Alternate Nucleic Acid Targets Can Be Used To Create a Composite Standard To Evaluate Clinical Performance of Nucleic Acid Amplification Tests

Julius Schachter,^a Max Chernesky^b

^aLaboratory Medicine, University of California San Francisco, San Francisco, California, USA

^bPediatrics, Pathology and Molecular Medicine, McMaster University/St Josephs Healthcare, Hamilton, Ontario, Canada

ABSTRACT Evaluating the clinical performance of a new nucleic acid amplification test (NAAT) for *Mycoplasma genitalium*, B. Kirkconnell, B. Weinbaum, K. Santos, T. Le Nguyen, et al. (J Clin Microbiol 57:e00264-19, 2019, <https://doi.org/10.1128/JCM.00264-19>) created 3 alternate NAATs that detected other unique *M. genitalium* gene targets. Lacking a reference standard, they used the consensus of results with those 3 NAATs as the comparator. This approach could be a new paradigm to evaluate new NAATs when there is no previously defined reference standard.

Mycoplasma genitalium is an emerging pathogen. It has been clearly associated with lower genital tract infection and disease in men and women, and there are data highly suggestive of a role in upper genital tract disease in women such as pelvic inflammatory disease (PID) (1, 2). *M. genitalium* is quite prevalent (3). Because the organism is extremely difficult to culture, there is a need for a diagnostic test to allow clinicians and public health officials to identify and treat these infections.

Nucleic acid amplification tests (NAATs) have provided the most sensitive diagnostic assays for sexually transmitted infections for approximately 25 years. Hologic Inc. (San Diego CA) has developed a new assay (Aptima AMG) using transcription-mediated amplification (TMA) to detect an *M. genitalium* 16S rRNA target. They were faced with the challenge of assessing performance of the AMG assay, because there was no reference standard available for comparative purposes. Other new assays for emerging infections have used comparisons to laboratory-derived NAATs developed in government laboratories (i.e., Zika virus) or seroconversion of infected patients (i.e., severe acute respiratory syndrome coronavirus [SARS-COV]). *M. genitalium* culture is so difficult that it is completely useless for that purpose, and there was no FDA-cleared assay that could be used as a comparator. Facing this challenge, they developed a new approach for evaluating the performance of their *M. genitalium* assay.

Kirkconnell et al. designed and evaluated 3 alternate TMA (Alt TMA) assays for the detection of *M. genitalium* (4). The 3 Alt TMA tests detect unique regions of 16S rRNA (Alt TMA-1) or 23S rRNA (Alt TMA-2 and Alt TMA-3) *M. genitalium* gene sequences. Analytical sensitivity, specificity, and strain inclusivity of the Aptima *M. genitalium* and the 3 alternate tests were determined using 9 *M. genitalium* laboratory strains and a variety of nontarget organisms. Clinical performance was assessed by testing residual vaginal swabs plus female and male urine specimens from 1,400 adult subjects. The use of the consensus from the 3 Alt assays as a reference standard to establish infected specimen status demonstrated 100% sensitivity, 99.9% specificity, and 99.9% overall agreement with the now FDA-cleared Aptima AMG *in vitro* diagnostic (IVD) assay.

They note that this approach demonstrates the utility of a molecular composite reference standard methodology for the clinical evaluation of future assays for *M.*

Citation Schachter J, Chernesky M. 2019. Alternate nucleic acid targets can be used to create a composite standard to evaluate clinical performance of nucleic acid amplification tests. J Clin Microbiol 57:e00661-19. <https://doi.org/10.1128/JCM.00661-19>.

Editor Erik Munson, Marquette University

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Julius Schachter, Julius.Schachter@ucsf.edu.

For the article discussed, see <https://doi.org/10.1128/JCM.00264-19>.

The views expressed in this article do not necessarily reflect the views of the journal or of ASM.

Accepted manuscript posted online 29 May 2019

Published 26 July 2019

genitalium and other pathogens. We agree there is the potential for wider application of this approach. It is not one which should be limited to just *M. genitalium*, but is an approach that could be applied in many other settings.

Modern genomics has greatly influenced our ability to identify organisms associated with diseases that previously were of unknown etiology or to define the microbiologic environment associated with some syndromes of uncertain etiology (i.e., “omics”) (5). This genomic research in “pathogen discovery” has allowed identification of organisms that could not be cultured or even identified by other technology.

It is certainly appropriate to use new assays for research purposes to determine the presence of recently identified organisms, but once we are faced with the reality of management and treatment of patients based on the results of the assays, one must consider the performance of these tests. The sensitivity and specificity (and ultimately of positive and negative predictive values) of assays are important for laboratorians and to clinicians using the test results for decision making of treatment options.

