

# Data spark new directions in cervical cancer

## William Check, PhD

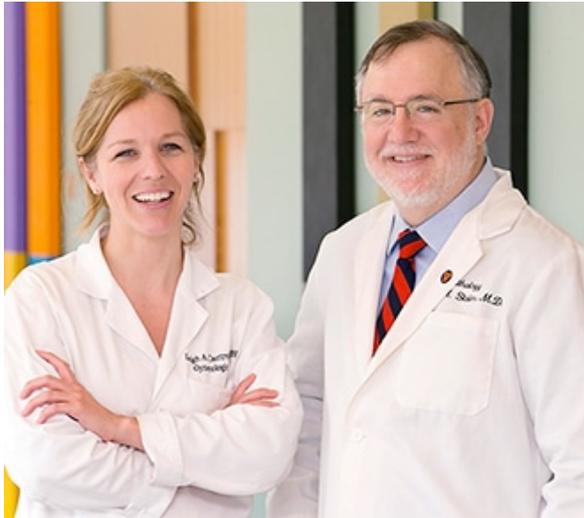
June 2014—When Mark Stoler, MD, stood up to speak at the 30th annual Clinical Virology Symposium on April 29, his topic was timely. Dr. Stoler was presenting three-year followup data from the ATHENA trial, in which a primary human papillomavirus screening algorithm based on the Roche Cobas HPV assay was compared with traditional cytology and a hybrid cotesting algorithm for their ability to prevent cervical cancer. On March 12, Dr. Stoler, a cytopathologist and professor emeritus of pathology and clinical gynecology at the University of Virginia, was part of the team that presented much of the same data to a 13-member FDA advisory committee, which unanimously recommended that the agency approve the primary HPV screening algorithm used in ATHENA. Six weeks later, on April 24, the FDA agreed with its committee.

Five days after the FDA approval, Dr. Stoler presented results from the trial of 40,901 women age 25 and older recruited at the time of undergoing routine screening for cervical cancer. Summarizing the conclusions from the ATHENA trial, Dr. Stoler said the data demonstrated that:

- Primary HPV testing was significantly superior to cytology and cotesting for detecting CIN3 or worse. It not only found more disease but also found it earlier.
- Specificity of HPV testing was at least equal to that of cytology when HPV genotyping and reflex cytology were added.
- Positive predictive value (PPV) and positive likelihood ratio (PLR) of the primary HPV algorithm were twice that of cytology.
- Negative predictive value (NPV) and negative likelihood ratio (NLR) of the primary HPV algorithm were improved over both cytology and cotesting.

Taken together, these findings showed that the HPV primary screening algorithm provides “a better balance of resource management,” Dr. Stoler said. Compared with cytology alone, it triggers more colposcopies but finds much more disease. Compared with hybrid cotesting, it triggers slightly more colposcopies but finds slightly more disease and requires far fewer total screening tests.

Dr. Stoler added that the reported positive and negative likelihood ratio values are important because these parameters are independent of prevalence, so the data can be applied to any population.



Primary HPV testing with genotyping and reflex cytology simplifies algorithms, says Dr. Mark Stoler, here with gynecologic oncologist Leigh Cantrell, MD, MSPH, of the University of Virginia.

At the March 12 FDA hearing, R. Marshall Austin, MD, PhD, had spoken in favor of FDA-approved cytology and HPV cotesting as the best cervical screening practice. In an interview with CAP TODAY, Dr. Austin, professor of pathology and director of cytopathology at the University of Pittsburgh Medical Center, said, “I think it is better to do cotesting. I talked to the medical staff about it this morning, and they were quite supportive.”

Dr. Austin’s primary reason for preferring cotesting: “You can get significantly more disease with cotesting than with just an HPV test alone.”

“I tell gynecologists and other clinicians, ‘What you could elect to do with primary HPV screening is still somewhat unproven. We will have a chance to see how this new test performs in the field. We will collect data to gauge protection from cervical cancer, which now we can only do by inference.’”

Barbara A. Crothers, DO, chair of the CAP Cytopathology Committee, also has reservations about the primary HPV algorithm. “At this point we need to be cautious about implementing this new algorithm on a widespread basis,” Dr. Crothers, who is pathology program director and director of cytopathology in the Department of Pathology and Area Laboratory Services at Walter Reed National Military Medical Center, said in an interview. (Dr. Crothers emphasized that she was representing the CAP Cytopathology Committee, not the military.) “We have some concerns about how it will be implemented in a clinical scale setting.”



