now accounts for approximately 10% of the cases in the United States (3).

SARS-CoV-2 is the etiologic agent of coronavirus disease 2019 (COVID-19), for which symptoms can vary from mild (e.g., fever, fatigue, dry cough) to severe illness (e.g., dyspnea, persistent pain or pressure in the chest, hypoxia, confusion), and can ultimately lead to death, particularly for those ≥65 years of age and those with certain underlying medical conditions, such as heart disease, lung disease, or diabetes (4). Even
more challenging, some are infected with the virus but do not display any signs or symptoms of illness, which likely has contributed to the high rate of transmission among humans (5).

Accurate and sensitive viral detection methods are key to quickly diagnose infections and mitigate transmission. Nucleic acid amplification tests (NAATs) are highly sensitive and specific methods for the detection of SARS-CoV-2 RNA in respiratory specimens. The majority of the SARS-CoV-2 NAATs available today are based on real-time reverse transcription PCR (RT-PCR) methods, including the BioFire Defense COVID-19 test (BioFire) and the Hologic Panther Fusion SARS-CoV-2 assay (Fusion) evaluated in the present study. Clinical comparative data have been obtained for the Fusion assay (6–9), but to our knowledge, there have been no comparative studies of the BioFire assay. Recently, Hologic has developed a second NAAT, the Aptima SARS-CoV-2 assay (Aptima), which has been granted Food and Drug Administration (FDA) Emergency Use Authorization (EUA); this NAAT is based on target capture and transcription-mediated amplification (TMA) technologies and is run on the Panther instrument. As with the BioFire assay, comparative evaluations have yet to be performed for this new molecular assay.

An increase in testing capability is critically needed to manage the current testing demands and to monitor and control the spread of the virus going forward. While the FDA has worked quickly to review and authorize more molecular diagnostic platforms to make more options available to meet demand, real-world clinical performance and comparative data are still lacking. These data are urgently needed to better understand the benefits and limitations of each test and how best to incorporate them into local and national testing strategies.

Here, we present an analytical and clinical evaluation of these three sample-to-answer NAATs for the qualitative detection of SARS-CoV-2 RNA in symptomatic patients: the Fusion assay, the Aptima assay, and the BioFire single-target COVID-19 assay.

**MATERIALS AND METHODS**

**Specimen collection and storage.** Nasopharyngeal (NP) swabs were obtained from patients with clinical signs/symptoms of COVID-19. Each collection used a sterile swab made from Dacron, rayon, or nylon, which was placed into 3 ml of sterile universal transport medium (UTM; from various manufacturers) after collection. Samples were then transported at room temperature, stored at 2 to 8°C for up to 72 h, and tested as soon as possible after collection. For the retrospective sample set, after routine patient testing, aliquots of samples were taken and stored at −80°C until testing of the three comparator NAATs could take place.

**Study design.** Samples were selected from specimens received for routine SARS-CoV-2 testing between April and May of 2020 that were initially tested by the Fusion assay. The selection included symptomatic patient samples that were selected at random without bias toward specific age or gender. A total of 150 NP samples (75 negative and 75 positive) were used for this study: 101 retrospective specimens (51 negative and 50 positive) and 49 prospective, fresh specimens (24 negative and 25 positive). Specimens were selected to represent our laboratory’s positivity rate at the time this study was designed (50 to 60% at beginning of April 2020) and also included positive specimens spanning the range of positivity, including those with low viral loads (characterized by high cycle threshold [Ct] values obtained by results from initial clinical testing). Ct values obtained from the Fusion assay during comparative testing are also shown in Fig. S1 in the supplemental material. The 101 retrospective specimens were initially tested per routine patient care and then immediately aliquoted and frozen at −80°C, remaining frozen until this study was performed. Retrospective sample aliquots were thawed and immediately tested on the Fusion, Aptima, and BioFire systems. Testing for prospective samples was performed within 48 h of collection. Prospective samples were run on the Fusion, Aptima, and BioFire systems using the original UTM sample and aliquoted directly into Hologic lysis tubes and into the BioFire sample injection vial. This work was conducted as a quality improvement activity in order to complete assay validation for the Aptima and BioFire assays.