While these problems evaluating diagnostic tests may be most obvious with newly discovered agents, it is also relevant for other infections where there is a problem of not having a suitable reference standard. In this context, it may be useful to review the historical situation as presented by the diagnosis of *Chlamydia trachomatis* infection using nonculture methods and discuss the various compromises that have been made in efforts to evaluate new assays. We have good experience with learning how to diagnose *C. trachomatis* infections, as industry has been introducing new assays for decades. *C. trachomatis* is the most common bacterial sexually transmitted pathogen, with the WHO estimating 131,000,000 new cases a year worldwide and CDC statistics indicating more than 1.7 million cases in 2017, making it the most common infection reported in the United States (6, 7). *C. trachomatis* infections are not only highly prevalent, they are clinically important. The organism is an important cause of urethritis in men, but more important, are the adverse reproductive consequences in women (PID, ectopic pregnancy, and tubal factor infertility). Identification of infection is immediately relevant, as highly effective antibiotic therapy is available. Given both the numbers and gravity of *C. trachomatis* infection, this has been an important target for industry in developing nonculture tests, because *C. trachomatis* isolation had its own challenges.

Historically, because *C. trachomatis* could not be cultured on artificial medium, cell culture systems were developed for isolation and propagation and became the original reference standard test for comparison. The isolation procedures are technically demanding, requiring centrifugation of inoculum and treatment of cells with cycloheximide to provide *C. trachomatis* with a metabolic advantage. Isolation attempts typically take 3 to 7 days to get a result based on microscopic visualization of inclusions. In the 1980s, several companies introduced antigen detection methods to diagnose *C. trachomatis* infection. Both direct fluorescent antibody tests (DFA) and enzyme immunoassays were widely used. When these immunoassays were compared to isolation by cell culture, it became obvious that the isolation procedure was a flawed reference standard, while highly specific cell culture was well less than 100% sensitive. Using different immunoassays with different antigenic targets, it became clear that at least some of the antigen-positive/culture-negative results actually were true positives (8).

In the 1990s, NAATs were introduced for detection of *C. trachomatis* infection of the genital tract and found many more positive results than were seen with either cell culture or antigen detection methods. Those evaluating the performance of these early NAATs were faced with the challenge of determining whether this increment in positive results reflected increased sensitivity or false-positive results. An early effort to determine the correct classification was through the use of discrepant analysis. The specimens that were NAAT positive and culture negative were retested using DFA at first and then with other methods. This showed that most of the excess NAAT positive results were true positives and resulted in excellent numbers for sensitivity and specificity for the new NAATs (9). Statisticians objected to the use of discrepant analysis as being

upwardly biased, resulting in an overestimate of sensitivity and specificity (10, 11). The test evaluators felt the bias was small and that it was more important to avoid the misclassification of up to 35% of the NAAT positive results as false positives had they not been confirmed by discrepant analysis.

The FDA was placed in an awkward position by these conflicting positions and opted not to accept simple use of discrepant analysis. This created a challenge for the NAAT manufacturers. Becton-Dickinson's strand displacement NAAT, ProbeTec ET, received FDA clearance, and their package insert shows the tortuous path needed to present the performance data of their assay (12). The first evaluation presented was of their assay versus culture, and specificity with swabs collected from men was 91.7% and with male urine samples was 89%. Next, they showed data for the patient infected status (PIS), with some improvement in performance (for example, specificity with male urine was 91.4%). Then, other NAATs were added to the evaluation, and finally, other tests were performed on discordant specimens. The unacceptable specificities of percentages in the low 90s were because all of the NAAT-positive culture-negative results were considered false positives. This would create a quandary for people using the tests, because they had to either be very cognizant of the problems involved in NAAT evaluation or they had to ignore the package insert. A real specificity of 90% to 92% would mean that false positives could easily occur at a rate in excess of the *C. trachomatis* prevalence in the populations being screened, and clinicians would be presented with the problem of having a positive predictive value of <50%. Having a multiple target comparator, as described by Kirkconnell et al., would have prevented this problem. The approach is much cleaner. The actual evolution of NAAT evaluation was more complicated.

Over time, more NAATs to detect *C. trachomatis* received FDA clearance, and our ability to evaluate new NAATs did improve. Cell culture was discarded as a comparator and other NAATs were used. In general, the excellent performance predicted by use of discrepant analysis was upheld, with NAATs typically having sensitivities in the low to mid 90% range and specificities of >98% (9). Still, some compromises were made to include the reality that multiple specimens were being tested on the same patient, and the FDA required the use of the PIS to evaluate newly introduced NAATs. The PIS standard required that positive results had to be obtained by two different NAATs on one or more specimens to classify the patient as infected. The specimens being evaluated from women included endocervical swabs, vaginal swabs, and urine specimens. Unfortunately, in this evaluation using PIS, once the woman was classified as infected, it was assumed that all specimens coming from her should be positive. This is not consistent with the biology of *C. trachomatis* infections, as some women are infected in one anatomic site and not at another. For example, a woman could be infected at just the urethra but not the cervix or vice versa. Thus, specimens that were coming from the uninfected site would be incorrectly classified as false negatives because positive results had been obtained from the other site (13). Another problem with the FDA requirement for 2 different NAATs being positive to declare a patient as infected is the assumption that the 2 assays used for this purpose will have relatively equal sensitivity and specificity. This has not always been the case (14). The single positives from the more sensitive assay would be recorded as false positives, and the assay under evaluation could receive a lower sensitivity than it deserves. In other words, the poorer-performing assay used to create a comparator would dictate the sensitivity. The use of alternative targets in test evaluation could have prevented all of these complications, and evaluations could be conducted on single specimens, making the use of a flawed PIS standard unnecessary.

