

Addressing the shortcomings of ANA testing by IFA

Charna Albert

March 2020—Standardizing indirect immunofluorescence testing for antinuclear antibodies is a critical task for the laboratory community, and it's more urgent now that new classification criteria make positive ANA a key factor in diagnosing lupus, said Mark H. Wener, MD, in a session at last year's American Association for Clinical Chemistry annual meeting.

The immunofluorescence assay (IFA) using the cultured HEp-2 cell line is the traditional preferred ANA screening method that the American College of Rheumatology has recommended, said Dr. Wener, professor in the Department of Laboratory Medicine and adjunct professor, Department of Medicine, University of Washington School of Medicine. In fact, the ACR considers it the reference screening method, as does the World Health Organization and the European Autoimmunity Standardization Initiative. According to international guidelines for autoantibody assessment (Agmon-Levin N, et al. *Ann Rheum Dis*. 2014;73[1]:17–23), “if the clinical suspicion is strong and the alternative method is negative, it's mandatory to perform IFA,” said Dr. Wener, who is also co-director of UW Medical Center's scleroderma clinic and director of immunology and the clinical labs at UW Medical Center.

Yet rheumatologists have expressed concerns about inconsistent results of ANA testing by IFA. Some of what's heard in the rheumatology community: “You're doing it wrong”; “We can't trust your results”; “Your ANAs are not always positive in lupus” (though they are expected to be).

“What is the laboratory community's response?” Dr. Wener asked. “Is IFA still the gold standard? Is it still useful or is it outmoded? Can it be improved or re-tooled? How do we integrate ANA IFA with other tests?”

“I think many would say ANA by IFA is not much of a gold standard,” he said, “and we have alternative ANA methods.” They include antigen-specific solid-phase technologies, like ELISA, multiplex bead systems, and solid-phase fluorescence immunoassays. But whether solid-phase assays should be used to screen for ANA in lieu of immunofluorescence remains an open and controversial question.

“There's a mixed response to this question,” Dr. Wener said. The ANA IFA test can be improved. “In fact, if I had my choice, we would be calling this the American Association for Clinical Alchemy, because I'm going to try to convince you that we can take this imperfect ANA IFA test and make it more perfect, if not 24-karat gold.”

The ACR convened a committee (of which Dr. Wener was a member) more than a decade ago to discuss the burgeoning use of non-IFA methods for ANA screening. Its members wrote a position paper that reaffirmed IFA as the gold standard in ANA testing at that time, and recommended that laboratories specify on ANA test reports the methods used for screening (Meroni PL, et al. *Ann Rheum Dis*. 2010;69[8]:1420–1422).

The ACR traditionally has objected to using non-IFA methods for ANA screening, Dr. Wener said, because enzyme and multiplex immunoassays aren't as sensitive. HEp-2 cells contain more antigens than solid-phase assays. Immunofluorescence has about 95 percent sensitivity for lupus detection; in comparison, the connective tissue disease multiplex screen is 69 percent sensitive, he said, citing a 2018 paper published in *Autoimmunity Reviews* (Bizzaro N, et al. 2018;17[6]:541–547). For scleroderma, sensitivity is 97 percent by IFA and 81 percent by CTD screen. “That's the rheumatologists' concern or complaint if you will,” Dr. Wener said: “*If we replace screening with something besides immunofluorescence, that's a problem.*”

Lupus is challenging to diagnose because, while the condition is relatively rare, its symptoms are common. “But there is a hallmark. A constant feature of lupus is a positive ANA, and sensitive tests for lupus and some other connective tissue diseases depend on a positive ANA.”

But ANA screening by IFA is far from perfect. One solution, Dr. Wener said, is to screen with both IFA and multiplex

testing.