**Hologic Panther Fusion SARS-CoV-2 assay.** The Panther Fusion SARS-CoV-2 assay (Fusion) (Hologic, Inc., San Diego, CA) was performed on the Fusion instrument according to the manufacturer’s instructions for use. This assay targets two unique regions of the open reading frame 1ab (ORF1ab) section of the SARS-CoV-2 viral genome. A 500-μl aliquot of the primary NP swab specimen is transferred into a specimen lysis tube containing 710 μl of lysis buffer and is then loaded onto the Fusion instrument. From this tube, the instrument removes 360 μl for extraction. Each specimen is processed with an internal control (IC), which is added via the working Fusion capture reagent-5. Nucleic acid is purified using capture oligonucleotides and a magnetic field and eluted in 50 μl, and 5 μl of the eluted nucleic acid is added to a Fusion reaction tube. The Fusion assay amplifies and detects two conserved regions of the
was performed according to the manufacturer’s instructions and is briefly described. A 500-ORF1ab section of the SARS-CoV-2 viral genome and is performed on the Panther instrument. All testing strategies for the isolation and amplification of SARS-CoV-2 RNA. This assay targets two unique regions of the ORF1ab section and one unique region of the ORF8 section performed on the BioFire FilmArray Torch 12-bay system according to the manufacturer’s instructions. The assay targets two unique regions of the ORF1ab section of the SARS-CoV-2 viral genome and is performed on the Panther instrument. All testing was performed according to the manufacturer’s instructions and is briefly described. A 500-µl aliquot of the primary NP swab specimen is transferred into a specimen lysis tube containing 710 µl of lysis buffer, and this tube is then loaded onto the Panther instrument. From the specimen lysis tube, 360 µl aliquot of lysis buffer to the sample injection vial, which is then inverted to mix. The sample mix is injected into the pouch sample port. Within the pouch, the sample is lysed by agitation (bead beating), and all nucleic acid is extracted and purified using magnetic bead technology. Nested multiplex PCR is performed, and endpoint melting curve data are used to detect and generate a result for each target assay of the BioFire COVID-19 test.

BioFire COVID-19 test. The BioFire COVID-19 test (BioFire) (BioFire Defense, Salt Lake City, UT) was performed on the BioFire FilmArray Torch 12-bay system according to the manufacturer’s instructions. This assay targets two unique regions of the ORF1ab section and one unique region of the ORF8 section of the SARS-CoV-2 viral genome. The test pouch (one per sample to be tested) is prepared by injecting hydration solution into the pouch hydration port. Using a provided transfer pipette, approximately 300 µl of sample is added to the sample injection vial, followed by the addition of a single-use sample buffer tube to the sample injection vial, which is then inverted to mix. The sample mix is injected into the pouch sample port. Within the pouch, the sample is lysed by agitation (bead beating), and all nucleic acid is extracted and purified using magnetic bead technology. Nested multiplex PCR is performed, and endpoint melting curve data are used to detect and generate a result for each target assay of the BioFire COVID-19 test.

Analytical sensitivity. The limit of detection (LOD) was determined using quantified, inactivated (gamma-irradiated) SARS-CoV-2 material from isolate USA-WA1/2020 (NR-52287; BEI Resources, Manassas, VA) and a SARS-CoV-2 synthetic RNA quantified control containing five gene targets (E, N, ORF1ab, RdRp, and S genes of SARS-CoV-2) from Exact Diagnostics, Fort Worth, TX (SKU COV019). The BEI Resources material was provided at a concentration of 4.1 × 10^5 genome equivalents (GE/ml) (2.8 × 10^5 50% tissue culture infective doses (TCID50)/ml) before inactivation of the virus, from which the following serial dilutions were prepared (in GE/ml): 1,000, 500, 250, 125, 62.5, and 31.2. Dilutions were prepared (in GE/ml): 1,000, 500, 250, 125, 62.5, and 31.3. Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The limit of detection (LOD) by positivity rate for each NAAT is highlighted in bold.

The number of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 GE/ml), 1 (500 GE/ml), 1 (250 GE/ml), 5 (125 GE/ml), 2 (62.5 GE/ml), and 0 (31.3 GE/ml). Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The numbers of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 copies/ml, 500 copies/ml, and 250 copies/ml), 2 (125 copies/ml), 6 (62.5 copies/ml), and 4 (31.3 copies/ml).

