SARS-CoV-2 *in vitro* diagnostics overview

Effective February 4th 2020, Human Health and Services (HHS) declared a public health emergency and determined that circumstances justify the emergency use authorization (EUA) of *in vitro* diagnostics (assays and devices) for the detection and/or diagnosis of SARS-CoV-2, the virus responsible for the COVID-19 pandemic. This is a temporary designation for clinical use granted by the FDA in an emergency. As such, these assays are not FDA-approved/cleared.

The aim of this communication is to build a common understanding of SARS-CoV-2 testing strengths, limitations and available approaches. Due to the novelty of the virus, it is important to note that we are still in the learning stages concerning the best specimen types, timing of collection in relation to course of disease and assay selection for optimal SARS-CoV-2 detection.

Direct detection

Molecular tests

Molecular-based testing is the preferred methodology for diagnosis due to the danger and slow turn-around-time of SARS-CoV-2 culture. In a non-pandemic setting, the FDA requires both a clinical and analytic evaluation to establish performance. Under a EUA, diagnostic test kits are not subject to clinical diagnostic performance trials or multi-site evaluations due to the emergent need to expand testing capacity. Therefore, the clinical sensitivity (true positive rate), specificity (true negative rate) and corresponding likelihood ratios are not applicable due to the novelty of the SARS-CoV-2 virus and the absence of both a clinical reference standard (comparator method) and clearly defined disease state.

In the absence of a recognized reference standard, labs evaluate clinical test performance by determining the percent agreement between the new test and authorized nucleic acid amplification test (NAAT) in a known set of positive and negative samples, expressed as positive percent agreement (PPA) and negative percent agreement (NPA). The overall percent agreement (diagnostic accuracy) is reported in lieu of sensitivity-specificity pair (PPA-NPA).

To establish the analytic performance, the FDA requires analytical sensitivity and specificity studies prior to clinical use. Analytic sensitivity or the limit of detection (LoD) determines the smallest amount of viral RNA (copies/mL) the assay can detect in 95 out of 100 tests (95% of the time). Analytic specificity evaluates cross-reactivity and presence of interfering substances. These analytical measures differ in meaning from clinical sensitivity and specificity, the latter of which cannot be established without a representative test population and clearly defined disease state.

Factors that influence molecular test performance and interpretation

Variables in specimen collection, such as anatomic site and timing of collection, along with the assay's LoD are the main factors that influence SARS-CoV-2 test performance (Figure 1). Nasopharyngeal (NP) swabs are the preferred upper respiratory specimen type per the Center for Disease Control and Prevention (CDC) and intended for use in symptomatic patients. Several studies have attempted to identify which upper respiratory specimen types have the highest viral load and at what point in the course of infection, albeit with sometimes conflicting results. Collectively, nasopharyngeal followed by nasal and throat swabs harbor the highest viral loads during the first 5 symptomatic days but detection declines significantly thereafter, particularly 8 days following symptom onset. The LoD varies depending on the commercial assay, how it was determined (intact virus or *in vitro* transcribed), specimen type and symptom onset. Studies indicate that NP swabs have a mean viral load in the range of 10^5 to 10^6 copies/mL of viral RNA, especially during early symptomatic infection and may be similar between asymptomatic and symptomatic infection. Thus, being asymptomatic may not increase the odds of having a false negative result when tested using an NP swab. Even less sensitive assays with a LoD in range of 1000(s) copies/mL should be easily detected during acute infection and minimize the chance of false negative results with a properly timed and collected specimen. Although some studies suggest that saliva has a high viral load near presentation, clinical labs are still awaiting guidance from the FDA to ascertain the value of saliva as a reliable specimen type for diagnosis.

As the course of illness progresses, lower respiratory tract secretions (BAL, sputum) become the best specimen type (Figure 1). For patients receiving invasive mechanical ventilation, collect a lower respiratory tract specimen when possible. The induction of sputum is not recommended.

Performance characteristics of PCH in-house molecular SARS-CoV-2 Tests

Currently, we are using a combination of 3 different in-house molecular platforms: DiaSorin, BioFire and Cepheid (coming soon). Independent verification/validation of claimed manufacturer performance is achieved using sets of known positive and negative patient samples. A pre-specified performance goal of ≥ 95% overall agreement is required for implementation. All in-house platforms have met or exceeded this criterion and we have rejected implementation of several other platforms due to poor performance. The analytic sensitivity or limit of detection is <u>excellent</u> for all platforms (DiaSorin: 250 cp/mL, BioFire: 500 cp/mL and Cepheid: 250 cp/mL). This is equivalent to dropping a skittle in a swimming pool and you can still taste it. Therefore, the probability of a false negative result due to analytical assay performance with our in-house platforms is extremely low.