Dr. Crothers

Dr. Crothers says there are “still unanswered issues and questions about the primary HPV screen for cervical cancer.” For example: “Will the results [from the ATHENA trial] apply in everyday clinical settings without the benefit of a controlled environment with patient recall and followup? And what effect will the subjectivity of colposcopy and cervical biopsy interpretation have on the correct identification of high-grade premalignant processes and followup for women?” These issues have not been sufficiently explored, she says.

Mona Saraiya, MD, MPH, was a member of the FDA advisory committee that recommended approval of the HPV

assay and algorithm. As an associate director in the Office of International Cancer Control at the Centers for Disease Control and Prevention, Dr. Saraiya works on assessing major medical practices in the U.S. with regard to cervical cancer screening, monitoring them to ensure that guidelines from major medical groups are being followed.



Dr. Saraiya

“Data from the three-year followup of ATHENA opens up an opportunity for and opens the door” to a primary HPV screening algorithm, Dr. Saraiya tells CAP TODAY. “The next step is for professional groups to evaluate the data.” Dr. Saraiya believes that guidelines should be based not just on ATHENA data, but that “ATHENA in conjunction with randomized clinical trials from Europe and other relevant data should be used to make decisions about the best way forward.” Even before ATHENA, she says, in the U.S. and Europe there was fairly strong evidence about the negative predictive value of primary HPV testing—“just some concern about how to manage the HPV positives.”

To Dr. Saraiya, the issue is the negative predictive value of HPV testing and what it means in terms of extending screening intervals in practice. Because of the high sensitivity of an HPV test, a woman with a negative test has a very low risk of developing cervical cancer in the next three to five years, raising the possibility of safely extending the recommended screening interval. Dr. Saraiya sees “a lot of promise with primary HPV testing,” saying, “We need to examine whether primary HPV testing with extended screening intervals can really play that role in the United States, a largely annual Pap-based opportunistic country.”



Dr. Massad

L. Stewart Massad Jr., MD, another member of the FDA advisory committee, says the three-year followup data demonstrated that HPV-based screening using the Cobas HPV test and the algorithm Roche proposed had good sensitivity and specificity, “with reassuringly low risk for CIN3+ out to the end of the three-year followup.”

“The vote of the FDA expert panel was strictly on the question of safety and effectiveness. Cost-effectiveness and comparative effectiveness are not part of the FDA review process and weren’t assessed,” says Dr. Massad, a professor of obstetrics and gynecology in the Division of Gynecologic Oncology at Washington University School of Medicine. Professional societies and the American Cancer Society will need to review those issues, he adds.

Dr. Stoler began his presentation at the Clinical Virology Symposium with a respectful but clear-eyed nod to cytology, which has accomplished so much in the past 60 years despite its limitations. “Cytology has been quite successful,” Dr. Stoler said, and screening with cytology at three-year intervals is still recommended for cervical cancer protection. More than half of the 12,000 cervical cancer cases each year in the U.S. occur in women who haven’t had a Pap test in five years or haven’t been screened at all.

Pap screening has reduced cervical cancer incidence despite the fact that cervical cytology is relatively insensitive for detecting high-grade cervical cancer precursors, Dr. Stoler continued. A 1999 review estimated the sensitivity

of the conventional Pap test at about 50 percent. In the four laboratories doing Pap reading in ATHENA, sensitivity for CIN3 or worse ranged from 42 percent to 73 percent, the same range found in six studies from around the world—44 percent to 74 percent. These figures document that Pap testing has not only low sensitivity but also low reproducibility, he said. Sensitivity in the ATHENA laboratories correlated directly with the rate of slides called  $\geq$  ASC-US and inversely with specificity, underscoring the tradeoff inherent in reading Pap tests.

In contrast, sensitivity of HPV in ATHENA was higher and confined in a much narrower band—88 percent to 90 percent. So the HPV assay is much more sensitive and much more consistent than cytology. This finding is consistent with a review that found that HPV testing had an average 35.7 percent increase in sensitivity for  $\geq$  CIN2 over cytology in six published studies (Whitlock EP, et al. *Ann Intern Med.* 2011;155:687-697). Based on this concept as part of the systematic literature review performed to inform the guidelines, screening recommendations were augmented in March 2012 to include cotesting every five years for women ages 30 to 65 as the “preferred” algorithm, with cytology becoming an “acceptable” method.

Dr. Stoler noted problems with cotesting: “It is expensive,” he said, and “it isn’t logical. It combines a relatively insensitive test with a highly sensitive test.” Moreover, he added, it leads to “an incredibly complex set of algorithms” for managing abnormalities (one guideline has 18 algorithms), and complexity can lead to mismanagement. Perhaps this is why cotesting hasn’t been widely adopted. “More than half of the women in this country who are screened are still only getting Pap smears,” Dr. Stoler said.