ORF1ab section of the SARS-CoV-2 viral genome. Both amplicons are detected by probes using the same fluorescent reporter, with amplification of either or both regions contributing to a single fluorescent signal and single fluorescence endpoint melting curve data are used to detect and generate a result for each target assay of the BioFire COVID-19 test.

Analytical sensitivity. The limit of detection (LOD) was determined using quantified, inactivated (gamma-irradiated) SARS-CoV-2 material from isolate USA-WA1/2020 (NR-52287; BEI Resources, Manassas, VA) and a SARS-CoV-2 synthetic RNA quantified control containing five gene targets (E, N, ORF1ab, RdRp, and S genes of SARS-CoV-2) from Exact Diagnostics, Fort Worth, TX (SKU COV019). The BEI Resources material was provided at a concentration of 4.1 × 10^5 genome equivalents (GE/ml) (2.8 × 10^5 50% tissue culture infective doses (TCID50)/ml) before inactivation of the virus, from which the following serial dilutions were prepared (in GE/ml): 1,000, 500, 250, 125, 62.5, and 31.2. Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The numbers of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 GE/ml), 1 (500 GE/ml), 1 (250 GE/ml), 5 (125 GE/ml), 2 (62.5 GE/ml), and 0 (31.3 GE/ml). Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The numbers of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 copies/ml, 500 copies/ml, and 250 copies/ml), 2 (125 copies/ml), 6 (62.5 copies/ml), and 4 (31.3 copies/ml).

ORF1ab section of the SARS-CoV-2 viral genome. Both amplicons are detected by probes using the same fluorescent reporter, with amplification of either or both regions contributing to a single fluorescent signal and single fluorescence endpoint melting curve data are used to detect and generate a result for each target assay of the BioFire COVID-19 test.

Analytical sensitivity. The limit of detection (LOD) was determined using quantified, inactivated (gamma-irradiated) SARS-CoV-2 material from isolate USA-WA1/2020 (NR-52287; BEI Resources, Manassas, VA) and a SARS-CoV-2 synthetic RNA quantified control containing five gene targets (E, N, ORF1ab, RdRp, and S genes of SARS-CoV-2) from Exact Diagnostics, Fort Worth, TX (SKU COV019). The BEI Resources material was provided at a concentration of 4.1 × 10^5 genome equivalents (GE/ml) (2.8 × 10^5 50% tissue culture infective doses (TCID50)/ml) before inactivation of the virus, from which the following serial dilutions were prepared (in GE/ml): 1,000, 500, 250, 125, 62.5, and 31.2. Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The numbers of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 GE/ml), 1 (500 GE/ml), 1 (250 GE/ml), 5 (125 GE/ml), 2 (62.5 GE/ml), and 0 (31.3 GE/ml). Dilutions for the Exact Diagnostics RNA are given in copies/ml.

The numbers of replicates at each dilution that gave initial equivocal results and required repeat testing were 0 (1,000 copies/ml, 500 copies/ml, and 250 copies/ml), 2 (125 copies/ml), 6 (62.5 copies/ml), and 4 (31.3 copies/ml).
NAAT was based on each manufacturer’s results interpretation algorithm. The positive percent agreement (PPA), negative percent agreement (NPA), positivity rate, kappa, and two-sided (upper/lower) 95% confidence interval (CI) were calculated using Microsoft Office Excel 365 MSO software (Microsoft, Redmond, WA). As a measure of overall agreement, Cohen’s kappa (\(\kappa\)) values were calculated, with values categorized as follows: >0.90, almost perfect; 0.90 to 0.80, strong; 0.79 to 0.60, moderate; 0.59 to 0.40, weak; 0.39 to 0.21, minimal; 0.20 to 0, none (10, 11). McNemar’s chi-square test was also calculated using R v3.6.3 (R Core Team, Vienna, Austria) (12).