Dr. Wener

"There's additional cost, but if there's high enough pretest probability, I think the combination is helpful," he said. If IFA and multiplex testing give confirmatory results, that are concordant, that provides a strong positive predictive value. In addition, information from multiplex testing can make IFA results more interpretable, and vice versa. For example, he said, low positive antibodies against anti-U1-ribonucleoprotein by multiplex is a fairly common result that is likely to be clinically important only if testing by IFA finds compatible ANA with a nuclear coarse speckled pattern (the pattern associated with MCTD). Low-titer U1-RNP with a negative ANA, on the other hand, probably isn't diagnostically significant and can be misleading.

"Sophisticated clinicians can understand that phenomenon, and the laboratory can help with that interpretation," he added.

For those who work in a clinical laboratory, there's another kind of synergy: The tests can be used "as a quality measure" for each other. "Since specific antibodies such as U1-RNP or centromere should have associated IFA patterns, if they are discordant, that's an alert to the laboratory," he said.

Discordance between different testing methods can be used "to help verify and potentially reset cutoffs of IFA and specific antibody results," he said, because the antibody tests "really ought to match." Long-term, he said, the lab-reported multiplex result can be dependent in part on IFA results—in other words, whether the ANA is positive or negative and the pattern.

An integrated testing approach could pose an opportunity for laboratory directors to expand the indeterminate region of multiplex results, something that's often a problem for interpretation, he said. "It could also be an alert to reconsider the cutoff of positivity for IFA ANA used in an individual lab." The results should be complementary, that is, and if there's too much mismatch over time, that's a signal to reconsider what's being done.

Repeatedly positive low-titer ribonucleoprotein with negative HEp-2 IFA, for instance, suggests that a laboratory could expand the indeterminate RNP range, he said. "We don't want to be too cavalier" about having individual labs change results. But "expanding the indeterminate zone for an analyte or reassessing IFA cutoffs" is something laboratories have the authority to do within their own populations.

"Many of our test results fall within this indeterminate range, and coordinating the results of the IFA with the multiplex result allows us to modify that result carefully as we think appropriate based on clinical data," he added.

A problem the lab community must face, Dr. Wener said, is the lack of standardization in ANA testing among different assays and different laboratories.

A 2018 study found that ANA negativity may vary by vendor and kit (Pisetsky DS, et al. *Ann Rheum Dis*. 2018;77[6]:911-913). Researchers tested sera from 103 patients with established lupus using three different IFA kits and found that "the frequency of ANA negativity varied from 5 to 23 of 103 samples," the authors write. Testing was also performed using ELISA and bead-based multiplex assays; 12 and 14 samples were negative, respectively. The authors say the results call into question whether ANA positivity should be used to determine eligibility for clinical trials.

Discordance in ANA results between laboratories also may have far-reaching clinical implications. Dr. Wener

referenced a 2016 study that measured agreement between paired ANA results from two commercial laboratories. The authors performed a sensitivity analysis to determine the degree of agreement using varying criteria. According to the most conservative definition of agreement—negative testing at both laboratories or positive titers within a twofold range of each other—agreement occurred in 18 percent of paired lab results. Forty-two percent of testing was in agreement according to the most lenient criteria, which defined ANA titers of less than 1:160 as negative and allowed less than or equal to a fourfold difference in titer to be acceptable agreement (Abeles AM, et al. *Clin Rheumatol*. 2016;35[7]:1713-1718).

The authors write, “This finding calls into question the reliability of ANA testing as it is currently performed and suggests that results may in part depend upon the laboratory center to which patients are referred.”

“That’s a problem for the lab community,” Dr. Wener said. “It’s easy to see something that’s very bright, or clearly black and negative, but where’s the endpoint? Where do we draw the line between positive and negative?” In fact, it’s a challenge for fluorescence microscopy in general.

There’s the problem within labs and between labs, “but there’s also a problem between kits and between reagents,” Dr. Wener said, citing a 2012 study published in the *American Journal of Clinical Pathology* (Copples SS, et al. 2012;137[5]:825-830). The authors compared results from five ANA IFA assays using serum samples from patients with a variety of connective tissue diseases, and from 100 healthy control patients. Overall, agreement between the five assays was 78 percent (the assays were considered to be in agreement when they exhibited the same titer and doubling dilution).