In contrast, primary HPV testing with genotyping and reflex cytology dramatically simplifies algorithms relative to hybrid cotesting, he said. If a woman has genotypes 16/18, she is referred for colposcopy; if she is positive for any of the 12 other high-risk genotypes contained in the HPV assay, she has reflex cytology. If the Pap test is  $\geq$  ASC-US, the woman is referred to colposcopy. Women negative on the initial HPV test continue routine triennial screening.

In a process called verification bias adjustment, or VBA, about 900 women in ATHENA who were HPV negative and negative for intraepithelial lesion or malignancy were also referred to colposcopy. The results were used to correct the raw sensitivities. “No test is perfect,” Dr. Stoler explained. “All miss some disease. This is a process for finding how much disease is still in people who you think test negative or normal.”

In the three-year followup ATHENA results, VBA sensitivities and specificities, respectively, looked like this:

- Primary HPV algorithm: 58.26 percent, 95.91 percent.
- Cytology alone: 41.71 percent, 96.88 percent.
- Hybrid cotesting: 53.22 percent, 95.80 percent.

It is from these numbers that NPV and PPV and the ratios—PLR and NLR—were calculated, showing that the HPV algorithm had significantly improved effectiveness and safety over cytology and behaved similar to or even better than cotesting.

Based on these performance characteristics, the following cross-sectional measures of clinical disease management were reported at the FDA:

- Primary HPV screening detects 232  $\geq$  CIN3 using 44,057 screening tests and 1,890 colposcopies, for an 8.1 colposcopy/CIN3+ ratio.
- Hybrid cotesting detects 211  $\geq$  CIN3 using 75,574 screening tests and 1,916 colposcopies, for a 9.1 colposcopy/CIN3+ ratio.

“The primary HPV algorithm with a three-year interval is better than cotesting at a five-year interval, which is the current guideline,” Dr. Stoler said. “The primary HPV algorithm with a three-year interval is about equivalent to cotesting at a three-year interval, which is not in the guideline.”

Dr. Stoler added one more idea, which he believes to be critical. “It is increasingly apparent that there is one key educational message that must be driven home, and it is one we have been trying to get out there forever,” he said. “Clinically valid HPV testing does not just detect the virus. That is the No. 1 misperception in the entire field. It detects the virus at a clinically valid cutoff that optimizes the detection of cancer and precancer while providing maximal reassurance to the people who are not at risk and can be followed at long intervals.”

Dr. Austin disagrees with the basic data and the main conclusion Dr. Stoler presented. “From ATHENA data on the FDA website, we can see that more disease was detected with cotesting,” he tells CAP TODAY. Dr. Austin was referring to a table in the summary of the advisory committee meeting, <http://j.mp/fda-cobas>, which provides the following figures for VBA sensitivity:

- 58.26 percent for the HPV algorithm (page 3).
- 61.16 percent for a cotesting algorithm (page 66, Table A8.2).

The figure for the primary HPV algorithm agrees with what Dr. Stoler presented, but for cotesting Dr. Austin’s VBA sensitivity, 61.16 percent, is much higher than the figure that Dr. Stoler cited, 53.22 percent. Can both numbers be describing the same measurement? As it happens, the answer is no. Here is the FDA’s explanation (pages 11–12) of the algorithm that gave rise to the figure for cotesting in Table A8.2:

“In Appendix 8 of this submission, this same algorithm [the standard cotesting algorithm] is presented with co-test results for everyone 25 and older. This is not a candidate algorithm as it does not represent a primary HPV screening claim, nor is it a comparator algorithm since it is not a current acceptable screening paradigm. It is provided simply to illustrate the impact of including the entire proposed screening population ( $\geq 25$  year old women) under the ASC-US Triage and NILM HPV 16/18 positive genotyping paradigm, which may help in evaluating the appropriateness of the proposed age range for the new indication.”

On page 3 of the FDA document, the agency makes the same conclusion about primary HPV screening (“Candidate algorithm”) relative to cotesting (“Additional Comparator”) that Dr. Stoler drew: “The Candidate algorithm is better than the Additional Comparator for women  $\geq 25$  years of age in the major performance characteristics (PPV, NPV, PLR and NLR) for both  $\geq$  CIN2 and  $\geq$  CIN3, and these improvements are statistically significant at the 95% confidence level.”