RESULTS

Analytical sensitivity. Both quantified, inactivated SARS-CoV-2 virus (BEI Resources) and Exact Diagnostics SARS-CoV-2 synthetic RNA quantified control were used to prepare serial dilution panels (1,000 to 31.3 [GE or copies]/ml, in 2-fold dilutions) to determine the LOD of each assay. The LOD was defined as the lowest dilution in which all replicates were detected (100% positivity rate), using the results interpretation algorithm per the instructions for use for each assay. Using these criteria, the LOD using inactivated SARS-CoV-2 virus was 1,000 GE/ml for the Fusion assay, 500 GE/ml for the Aptima assay, and 500 GE/ml for the BioFire test (Table 1). In addition, the LOD as determined using the Exact Diagnostics synthetic RNA transcript was 62.5 copies/ml for both the Fusion and Aptima assays and 125 copies/ml for the BioFire test.

Clinical performance. After testing of the 150 clinical specimens on all three platforms, consensus results were determined for each sample (consensus positive is a positive result in \(\geq 2\) of 3 NAATs; consensus negative is a negative result in \(\geq 2\) of 3 NAATs), and results from each NAAT were compared to the consensus result. The Fusion and BioFire tests exhibited the highest PPA of 98.7%, while the Aptima assay had a PPA of 94.7% (Table 2). NPAs were 100% for all three NAATs, with no false-positive results for any platform. Cohen’s kappa values were 0.987, 0.947, and 0.987 for the Fusion, Aptima, and BioFire tests, respectively, all of which indicate an “almost perfect” level of agreement with the consensus result. In addition, McNemar’s chi-square test was also performed and showed no significant difference between each assay and the consensus result (Table 2).

Initial equivocal or invalid results occurred for three samples out of the sample set: two samples with equivocal results for the BioFire test and one sample with an invalid result for the Fusion assay. The initial BioFire results for the two samples were one of three targets detected (one of the ORF1ab targets); repeat testing gave the same result for both samples, for an overall interpretation of “detected” as described by the manufacturer. The initial invalid result by the Fusion assay repeated as invalid, and this sample was removed from the overall agreement analysis for the Fusion assay; since this sample was negative by both the Aptima and BioFire tests, giving a consensus negative result, this sample was kept in the overall agreement analyses for the Aptima and BioFire tests.

---

**TABLE 2** Clinical performance comparison of three SARS-CoV-2 NAATs for 150 NP swab specimens

| Molecular assay and result | Consensus result | Kappa (\(\kappa\)) (±95% CI) | McNemar’s test (P value) | PPA (%) (±95% CI) | NPA (%) (±95% CI) |
|-----------------------------|-----------------|--------------------------|----------------------|-----------------|-----------------|
| Fusion\(^a\)                |                 |                          |                      |                 |                 |
| Positive                    | 74              | 0                        | 0.987 (0.96–1.00)    | 1               | 98.7 (0.928–0.998) |
| Negative                    | 1               | 74                       |                      |                 | 100 (0.951–1)   |
| Aptima                      |                 |                          |                      |                 |                 |
| Positive                    | 71              | 0                        | 0.947 (0.895–0.998)  | 0.1336          | 94.7 (0.871–0.979) |
| Negative                    | 4               | 75                       |                      |                 | 100 (0.951–1)   |
| BioFire                     |                 |                          |                      |                 |                 |
| Positive                    | 74              | 0                        | 0.987 (0.961–1.00)   | 1               | 98.7 (0.928–0.998) |
| Negative                    | 1               | 75                       |                      |                 | 100 (0.951–1)   |

\(^a\): upper/lower 95% confidence interval (CI).

\(^b\): One sample gave an invalid result and was excluded from the Fusion agreement analysis.

---

Smith et al. Journal of Clinical Microbiology September 2020 Volume 58 Issue 9 e01134-20 jcm.asm.org 4
There were no discordant results among the prospective, fresh sample set; all of the discordant samples occurred among the retrospective sample set. Of the three NAATs evaluated, the Aptima assay had the most discordant results with the consensus (n = 4), followed by the Fusion and BioFire tests, each with one discordant result. Discordant sample results are shown in Table 3. The four discordant samples for Aptima had Ct values of ≥37.3 by the Fusion assay, indicating lower viral titers in these samples. The discordant BioFire sample had a Ct value by the Fusion assay of 35.7. For the five Fusion/Aptima discordant samples, the BioFire test detected only two of three targets in each sample (samples GSD-3, GSD-6, GSD-23) or detected one of three targets in each sample twice for an overall result of “detected” (GSD-4, GSD-48) (Table 3).

