COVID-19 Evidence Accelerator Collaborative

Lab Meeting #14

Thursday, July 23, 2020, 3:00-4:00 pm ET

Call Summary

Introduction to Lab Meeting 14

The theme of this week's lab meeting was "the implications of COVID-19 diagnostic testing accuracy in COVID-19 research and how to manage it." As the COVID-19 pandemic has evolved, members of the public health community have continued to point out the necessity of diagnostic testing for improving understanding of COVID-19 and its impact. As the community begins to have comparative discussions, however, it will be ever more important to have a broad understanding of diagnostic testing factors such as the availability of testing, the reliability of testing, the timepoint of testing within the disease course, and the performance of tests in the real-world before conclusions can be appropriately made.

The lab meeting included a total of five presentations and discussions from members of the diagnostic community:

1. Presentation on A Primer on Diagnostics for COVID-19 (William Morice, Mayo Clinic; Nilay Shah, Mayo Clinic)
2. Discussion on Linking the Test to the Results to the Clinical Dataset (Ed Michael, Lionbird)
3. Presentation on Imperfect SARS-CoV-1 RNA Test Status and Implications for RWE Research (Sebastian Schneeweiss, Harvard)
4. Discussion on the Implications for Clinical Trial Patient Selection (Rob Califf, Alphabet)
5. Presentation on the Implications of COVID-19 Diagnostic Test Accuracy with an FDA Lens (Bob Ball, FDA)

Lab Meeting Presentations

Presentation on A Primer on Diagnostics for COVID-19

- When talking about laboratory testing performance metrics, we often hear about analytical performance and clinical performance. In the context of the COVID-19 pandemic, it is more important to think in terms of the clinical performance of the test.
- Currently, there are a number of different testing options for SARS-CoV-2, each with its own performance considerations as outlined in the below table:

	Diagnostic RT-PCR Swab Test	Enzyme-linked Immunosorbent Assay	Rapid Serology Antibody Test
Sample	Nasopharyngeal (NP) swab sample	Serum or plasma sample	Serum or plasma sample; finger prick
Result	Detection of viral SARS-Cov-2 RNA	Detection of IgM/IgG antibodies	Detection of IgM/IgG antibodies
Advantages	Gold standard diagnostic test; most accurate results early in disease presentation; directly detects virus presence	Accurate detection of seroconversion in laboratory setting several days after onset or sooner	Point-of-care or at home collection; rapid results
Limitations	Labor and reagent intensive; **sensitive to sample collection error**; loses accuracy after ~5 days of symptom onset	Laboratory-based testing; assessing cross-reactivity with other immune response	Variability in test sensitivity and specificity
Performance of available tests	Sensitivity: 75-100% Specificity: 98-100% PPV: 47-100%	Sensitivity: 85-100% Specificity: 96-100% PPV: 59-100%	Sensitivity: 88-100% Specificity: 95-100% PPV: 50-100%

- There are multiple ways to think about the performance of a diagnostic test. In the COVID-19 context, when we think about a "gold standard" the key measure being considered is typically the sensitivity of the test.
    - Sensitivity measures the ability of a test to correctly identify a person with the disease (true positive).
    - Oppositely, specificity measures the ability of a test to correctly identify a person without the disease (true negative).
- The ability of diagnostic tests to detect COVID-19 can be influenced by the timing of the test and amount of time which has elapsed since a person was infected.
- Other, non-routine testing options for SARS-CoV-2 include rapid antigen tests, point-of-care PCR tests, saliva PCR tests, and IgM/IgA serology tests.
    - Rapid antigen tests can produce rapid results, are easy to use, and are typically of lower cost, but often have lower sensitivity as compared to PCR tests.
    - Point-of-care PCR tests can produce rapid results but often have lower test throughput and variable sensitivity across tests.
    - Saliva PCR tests are easier and more convenient for the patient to collect compared to a nasopharyngeal swab but have lower sensitivity.
    - IgM/IgA serology tests may allow for earlier detection of those who have been exposed to SARS-CoV-2 but typically have lower specificity.
- All diagnostic tests are measuring a dynamic process with biologic variability. The performance metrics of these tests are influenced by numerous factors unrelated to the test itself such as the availability of tests and when in the course of disease, a patient was tested.
- Ongoing initiatives in the SARS-CoV-2 diagnostic community include:
    - Quantitative SARS-CoV-2 PCR—this quantitation may be an important correlate to which COVID-19 patients are infectious.