Dr. Austin posed other arguments for the superiority of FDA-approved cotesting of women 30 and older. In a Kaiser Permanente Northern California study published in 2011, cytology, cotesting, and primary HPV screening were evaluated in more than 330,000 actual patients followed for five years (see “Kaiser study” at right). “In this paper they estimated the amount of invasive cancer there would have been if they had acted on Pap alone or HPV alone versus cotesting,” Dr. Austin said in an interview. “The lowest estimated cancer rate was with cotesting. It was 10 percent to 15 percent lower than if they had just relied on HPV results.”

Not surprisingly, Dr. Stoler has a different view. “The purpose of screening is not to find cancers,” he told CAP TODAY. “With screening we are trying to find precancers, CIN3 or worse, and treat them. That is why invasive cancer decreased with cytology: We find precancer and treat it before it develops into cancer.” In both ATHENA and the Kaiser study, detection of precancer with HPV screening was equal to or better than with cotesting. That may be why the Kaiser investigators concluded:

“Furthermore, our findings suggest . . . that HPV testing without adjunctive cytology might be sufficiently sensitive for primary cervical cancer screening.”

In the Kaiser data, Dr. Austin says, among women who developed invasive cancer in the next five years after cotesting, 31 percent had a baseline negative HPV result. If persistent HPV infections—10 years and longer—are assumed and if FDA-approved HPV tests are reliable, then why do women developing cervical cancer over the next several years test negative at baseline? Sampling challenges and low and fluctuating viral load in cancers probably account for this, he says, “and the false-negative results represent a patient safety hazard for patients screened

with extended intervals.”

Dr. Austin also argued that a short-term trial like ATHENA cannot establish that a screening strategy can prevent cancer. “A lot of people don’t even think about the fact that screening is all about protection from cervical cancer,” he says. “They look only at rates of detection of CIN2 and CIN3. The only way to document protection from cervical cancer is long-term observational studies on populations. You can’t conclude anything about protection from cancer in a three-year trial.”

FDA advisory committee member Dr. Massad says cancer protection was addressed: “CIN3+ is the standard surrogate marker for cancer throughout the world. HPV screening has been shown to prevent cancer in the Ronco trial, and early detection/treatment of CIN3 worked in that study” (Ronco G, et al. *Lancet Oncol.* 2010; 11:249-257).

Commenting on the deliberation phase that will follow the FDA’s approval, Dr. Crothers says one trial is not enough. “Most of the time, CAP and other professional organizations like to have lots of solid evidence, that maybe doesn’t come from the vendor, that this is an appropriate algorithm for our patients. At this time that evidence is lacking.”

In addition to the Kaiser study, several European trials have demonstrated the efficacy of a primary HPV screening approach (e.g., Naucler P, et al. *J Natl Cancer Inst.* 2009;101:88-99; Dillner J, et al. *BMJ.* 2008;337:a1754; also the Ronco trial cited previously). Do these qualify as part of the evidence base? “Those were done in national screening programs where women receive notification that they require followup and have access to that health care at very low cost,” Dr. Crothers says.

She notes another possible drawback to the proposed program. “One of the advantages of implementing primary HPV screening is the cost savings, since you can extend the screening interval to three years,” she says. But the United States doesn’t have a national screening program. Without such a program, she notes, “we cannot be certain that women will have access to and return at appropriate intervals for screening, and longer intervals could result in women inadvertently delaying appropriate screening.”

Dr. Massad notes that the currently recommended interval for Pap screening is three years. “Both the U.S. Preventive Services Task Force and the American Cancer Society recommend against annual cervical screening based on extensive evidence assessment,” he says. “Annual screening will increase the detection of lesions fated to regress if left undiagnosed. Women and the clinicians who care for them will believe they’ve averted cancer by treating these, without actually having any impact.”

“In my opinion,” Dr. Massad says, “based on data from trials other than ATHENA, the optimal interval for primary HPV screening is five years, not three years. European studies suggest that five-year followup is optimal, with strong NPV out to at least five years following a negative HPV test, as do data presented from the Kaiser Permanente Northern California database used to calculate risk estimates for the American Society for Colposcopy and Cervical Pathology’s management guidelines. These studies were with other assays, but the performance of the Cobas HPV test at three years seems to be at least as good.”

Dr. Crothers says several organizations have assembled a working group to look at interim guidelines. The Society of Gynecologic Oncology and the ASCCP are leading the effort to review the efficacy of primary HPV screening and to suggest how to incorporate the HPV test into screening practice. “I believe that primary HPV screening will be an option for a subset of women,” Dr. Crothers says.

If primary HPV screening does move into practice, it will have a major impact on cytopathology practice. In the core population—25 to 65 years—15 percent of women will be HPV positive. “So 85 percent of Pap smears will drop out,” Dr. Stoler says. Authors of the Kaiser study estimated a 95 percent reduction (see page 48).