“This was a carefully done study within a single lab using a single criterion for cutoff, with different technologists looking at the slides,” Dr. Wener said. “Among lupus patients, the same specimen, depending on what kit was used, might be called 1:80 positive or 1:2560 positive. The same specimen might be called ANA negative, or positive at a titer as high as 1:320.”

Titer quantification is a challenge, he said, but important for clinical interpretation, clinical trials, epidemiologic classification, and drug prescription and payment. “The latter may depend on autoantibody results.”

It’s set to have even greater significance for lupus diagnosis because in 2019, the ACR and European League Against Rheumatism released new lupus classification criteria that establish a positive ANA at a titer of 1:80 as decisive for lupus diagnosis (Aringer M, et al. *Arthritis Rheumatol*. 2019;71[9]:1400-1412).

“Using the previous classification criteria, a positive ANA was one of several equally weighted clinical and laboratory features,” Dr. Wener said. “With the new criteria, unless a patient has a kidney biopsy diagnostic for the presence of lupus nephritis, there must be a positive ANA at a titer of 1:80 or above the 95th percentile for a reference population.”

“It’s almost as though you can’t have lupus by classification criteria if you don’t have a positive ANA of at least 1:80.”

The 2019 criteria, Dr. Wener said, give ANA positivity “a central role—not just a participating role—in lupus classification. In turn, that puts more burden on laboratories to know their ANA method’s 95th percentile reference range cutoff and encourages laboratories to convey that information to clinicians.”

With 1:80 now the entry criteria for lupus, that raises the question, what is a 1:80?

Recommendations published in 2014 said “a proper [screening] ANA by IFA is dependent on reagents, equipment, and other local factors; thus, the screening dilution should be defined locally” (Agmon-Levin N, et al. *Ann Rheum Dis*. 2014;73[1]:17-23). And “an abnormal ANA should be the titer above the 95th percentile of a healthy control population. In general, a screening dilution of 1:160 on conventional HEp-2 substrates is often suitable for ANA detection,” Dr. Wener said, citing the 2014 ANA assessment guidelines.

That recommendation is in conflict with the newly established entry criteria for lupus diagnosis, he said. “On the

one hand, we're saying labs should pick 1:160 for the screening titer. Oh, but by the way, 1:80 is the entry criteria for lupus. Clearly, as a profession we need to clarify what we mean by a positive ANA."

The ACR and EULAR proposed the 1:80 ratio based on a systematic literature review and meta-regression of diagnostic data on the performance of ANA for classifying lupus, Dr. Wener said (Leuchten N, et al. *Arthritis Care Res.* 2018;70[3]428-438). "But the implicit assumption with this analysis is that all IFA assays give the same result."

Given the evidence, he said, "I just don't think that's likely to be true. So there's a heightened need to standardize if we're going to be supporting the clinical groups and epidemiologic groups that are using this titer."

The good news, he said, is that a number of approaches are underway to improve consistency of ANA reporting. For example, the CAP Diagnostic Immunology and Flow Cytometry Committee, to which he is AACC liaison, is "looking into [this] in a more formal way."

Organizations and industry would need to coordinate efforts, he said, adding, "I would think organizations like ACR, EULAR, AACC, and CAP might work with the FDA to do this." He noted a couple of examples—INR for prothrombin time normalization, efforts to standardize tests like cholesterol—and said, "I think it's time for us to think about how to do this for ANA."

Laboratory directors and staff can improve consistency of reporting at individual labs by "knowing the ANA population prevalence using your lab's method," he said. Laboratories should also report the ANA method used for screening—in fact, in 2019 the CAP added a new Laboratory Accreditation Program checklist requirement that says laboratories should include on the ANA report a description of the method used for ANA screening (if the method is not explicit in the test name) (IMM.39700).