**Workflow.** Workflow parameters along with basic assay characteristics are presented in Table 4. Maximum sample throughput in 8-h and 24-h time periods were considerably higher on the Fusion and Aptima platforms (Fusion, 335 samples/8 h and 1,150 samples/24 h; Aptima, 275 samples/8 h and 1,020 samples/24 h), while throughput for BioFire was 72 samples/8 h and 216 samples/24 h. Hands-on time (HoT) per specimen for the both the Fusion and Aptima assays is 1 min, whereas the BioFire HoT is 3 min per specimen. Sample preparation for the Fusion and Aptima assays consists of aliquoting 500 μL of sample into a lysis tube. Sample preparation for BioFire consists of five steps, including pouch hydration, sample and buffer mixing, addition of sample/buffer to the pouch, and loading onto the instrument.

### TABLE 3 Details of discordant samples

| Sample no. | Consensus result | Individual test result* | Fusion (Ct value) | Aptima | BioFire (gene target) |
|------------|------------------|-------------------------|------------------|--------|----------------------|
| GSD-3      | Positive         | Positive (38.1)         | Negative         | Detected (2d/2e) |
| GSD-4      | Positive         | Positive (38.1)         | Negative         | Equivocal (2a) twice, overall detected |
| GSD-6      | Positive         | Positive (40.5)         | Negative         | Detected (2a/2e) |
| GSD-7      | Positive         | Positive (35.7)         | Positive         | Not detected |
| GSD-23     | Positive         | Negative (NA)           | Positive         | Detected (2a/2e) |
| GSD-48     | Positive         | Positive (37.3)         | Negative         | Equivocal (2a) twice, overall detected |

*Discordant sample results are highlighted in bold. NA, not applicable.

### TABLE 4 Basic assay characteristics and workflow parameters of three EUA SARS-CoV-2 NAATs*

| Characteristic                          | Fusion                  | Aptima                  | BioFire                  |
|-----------------------------------------|-------------------------|-------------------------|--------------------------|
| Detection platform/system                | Panther Fusion          | Panther or Panther Fusion | BioFire FilmArray Torch system, 12-bay tower |
| Sample type                             | NP, OPS, LRT, NS        | NP, OPS, NS, nasal wash/aspirate | NP |
| Sample vol required                     | 500 μL (250 μL for LRT) | 500 μL                  | 300 μL                   |
| Target region of SARS-CoV-2             | Two regions of ORF1ab   | Two regions of ORF1ab   | Two regions of ORF1ab and ORF8 |
| Analytical sensitivity per claim         | 0.01 TCID<sub>50</sub>/ml | 0.01 TCID<sub>50</sub>/ml | 0.022 TCID<sub>250</sub>/ml (330 GE/ml) |
| Observed analytical sensitivity for inactivated virus | 1,000 GE/ml           | 500 GE/ml               | 500 GE/ml               |
| Observed analytical sensitivity for RNA transcript | 62.5 copies/ml         | 62.5 copies/ml          | 125 copies/ml           |
| High-throughput processing              | Yes                     | Yes                     | No                       |
| Throughput (no. of samples per run)     | 120 (with continuous loading) | 120 (with continuous loading) | 12 |
| HoT/sample                              | 1 min                   | 1 min                   | 3 min                    |
| HoT/run                                 | 2 h/120 samples         | 2 h/120 samples         | 36 min/12 samples        |
| Assay processing time/run               | 4 h 35 min/120 samples  | 5 h 30 min/120 samples  | 45 min/12 samples        |
| Time to first result                    | 2 h 25 min              | 3 h 30 min              | 45 min                   |
| Overall TAT/run                         | 6 h 35 min/120 samples  | 7 h 30 min/120 samples  | 1 h 21 min/12 samples    |
| Maximum sample throughput in 8 h/24 h   | 335/1,150               | 275/1,020               | 72/216                   |

*NP, nasopharyngeal swab; NS, nasal swab; OPS, oropharyngeal swab; LRT, lower respiratory tract; HoT, hands-on time; TAT, turnaround time.
These additional steps for BioFire account for the increased HoT per specimen, as shown in Table 4.