False negative rates: facts, folklore and something in-between

Reports of 20-30% false negative rates for RT-PCR are derived mostly from pooled data collected from multiple studies, across various parts of the world, and compiled to fit a statistical model. Several of the pooled data sets came from Chinese studies where the high false negative rate is likely attributable to a combination of poor assay design and performance, coupled with a different SARS-CoV-2 strain (Wuhan strain) than the U.S. Due to heterogeneity in study designs and imprecise nature of pooled results, statistical modeling studies should not be superimposed broadly onto all molecular-based assay results or considered dogma for test interpretation. Instead, these modeling studies highlight that the predictive value of diagnostic tests vary with time from exposure and symptom onset. However, it is important to remember these studies do not represent a blanket failure of molecular testing.

False positive rates: facts, folklore and something in-between

Beyond day 8-10 of symptom onset, attempts to culture SARS-CoV-2 have been unsuccessful suggesting a decline in infectivity beyond the first week of illness. Further evidence suggests that persons with mild to moderate COVID-19 are no longer infectious (replication-competent) at 10 days post-symptom onset. However, due to the high analytic sensitivity of molecular-based tests, viral RNA can be detected in recovered persons up to 3 months following symptom onset. Therefore, CDC recommends a symptoms-based approach to demonstrate inpatient recovery. Isolation precautions can be discontinued 10 days after symptom onset, resolution of fever ≥24 hours and obviates the need for two consecutive negative RT-PCR results. However, prolonged shedding can occur in the severely immunosuppressed and a test-based strategy should be considered in consultation with infectious disease experts.

Other non-molecular-based tests

FDA recommends selecting a comparator assay that has an established high analytic sensitivity. Thus, molecular-based tests become the recognized “gold-standard” as a comparator method for test performance and enable estimates of diagnostic accuracy (sensitivity-specificity pair or PPA-NPA pair). Although the U.S. prevalence of SARS-CoV-2 positive individuals is unknown, a contrived prevalence can be used to estimate the positive predictive value (PPV) and negative predictive value (NPV).

Rapid antigen tests

SARS-CoV-2 rapid antigen-based test kits exhibit suboptimal sensitivity and specificity. The overall sensitivity of direct antigen-based testing hovers around 50% with significant heterogeneity in assay performance depending upon the evaluated test kit and viral burden. Significant concerns with the false negative rate (15% sensitivity with low viral load samples) due to either low or variable viral burden in patient samples. Consequently, FDA recommends that negative results reflex to a molecular-based test due to the high risk of false negative results. Using a contrived prevalence, the estimated PPV is around 46% meaning these tests are very prone to false positive results.

Clinical utility of rapid antigen tests

In response to severe supply shortages, some hospitals are utilizing rapid antigen tests to screen asymptomatic peri-operative patients. The rationale is that molecular-based tests are too sensitive, thus patients with viral burdens below the level of transmissibility would not be detected when using a test with lower sensitivity. However, clinical hospital labs should exercise caution when considering this diagnostic testing option since we do not know the minimal viral burden associated with viral transmission.

Indirect detection

Serologic tests

In contrast to nucleic acid testing, which directly detects the virus, serology tests are used to identify exposure and antibodies to SARS-CoV-2. Serologic testing is retrospective in nature and should not be used to establish the presence or absence of SARS-CoV-2 infection or reinfection. In certain scenarios, serology can be used in conjunction with but not in place of a molecular test.

Factors that influence serologic test performance and interpretation

By using nucleic acid amplification tests as the "gold-standard" or reference comparator (assume a positive PCR result = COVID), estimates of diagnostic accuracy can be used to describe the performance of serology tests. The sensitivity of serologic tests are highly dependent on time since symptom onset (>14 days is optimal; 95% will be positive) and age (titers are weaker <40 yrs). Development of a detectable antibody response occurs somewhere between 5-10 days. IgM may be negative (non-reactive) if timing of symptom onset is incorrect, thus the recommendation is to test between 10-14 days. Furthermore, the temporal kinetics of IgM and IgG SARS-CoV-2 antibodies overlap (IgM/IgG appear concurrently) and do not have clearly defined intervals, thus confounding the interpretation of an IgM result (Figure 1). The recommendation is to test IgG alone where >95% symptomatic individuals test positive by day 14 following symptom onset or total Ab (IgM/IgG) where 90% symptomatic individuals test positive by days 11-24 to determine exposure.