Taking Pap tests and HPV tests together, Dr. Stoler says, “Integrated labs will do about the same number of tests. At any academic institution you have a cytology lab and a clinical molecular pathology lab or a cytology lab doing

HPV testing. The only places that will lose out on test volume are labs that only do cytology.”

Dr. Crothers notes the importance of any situation in which the result of only one test is relied on to make a primary decision. “[The Roche assay] is the only FDA-approved test for primary HPV screening. That is critically important,” she says. “Most clinicians and even some pathologists may not be aware of that. There are a number of platforms for HPV testing. However, we should not be using other methods at this point, only this Roche method, since it has been validated in a clinical trial. We don’t know how those other methods perform in a clinical setting.”

Dr. Stoler agrees. He was asked about this issue after his symposium presentation. “We can’t say, Do HPV screening with any test you want,” he responded. While some HPV tests have been shown to have 90 percent to 95 percent equivalence to the Roche Cobas assay, he said, “The bar for HPV tests in cervical screening is higher than that.” In an interview, he added: “In primary HPV screening, where you don’t have cytology as your backup, you really want to know you are performing like the HPV assay performed in the ATHENA trial.”

“The FDA held Roche to this very high standard of doing a clinical trial and three-year followup before considering a claim of primary screening,” he continues. “The question for other FDA-approved tests is whether they will perform in the same manner. The answer is maybe, but we don’t know for sure. Each manufacturer will have to prove its case.”

Primary HPV testing is unlikely to be adopted quickly in the United States. However, FDA endorsement of the test in a cervical cancer screening context may give a boost to those who use cotesting. Once they become comfortable with HPV test performance in that setting, eventually they may switch over for cost reasons.

“No matter how you analyze it,” Dr. Stoler says, “a good, clinically validated HPV-driven algorithm outperforms cytology and is about equal to cotesting with minor adjustments by interval.” For him, at least, it’s about “doing better for patients or even as good while taking out subjectivity, redundancy, and complexity.”

*William Check is a writer in Ft. Lauderdale, Fla.*

---

## **Kaiser study: single negative HPV test is sufficient reassurance**

Many of the conclusions from the ATHENA study were mirrored independently in an even larger study of a different kind published a few years earlier.

Kaiser Permanente Northern California in 2003 adopted a screening program for cervical cancer based on cotesting, with extended screening intervals for women with normal cytology who test negative for HPV. After the program had been in operation for several years, Kaiser physicians teamed up with epidemiologists from the National Cancer Institute to evaluate its efficacy. They assessed the five-year cumulative incidence of cervical cancer and CIN3 or worse among more than 330,000 women 30 and older who enrolled in the program (Katki HA, et al. *Lancet Oncol.* 2011;12: 663-672).

As they wrote in the introduction, they believed that the results would have wider application:

“The KPNC experience serves as a large-scale demonstration project of what could realistically be achieved in routine clinical practice, where providers receive no special training and do not need any special qualifications to participate and that no provider, provider group, patient, or group of patients is excluded.”

Here are the core findings from the assessment:

“In 315,061 women negative by HPV testing, the five-year cumulative incidence of cancer was 3.8 per 100,000 women per year, slightly higher than for the 306,969 who were both negative by HPV and Pap testing (3.2 per 100,000), and half the cancer risk of the 319,177 who were negative by Pap testing (7.5 per 100,000).”

While both HPV testing and cotesting were highly superior to Pap testing, the authors considered the cancer incidence among the HPV and cotesting groups to be equivalent and concluded that primary HPV screening as a single test was as effective as cotesting:

“A single negative test for HPV was sufficient to reassure a woman of extremely low risk of CIN3 or cancer for 5 years. We identified that negative cytology provided no extra reassurance against cancer beyond that conferred by a negative HPV test result.”

While performing a Pap test along with an HPV test identified an additional 27 (four percent) of the 747 cases of CIN3/AIS (adenocarcinoma in situ), over the longer term this made no practical difference. Doing 330,000 additional Pap tests to detect an additional 27 cases of CIN3/AIS would not be an effective use of health care resources. Therefore, they wrote:

“Our findings strongly suggest that primary HPV testing, with a positive test for HPV triaged by cytology (or other tests with high specificity), a strategy that might preserve nearly all the safety of co-testing while reducing the number of Pap tests by 95% in our population, could be more efficient than co-testing—as has been suggested by others.”

In the study’s discussion section, the authors reiterated their broader conclusion:

“[W]e believe that the KPNC experience serves as a large-scale demonstration project of what could realistically be achieved in real-life clinical practice.” —*William Check, PhD*