**DISCUSSION**

The present study compared the analytical sensitivities (LODs), clinical performance, and workflows of three SARS-CoV-2 NAATs for 150 NP swab specimens. Our two independent LOD analyses revealed that while all three assays had an LOD that was within 1 dilution factor of each other within a given control material, absolute LODs with quantified, inactivated virus (500 to 1,000 GE/ml) were severalfold higher than the value obtained when using quantified synthetic RNA (62.5 to 125 copies/ml). This difference in absolute LOD values reflects the inherent difficulty in comparing standards that have been prepared and quantitated differently and that have very high stock concentrations requiring significant dilutions for LOD panel testing (i.e., any quantification error and/or pipetting error of the stock will be magnified in a dilution series). Of interest, the LOD calculations for BioFire using the synthetic RNA standard showed amplification in only two of three gene targets, since ORF1ab is included in the Exact Diagnostics control but ORF8 is not included. Considering this discrepancy, the increased LOD of the Biofire assay relative to the Fusion and AptaIma assays when using the quantified synthetic RNA may possibly be due to the lack of ORF8 target material. Our clinical correlation showed that the Fusion and Biofire assays had similar PPAs (98.7%), while the Aptima assay showed a slight decrease in PPA, at 94.7%, due to four false-negative results among samples in the frozen retrospective set. The differences seen among all three assays and the consensus result were not statistically significant when analyzed using McNemar’s chi-square test. Cohen’s kappa values also showed almost perfect agreement (range, 0.947 to 0.987) between each assay and the consensus result. NPAs were 100% for all NAATs, suggesting that each of these assays has high specificity.

All discordant results were false negatives compared to the consensus result, with Fusion and BioFire each demonstrating one missed positive and Aptima demonstrating four missed positives. A closer analysis of discordant results showed that the sample missed by the Fusion assay (GSD-23) had only two gene targets detected by BioFire (2a/2e), and the sample missed by the Biofire assay (GSD-7) had a C<sub>t</sub> value of 35.7 by the Fusion assay. All four samples missed by the AptaIma assay (GSD-3, -4, -6, and -48) had C<sub>t</sub> values ranging from 37.3 to 40.5, suggesting that the Aptima assay may be slightly less sensitive than the Fusion and Biofire assays. Of note, two of the specimens missed by the AptaIma assay (GSD-4 and GSD-48) were also each equivocal twice by the BioFire assay and resulted as positive per the EUA instructions for use (IFU), suggesting that these were also weak positives by the BioFire assay.

Comparisons of workflows between the Fusion and the AptaIma assays showed that they were quite similar, both being optimal in high-volume testing situations (>1,000 tests/24 h), whereas the BioFire assay had an advantage of fast sample-to-answer time, allowing for the faster detection and diagnosis that is useful in urgent situations.

Other comparator studies of the Fusion assay have recently been reported, including the following: a study comparing the Fusion assay to the modified CDC assay, GenMark ePlex, and DiaSorin Simplexa (6); a study comparing Cepheid, ID NOW, and GenMark assays using Fusion as the reference standard (7); a study comparing Fusion, Cepheid, DiaSorin, and cobas assays and a laboratory-developed test (LDT) (8); and a study comparing Fusion to an in-house LDT targeting the envelope gene (9). In general, these reports show that the Fusion SARS-CoV-2 assay is comparable, if not superior, to other molecular platforms available for SARS-CoV-2 testing. In particular, these evaluations showed that the Fusion, DiaSorin, and Cepheid assays tended to have better analytical sensitivity and fewer false negatives in clinical testing than other assays. The low rate of false negatives is a critical performance aspect, as missed positive results allow for further spread of the disease and potential patient mismanagement. These comparison studies, while not clinical trials as would be typical of new FDA-cleared in
vitro diagnostic assays, have provided a window into the relative real-world clinical performance characteristics of these assays, which has been a significant knowledge gap since these SARS-CoV-2 EUA molecular assays first became available. Among these knowledge gaps are relative performance data for the Aptima and BioFire tests. To our knowledge, this is the first report comparing both the Aptima and BioFire assays to any other NAAT for SARS-CoV-2 detection.