By testing serum from uninfected individuals, studies suggest that COVID-19 antibody tests have high analytic specificity and the probability of a false positive due to cross-reactivity with other coronavirus antibodies is relatively low.

Prevalence is the other the main factor that influences serologic test interpretation. Currently, we do not know the U.S. prevalence of SARS-CoV-2 antibody positive individuals. Under a serology EUA, the FDA bases estimates for predictive value (PPV and NPV) calculations on an assumed prevalence of 5%. Consequently, the predictive values stated by manufacturers are truly estimated claims based on a contrived prevalence. CDC recommends employing two independent antibody tests in sequence for a population with a 5% prevalence to increase the PPV of a single test from 49% to 95% to mitigate the risk of false positive results.

Clinical utility of serology tests

Antibody-based tests are not recommended for primary diagnosis, except in rare instances when patients present late with a viral load below the LoD, and when sampling of the lower respiratory tract is not possible.

Pediatric populations and individuals with mild SARS-CoV-2 infection may not mount a detectable or effective immune response. Immunosuppressed patients will likely test negative.

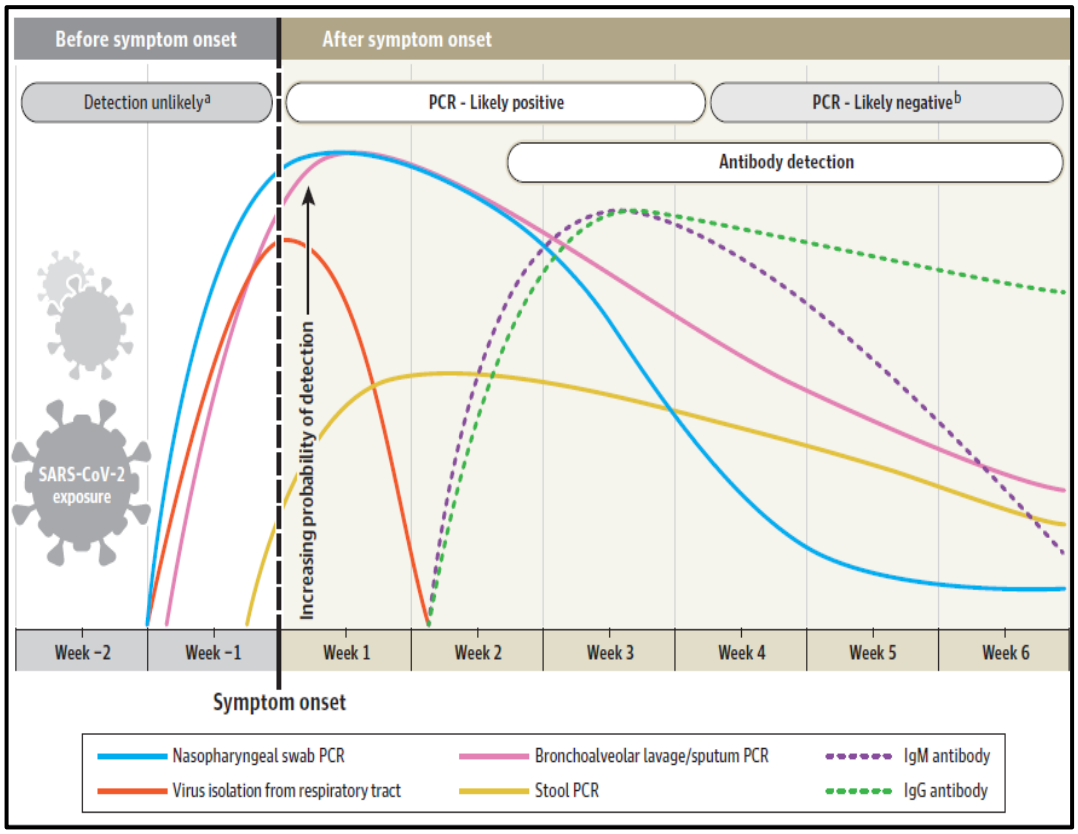
Remember, when testing a population with low prevalence, regardless of the high sensitivity of a test, the probability of false positive results increases due to the low positive predictive value of the test.