Automation is another path to improved consistency among laboratories, Dr. Wener said. "Automated instruments set thresholds for positivity based on fluorescence intensity. Essentially, a single point calibration above or below the cutoff is what's considered positive. This is coordinated with fluorescence light intensity."

But individual labs can do nearly the same thing by having an endpoint calibrator or single point calibration above or below a threshold, he said. "The current positive and negative controls are rarely used at this threshold level. But labs can develop or purchase endpoint calibrators that would serve this role."

The pivotal question is whether advanced automation can be used to address ANA by IFA testing's shortcomings, said Melissa Snyder, PhD, co-director of Mayo Clinic's antibody immunology laboratory, who presented during the same AACC session on whether automation can bring ANA testing "out of the dark room and into the modern laboratory."

The 2014 study by Bizzaro, et al., that compared six automated platforms—Aklides, EuroPattern, Nova View, Helios, Zenit G-Sight, and Image Navigator—found about 90 percent agreement on positivity (*Autoimm Rev.* 2014;13[3]:292-298). The authors sent 144 ANA sera to six laboratories for manual ANA IFA testing, identified a consensus result for each sample (excluding 17 positive and six negative samples for which no consensus could be reached), and then repeated testing on the six automated platforms. There was more variability among the systems on the negative samples, ranging from 79 percent to 94.1 percent agreement.

"So good consensus on the positive agreement, a little less on the negative agreement," Dr. Snyder said.

Bizzaro, et al., also compared estimated titer to manual titer and automated pattern interpretation to manual interpretation, Dr. Snyder said. Titer agreement (among five platforms only), which the authors measured using a Spearman's rho calculator, ranged from .627 to .839. The platforms had greater variability with regard to pattern agreement (four were compared), ranging from 50 to 80 percent.

They also looked at a comparison of light intensity unit as a positive/negative cutoff, and how varying the light intensity unit affected sensitivity and specificity as compared with the consensus result. There was "fairly decent

standardization there as well.”

“What I take away from this is that these systems can give us a bit of help with standardization on positive/negative agreement,” Dr. Snyder said. Pattern and titer agreement, on the other hand, “is something we could still work on.”

Dr. Snyder’s laboratory uses an advanced automation platform to perform ANA testing. The system automates not only slide and sample processing but also slide interpretation, pattern identification, and titer estimation, “based on the fluorescence intensity read from the digital image” rather than serial dilution.

While “we certainly have reduced our technologist’s time in terms of reading, it’s critical to note that you still need technologist expertise with these systems,” she said. Mayo Clinic technologists review results from the automated readers, focusing on positive/negative interpretation and pattern. “They might agree with what the computer calls, or they might disagree, at which point they would make a change.”

To assess the performance of the automated system, her laboratory collected data on 1,559 ANA samples submitted for IFA testing and compared the automated slide reader’s interpretations with the results that were eventually released to the clinical record.

The laboratory’s technologists and the automated slide reader agreed on almost 100 percent of the negative samples. (The slide reader identified 909 samples as negative; technologists didn’t identify any of those samples as positive but repeated testing on two samples.) However, they confirmed as negative 26 percent of the samples the computer identified as positive.

“The cutoff for positive/negative on the computer may be set a little on the low side,” Dr. Snyder speculated. Overall, positive/negative agreement between the manual and automated interpretation was 86.6 percent.

Pattern agreement was 45 percent. “This is very much in line with what was studied by Bizzaro’s group,” Dr. Snyder said. In the majority of cases in which technologists disagreed with the automated slide reader on pattern, the sample was ultimately determined to be negative.

Overall, automated systems can lead to improved qualitative agreement, she said, while improvements in pattern and titer agreement “have not yet been realized.”

“We are seeing a little bit more objectivity in our interpretation, particularly in our positive/negative agreement.” But the expertise of the technologists is still a critical component to performing ANA by IFA testing, she said, even with automation of slide reading.□

Charna Albert is CAP TODAY associate contributing editor.