It is worth noting that the LOD for the Fusion SARS-CoV-2 assay determined in this study was 1,000 GE/ml when inactivated virus was used, while the LOD determined in a previous study by our laboratory was 83 ± 36 copies/ml (6) and an additional LOD study performed as part of this study, once again using the same synthetic quantified RNA standard, also previously published, showed a comparable LOD of 62.5 copies/ml. This is an important additional new set of data that exhibits the difficulty of comparing absolute LOD values when using different standards, specifically inactivated virus versus synthetic RNA standards.

This study does have limitations, being a single-site study with a limited number of NP swab specimens (n = 150) included in the clinical evaluation. However, this specimen set was selected to be representative of our patient population (including samples selected without bias toward specific age or gender) and positivity rate (50 to 60% in the beginning of April 2020) and included samples with a range of viral loads (low, moderate, high [see Fig. S1 in the supplemental material]). One additional potential limitation is that the Fusion assay was used to select positive results, but any bias introduced by this selection should be minimal, considering that the proportion of positives analyzed mirrored our true positivity rate at the time of this study. Some initial equivocal or invalid results (n = 3) by BioFire and Fusion required retesting, but only one remained unresolved on retesting and was removed from the agreement analysis for the Fusion assay.

In conclusion, our data show that the Fusion, Aptima, and BioFire SARS-CoV-2 NAATs exhibit similar LODs, even when tested using two different quantified standards. In addition, the Fusion and BioFire assays have comparable clinical performance for detection of SARS-CoV-2 in NP swabs, with a PPA of 98.7%, while the Aptima assay showed a slightly lower PPA, at 94.7%. All three assays demonstrated 100% NPA, suggesting high specificity. These performance characteristics, as well as testing volume and workflow requirements, should be considered when making testing platform decisions.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.4 MB.

ACKNOWLEDGMENTS

Gregory Berry has previously given education seminars for Hologic, Inc., and BioFire Diagnostics and has received honorariums.

REFERENCES

1. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X, Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM, Pilail SK, Washington State 2019-nCoV Case Investigation Team. 2020. First case of 2019 novel coronavirus in the United States. N Engl J Med 382:929–936. https://doi.org/10.1056/NEJMoa2001191.

2. Centers for Disease Control and Prevention. 2020. Coronavirus disease 2019 (COVID-19) situation summary. Centers for Disease Control and Prevention, Atlanta, GA. Updated 19 April 2020. https://www.cdc.gov/coronavirus/2019-ncov/coronavirus/2019-ncov/situation-summaries.html. Accessed 13 May 2020.

3. John Hopkins University. 2020. Coronavirus COVID-19 global cases by the Center for Systems Science and Engineering at John Hopkins University. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6. Accessed 18 June 2020.

4. Centers for Disease Control and Prevention. 2020. Coronavirus disease 2019 (COVID-19): people who are at higher risk for severe illness. Updated 15 April 2020. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html. Accessed 13 May 2020.

5. WHO-Joint Mission on Coronavirus Disease. 2020. Report of the WHO-China Joint Mission on Coronavirus Disease. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf. Accessed 13 May 2020.

6. Zhen W, Manji R, Smith E, Berry GJ. 2020. Comparison of four molecular in vitro diagnostic assays for the detection of SARS-CoV-2 in nasopharyngeal specimens. J Clin Microbiol. https://doi.org/10.1128/JCM.00743-20.

7. Zhen W, Smith E, Manji R, Schron D, Berry GJ. 2020. Clinical evaluation of
three sample-to-answer platforms for the detection of SARS-CoV-2. J Clin Microbiol https://doi.org/10.1128/JCM.00783-20.

8. Lieberman JA, Pepper G, Naccache SN, Huang M, Jerome KR, Greninger AL. 2020. Comparison of commercially available and laboratory developed assays for in vitro detection of SARS-CoV-2 in clinical laboratories. J Clin Microbiol https://doi.org/10.1128/JCM.00821-20.

9. Hogan CA, Sahoo MK, Huang C, Garamani N, Stevens B, Zehnder J, Pinsky BA. 2020. Comparison of the Panther Fusion and a laboratory-developed test targeting the envelope gene for detection of SARS-CoV-2. J Clin Virol 127:104383. https://doi.org/10.1016/j.jcv.2020.104383.

10. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174. https://doi.org/10.2307/2529310.

11. McHugh ML. 2012. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22:276–282.

12. R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.