Antibody-based testing does not identify individuals with neutralizing antibodies or protective immunity.

Negative results do not rule out and false negative results can be associated within the first 2 weeks of infection and influenced by time of collection (viral exposure and symptom onset).

Thus, the utility of SARS-CoV-2 antibody testing is at the community rather than individual level and best suited for public health sero-surveillance studies and for the identification of convalescent plasma donors.

Figure 1



Estimated time intervals and rates of viral detection are based on data from several published reports. Because of variability in values among studies, estimated time intervals should be considered approximations and the probability of detection of SARS-CoV-2 infection is presented qualitatively. SARS-CoV-2 indicates severe acute respiratory syndrome coronavirus 2; PCR, polymerase chain reaction.

[a] Detection only occurs if patients are followed up proactively from the time of exposure.

[b] More likely to register a negative than a positive result by PCR of a nasopharyngeal swab.

Summary of key points for interpretative aid

No test is 100% perfect or accurate (no false positive/negative results).

Proper specimen collection, specimen type (upper and/or lower respiratory tract), time of collection in relation to course of disease is key to having accurate test performance and interpretation.

SARS-CoV-2 tests are not FDA-approved/cleared but authorized for emergency use only.

Viral RNA-based (molecular) testing is the preferred method of detecting and diagnosing SARS-CoV-2.

In the absence of a clinical reference standard for diagnosing COVID-19, the true clinical sensitivity and specificity of SARS-CoV-2 diagnostic testing remains unknown.

- Molecular-based assay performance is determined by overall percent agreement with a comparator method and limit of detection (analytic sensitivity).
- RT-PCR is considered the "gold-standard" reference method for which all other assay types (direct antigen or antibody-based) are tested against and thus, estimates of diagnostic accuracy (sensitivity-specificity pair) and predictive values (PPV, NPV) are generated under the assumption molecular tests are correctly identifying the true disease state.

Reports of 20-30% false negative rates for RT-PCR are derived mostly from pooled data collected from multiple studies that are compiled to fit a statistical model and should not be interpreted as dogma and broadly applied to all molecular-based assays. These statistical models should be used to highlight the interpretative value of diagnostic tests vary with time from exposure and symptom onset and does not represent a blanket failure of molecular testing.

Treat the patient not the test. If clinical suspicion is high, infection cannot be ruled out based on RT-PCR alone and test results should used in conjunction with the clinical and epidemiological situation.

NP swabs, the preferred specimen type, have a mean viral load in the range of 10^5 to 10^6 copies/mL of viral RNA, which may be similar between asymptomatic and symptomatic infection. This is well above limit of detection (analytic sensitivity) of most RT-PCR assays. Thus, the chance of false negative results with a properly collected and timed specimen is very low.

A positive PCR result reflects the detection of viral RNA but does not indicate viable virus. Evidence suggests that persons with mild to moderate COVID-19 are no longer infectious (replication-competent) at 10 days post-symptom onset.

Rapid antigen-based SARS-CoV-2 test kits exhibit poor sensitivity and specificity and clinical hospital labs should exercise caution when considering this diagnostic testing option.

Antibody-based tests are not recommended for primary diagnosis, except in rare instances when patients present late with a viral load below the LoD, and when sampling of the lower respiratory tract is not possible. In this scenario, it is important to remember that serology can be used in conjunction with but not in lieu of a molecular test.

The temporal dynamics of IgM and IgG SARS-CoV-2 antibodies overlap (IgM/IgG appear concurrently) and do not have clearly defined intervals, thus confounding the interpretation of an IgM result. Therefore, the recommendation is to test IgG alone where >95% symptomatic individuals test positive by day 14 following symptom onset or total Ab (IgM/IgG) where 90% symptomatic individuals test positive by days 11-24 to determine exposure.

Antibody-based testing does not identify individuals with neutralizing antibodies or protective immunity. The level and duration of protective immunity against re-infection remains unknown. Thus, return of healthcare workers based on serologic testing may give a false sense of security.

The utility of antibody-based testing is at the community rather than individual level and best suited for public health surveillance.