- High throughput, high sensitivity SARS-CoV-2 antigen test—these methods would lend themselves well to automation and alleviate supply chain constraints which limit PCR testing
- Rapid POC molecular SARS-CoV-2 detection—the time from detection to informing an individual is an important time period in terms of containment.
- Neutralizing antibody analysis
- Measures of immune competence/response

Discussion on Linking the Test to the Results to the Clinical Dataset

- There are a vast number of diagnostic tests with different performance metrics entering the community.
- Even within categories of tests such as "PCR tests," there are a number of differentiating factors such as probe and primer sets, sampling, transport media, and handling instructions.
- The sensitivity of tests can vary significantly over the time elapsed since the onset of symptoms.
- We do not know enough about the immune response to COVID-19. While we may be able to say with confidence what the analytical performance metrics of an antibody test are, we cannot say much about the clinical performance of these tests.
  - What does it mean to have antibodies present? Does it mean there has been an immune response? Do they provide protection?
  - Does the quantity of antibodies matter?
  - Does immunity go away over time?

Discussion Insights

As we think about the interplay between test accuracy and clinical studies that incorporate data that is reliant on these diagnostic tests, there is a key task to make sure that information about these tests (e.g. test used, manufacturer, lot number and expiration date, quality controls) is embedded into our datasets. These details can help one explore what might have gone wrong if results start to look funny. There is an opportunity to begin collecting and incorporating this data as diagnostic test demand continues to increase.

Presentation on Imperfect SARS-CoV-1 RNA Test Status and Implications for RWE Research

There are a variety of research questions that rely on SARS-CoV-2 test status and thus SARS-CoV-2 diagnostic tests. What are the implications of having an imperfect test?

- SARS-CoV-2 testing to identify the <u>study population</u> to evaluate therapeutic effectiveness: want to make sure all participants are truly positive before treatment.
  - Ideally 100% positive predictive value (PPV) = 0% false positives: false identification of patients will dilute the signal and bias the results.
  - High sensitivity helps efficiency of trial recruitment
- SARS-CoV-2 testing to identify the <u>study population</u> to evaluate vaccine effectiveness: want to make sure all participants are truly negative before treatment.

- o Ideally 100% negative predictive value (NPV) = 0% false negatives.
- o If disease prevalence is low, a reasonable specificity will suffice to achieve high NPV.
- o We would also assume IgG sero negativity.
- SARS-CoV-2 testing to identify <u>newly occurring cases</u> for a vaccine research question: need to rely on accurate testing to identify all new cases.
  - o Ideally 100% specificity
  - o High sensitivity improves statistical efficiency
  - o Even if sensitivity is less than 100% but specificity = 100%, the relative risk estimates are still unbiased.
- SARS-CoV-2 testing to identify <u>newly occurring cases</u> for disease surveillance: need to rely on accurate testing to identify all new cases.
  - o 100% specificity and 100% sensitivity are important
  - o The disease prevalence, frequency, and comprehensiveness of testing strategy may matter as much if not more than the actual test performance.
  - o Knowing test performance allows for correction of test bias.
- SARS-CoV-2 testing to determine <u>exposure status</u> for natural history of disease study: need to rely on accurate testing to distinguish between COVID+ and COVID-.
  - o 100% specificity and 100% sensitivity
  - o Random misclassification of exposure leads to an underestimation of relative risk estimates
  - o In a low sensitivity situation, some classified as COVID-19 negative will actually have the virus, so historical controls from before the COVID-19 pandemic are useful.
- For most COVID-19 research questions that we have, specificity is most important. The unknown variation of test performance characteristics between manufacturers even under optimal circumstances is a concern. In real world setting additional performance variations comes from differences in local disease prevalence, specimen collection, test kit handling.
- In terms of next steps, we need accurate performance characteristics to correctly study findings that rely on SARS-CoV-2 testing (under optimal circumstances, by manufacturer, over time as technology evolves).
  - o Manufacturers should make information publicly available.
  - o The performance of diagnostic tests is almost certainly worse in the field than as reported by manufacturers under ideal circumstances. Therefore, need to know actual test performance in the field.
  - o Guidance on the analytic methods to correct study results based on SARS-CoV-2 test misclassification would benefit all

Discussion on the Implication for Clinical Trial Patient Selection

- When thinking about prospective studies, there are six key factors that matter in clinical trials:
  - o Did you enter the intended population?
  - o Was randomization done as intended?
  - o Did people get the treatment as expected when assigned?
  - o Did you measure outcomes that matter?
  - o Did you measure known important risk factors and subgroup indicators
  - o Was follow-up complete?