This paradigm shift to using multiple targets on the same genes or within the same genome as a composite reference standard for the evaluation of new NAATs is a way to avoid these kinds of complications and have an independent assessment of test performance. It is an approach that should see considerable use in the future.

REFERENCES

1. Jensen AJ, Kleaveland CR, Moghaddam A, Haaheim H, Hjelmevoll SO, Skogen V. 2013. *Chlamydia trachomatis*, *Mycoplasma genitalium* and *Ureaplasma urealyticum* among students in northern Norway. *J Eur Acad Dermatol Venereol* 27:e91–e96. <https://doi.org/10.1111/j.1468-3083.2012.04528.x>.
2. Lis R, Rowhani-Rahbar A, Manhart LE. 2015. *Mycoplasma genitalium* infection and female reproductive tract disease: a meta-analysis. *Clin Infect Dis* 61:418–426. <https://doi.org/10.1093/cid/civ312>.
3. Getman D, Jiang A, O'Donnell M, Cohen S. 2016. *Mycoplasma genitalium* prevalence, coinfection, and macrolide antibiotic resistance frequency in a multicenter clinical study cohort in the United States. *J Clin Microbiol* 54:2278–2283. <https://doi.org/10.1128/JCM.01053-16>.
4. Kirkconnell B, Weinbaum B, Santos K, Le Nguyen T, Vinluan B, Astete S, Wood GE, Totten PA, Getman DK. 2019. Design and validation of transcription-mediated-amplification nucleic acid amplification tests for *Mycoplasma genitalium*. *J Clin Microbiol* 57:e00264–19. <https://doi.org/10.1128/JCM.00264-19>.
5. Bullman S, Meyerson M, Kostic AD. 2017. Emerging concepts and technologies for the discovery of microorganisms involved in human disease. *Annu Rev Pathol Mech Dis* 12:217–244. <https://doi.org/10.1146/annurev-pathol-012615-044305>.
6. Newman L, Rowley J, Vander Hoorn S, Wijesooriya NS, Unemo M, Low N, Stevens G, Gottlieb S, Kiarie J, Temmerman M. 2015. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS One* 10:e0143304. <https://doi.org/10.1371/journal.pone.0143304>.
7. CDC. 2017. Chlamydia. Centers for Disease Control and Prevention, Atlanta, GA. <https://www.cdc.gov/std/stats17/chlamydia.htm>. Accessed 1 May 2019.
8. Moncada J, Schachter J, Bolan G, Engelman J, Howard L, Mushahwar I, Ridgway G, Mumtaz G, Stamm W, Clark A. 1990. Confirmatory assay increases specificity of the chlamydiazyme test for *Chlamydia trachomatis* infection of the cervix. *J Clin Microbiol* 28:1770–1773.
9. Schachter J, Stamm WE, Quinn TC, Andrews WW, Burczak JD, Lee HH. 1994. Ligase chain reaction to detect *Chlamydia trachomatis* infection of the cervix. *J Clin Microbiol* 32:2540–2543.
10. Hadgu A. 1999. Discrepant analysis: a biased and an unscientific method for estimating test sensitivity and specificity. *J Clin Epidemiol* 52: 1231–1237. [https://doi.org/10.1016/S0895-4356\(99\)00101-8](https://doi.org/10.1016/S0895-4356(99)00101-8).
11. Miller WC. 1998. Bias in discrepant analysis: when two wrongs don't make a right. *J Clin Epidemiol* 51:219–231. [https://doi.org/10.1016/S0895-4356\(97\)00264-3](https://doi.org/10.1016/S0895-4356(97)00264-3).
12. Becton, Dickenson and Company. 2019. ProbeTec ET Chlamydia trachomatis and Neisseria gonorrhoeae amplified DNA assays. Becton, Dickenson and company, Sparks, MD. [legacy.bd.com/ds/technicalCenter/inserts/3300754JAA\(04\).pdf](https://legacy.bd.com/ds/technicalCenter/inserts/3300754JAA(04).pdf). Accessed 1 May 2019.
13. Moncada J, Schachter J, Hook EW, III, Ferrero D, Gaydos C, Quinn TC, Willis D, Weissfeld A, Martin DH. 2004. The effect of urine testing in evaluations of the sensitivity of the Gen-Probe Aptima Combo 2 assay on endocervical swabs for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*: the infected patient standard reduces sensitivity of single site evaluation. *Sex Transm Dis* 31:273–277. <https://doi.org/10.1097/01.OLQ.0000124611.73009.D5>.
14. Schachter J, Hook EW, Martin DH, Willis D, Fine P, Fuller D, Jordan J, Janda WM, Chernesky M. 2005. Confirming positive results of nucleic acid amplification tests (NAATs) for *Chlamydia trachomatis*: all NAATs are not created equal. *J Clin Microbiol* 43:1372–1373. <https://doi.org/10.1128/JCM.43.3.1372-1373.2005>.