- Sample size can be used to help decrease the impact of false test results if the bias is random.
- When the operating characteristics of diagnostic tests change, trials must be adjusted. Two examples from cardiology include:
  - Introduction of EKG as a critical condition for entering trial became important to guarantee that patients were not being harmed by giving them therapies where there is likely a chance for risk and no benefit.
  - A more recent example is the measurement for risk of heart attack as an entry criteria and outcome. The number of people with "positive test" for a heart attack magnified many-fold after a change in test characteristics which had profound implications for trial entry and interpretation of endpoints.

Discussion Insights

Diagnostic test performance metrics such as sensitivity and specificity should not be thought of as fixed characteristics but as characteristics which are impacted by a number of factors and and have substantial variation in the real-world.

Presentation on the Implications of COVID-19 Diagnostic Test Accuracy with an FDA Lens

Overarching approach

- The FDA's overarching approach for the management of COVID-19 diagnostic testing accuracy in COVID-19 therapeutic research involves a series of steps:
  - Describe
    - Transparent description of the problem (e.g. limitations)
    - Identify problems to be solved
  - Analyze
    - Develop analysis approaches for RWD that accommodate varied sensitivity/specificity
    - Determine real-world performance of diagnostic tests by manufacturer/test
  - Data
    - Incorporate discrete test information in RWD sources including performance data
    - Dynamically incorporate test performance into analyses
- Traditionally diagnostic testing accuracy has not been a major consideration in RWE generation.
- Recognition of importance of valid tests has been emphasized in the COVID-19 era in the context of RWE generation and current limitations. For example:
  - Among those who are tested, false negatives may occur due to imperfect test performance, the timing of the test, or the technique of the person administering the test.
  - Performance characteristics of diagnostic tests are not currently close to "perfection," vary by type (e.g. PCR vs. antibody) and manufacturer of test, and will likely improve over time

- We should be transparent about these limitations and begin work to understand and improve the current state.
- The main implication for having imperfect diagnostic testing for RWE studies of therapeutic effectiveness is misclassification of cases.
  - This can bias results toward the null and indicate a smaller effect size than present.
  - If misclassification is different in different arms, bias will result.
- For RWE studies of therapeutic effectiveness, false negatives are primarily of concern.
  - Unknown if those who falsely test negative are different from those who test positive in such a way that might bias any results.
  - If only patients who test positive are being treated with drugs specific for COVID-19, it might be worth considering restricting RWE studies to those with a positive test.
- The ability to identify asymptomatic cases is less important for RWE studies of therapeutic effectiveness because these patients are generally not treated but might be important for studies of prevention (e.g., vaccines).
- The FDA is working on an algorithm validation project through its Sentinel System. The purpose is to validate five ICD-10-based algorithms, representing different combinations of ICD-10 diagnosis codes, for identification of hospitalized COVID-19 cases in claims-based data sources using a positive diagnostic, molecular laboratory as the reference standard.
- Diagnostic test performance concerns for this initiative include:
  - Will the use of an imperfect reference standard result in incorrect accuracy estimates and lead to choosing the "wrong" algorithm?
    - Unlikely that the rate of false negatives will differ across algorithms.
  - Information on the specific diagnostic test product/manufacturer is not readily available in RWD sources, although LOINC codes are often available to identify the type of test
    - The underlying EHR data might have information on the test manufacturer, but will require further exploration
- Next steps for the FDA include:
  - Create linkage between data on tests (e.g. manufacturer, performance characteristics) and clinical data
  - Create cohort of patients with diagnostic testing, serology, and clinical information and incorporate this information into analyses

---

Discussion Insights

In the context of observational studies, epidemiological studies, or every day decision making, the consequences of a false negative test or how to interpret test results in low prevalence environments becomes very complicated.

- Not only requires analytical operating characteristics but clinical operating characteristics as well.